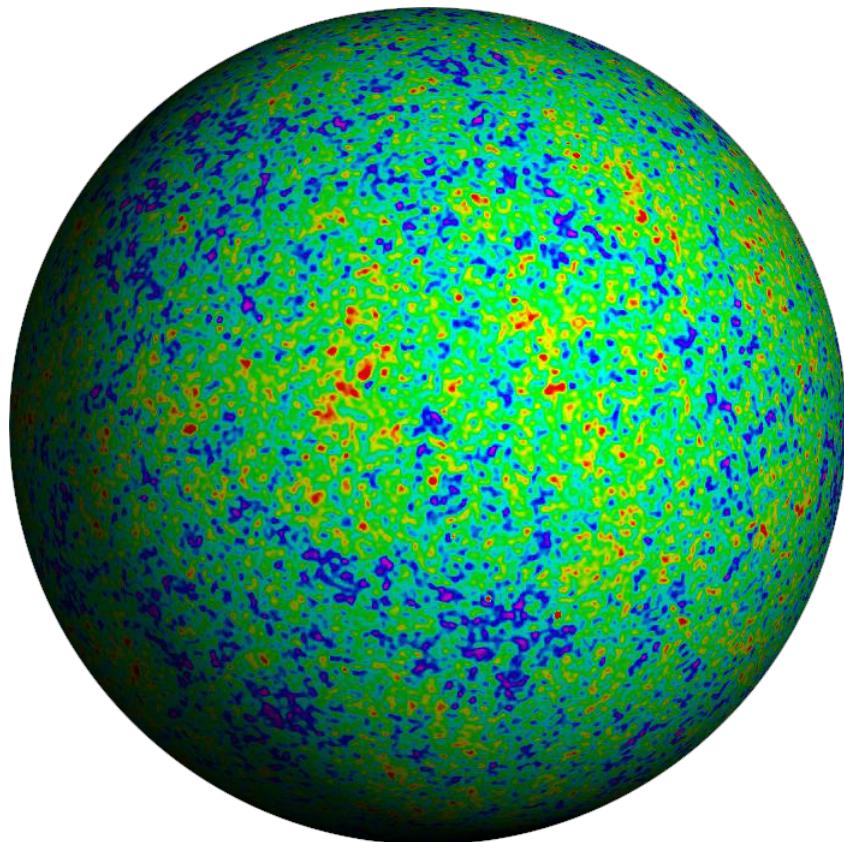


Cosmology

Christoph Weniger, adopted from Daniel Baumann¹

*Institute of Theoretical Physics, University of Amsterdam,
Science Park, 1090 GL Amsterdam, The Netherlands*



¹The course notes are largely identical to the ones from Daniel Baumann last year, see <http://cosmology.amsterdam/education/cosmology/>. However, some of the material changed or was added new. Only content of the *current* course notes is relevant for the exam.

Contents

Preface	1
I The Homogeneous Universe	3
1 Geometry and Dynamics	4
1.1 Geometry	4
1.1.1 Metric	4
1.1.2 Symmetric Three-Spaces	5
1.1.3 Robertson-Walker Metric	7
1.2 Kinematics	9
1.2.1 Geodesics	9
1.2.2 Redshift	12
1.2.3 Distances	13
1.3 Dynamics	16
1.3.1 Matter Sources	16
1.3.2 Spacetime Curvature	20
1.3.3 Friedmann Equations	22
2 Inflation	26
2.1 The Horizon Problem	26
2.1.1 Particle Horizon	26
2.1.2 Hubble Radius	27
2.1.3 Why is the CMB so uniform?	28
2.2 A Shrinking Hubble Sphere	29
2.2.1 Solution of the Horizon Problem	29
2.2.2 Conditions for Inflation	30
2.3 The Physics of Inflation	31
2.3.1 Scalar Field Dynamics	31
2.3.2 Slow-Roll Inflation	33
2.3.3 Reheating	35
3 Thermal History	36
3.1 The Hot Big Bang	36
3.1.1 Local Thermal Equilibrium	36
3.1.2 Decoupling and Freeze-Out	38
3.1.3 A Brief History of the Universe	39
3.2 Equilibrium	41
3.2.1 Equilibrium Thermodynamics	41
3.2.2 Densities and Pressure	44
3.2.3 Conservation of Entropy	49

3.2.4	Neutrino Decoupling	51
3.2.5	Electron-Positron Annihilation	52
3.2.6	Cosmic Neutrino Background	53
3.3	Beyond Equilibrium	55
3.3.1	Boltzmann Equation	55
3.3.2	Dark Matter Relics	56
3.3.3	Recombination	59
3.3.4	Big Bang Nucleosynthesis	63
II	The Inhomogeneous Universe	70
4	Cosmological Perturbation Theory	71
4.1	Metric Perturbations	71
4.2	Matter Perturbations	74
4.3	Equations of Motion	76
4.3.1	Conservation Equations	76
4.3.2	Einstein Equations	79
4.3.3	Metric Evolution	82
4.4	Initial Conditions	83
4.4.1	Adiabatic Fluctuations	83
4.4.2	Curvature Perturbations	84
4.4.3	Statistics	85
4.5	Summary	86
5	Structure Formation	88
5.1	Gravitational Instability	88
5.1.1	Jeans' Instability	88
5.1.2	Clustering of Dark Matter	89
5.1.3	Matter Power Spectrum	91
5.2	Acoustic Oscillations	92
5.2.1	Radiation Fluctuations	92
5.2.2	Primordial Sound Waves	94
5.3	CMB Anisotropies	95
5.3.1	Motion of the Solar System	95
5.3.2	Perturbed Photon Geodesics	96
5.3.3	Line-of-Sight Solution	98
5.3.4	CMB Power Spectrum	100
6	Quantum Initial Conditions*	102
6.1	From Quantum to Classical	102
6.2	Classical Oscillators	104
6.2.1	Mukhanov-Sasaki Equation	104
6.2.2	Subhorizon Limit	106
6.3	Quantum Oscillators	106
6.3.1	Canonical Quantisation	106

6.3.2	Choice of Vacuum	107
6.3.3	Zero-Point Fluctuations	108
6.4	Quantum Fluctuations in de Sitter Space	109
6.4.1	Canonical Quantisation	109
6.4.2	Choice of Vacuum	110
6.4.3	Zero-Point Fluctuations	111
6.4.4	Quantum-to-Classical Transition*	112
6.5	Primordial Perturbations from Inflation	112
6.5.1	Curvature Perturbations	112
6.5.2	Gravitational Waves	114
6.6	Observations	115
6.6.1	Matter Power Spectrum	115
6.6.2	CMB Anisotropies	116
A	Elements of General Relativity	118
A.1	Gravity as Geometry	118
A.1.1	The Happiest Thought	119
A.1.2	Bending of Light	119
A.1.3	Gravitational Redshift	119
A.1.4	Gravitational Time Dilation	120
A.1.5	Curved Spacetime	121
A.2	Geodesic Equation	123
A.2.1	Timelike Geodesics	123
A.2.2	Null Geodesics	125
A.3	Einstein Equation	126
A.3.1	Curvature	126
A.3.2	Energy-Momentum	128
A.3.3	Einstein Equation	130

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 the large-scale structure of the universe. 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.

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 $ds^2 = dt^2 - d\mathbf{x}^2$ for Minkowski space. Spacetime four-vectors will be denoted by capital letters, e.g. X^μ and P^μ , where the Greek indices μ, ν, \dots run from 0 to 3. We will use the Einstein summation convention where repeated indices are summed over. Latin indices i, j, k, \dots will stand for spatial indices, e.g. x^i and p^i . Bold font will denote spatial three-vectors, e.g. \mathbf{x} and \mathbf{p} . We will use η for conformal time and τ for proper time.

Acknowledgements Thanks to Mustafa Amin and Paolo Creminelli for comments on a previous version of these notes. Adam Solomon helped designing the problem sets and writing some of the solutions. Thanks to the many students worldwide who emailed me questions and corrections.

Material marked with an asterix (*) is not relevant for the exam.

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 *homogeneous* and *isotropic*, i.e. independent of position and direction. As we will see, in §1.1, homogeneity and isotropy single out a unique form of the spacetime geometry of the universe. We will discuss how particles and light propagate in this spacetime in §1.2. Finally, in §1.3, we will show how the equations of general relativity relate the rate of expansion of the universe to its matter content.

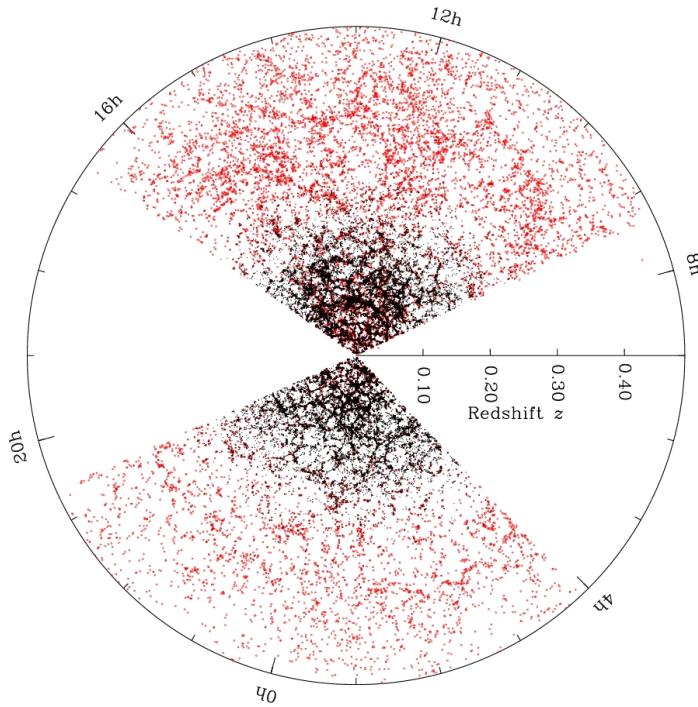


Figure 1.1: The distribution of galaxies is clumpy on small scales, but becomes more uniform on large scales and at early times.

1.1 Geometry

1.1.1 Metric

I am assuming you have seen a metric before. (Otherwise, we will be in trouble.) Just to remind you, the metric is an object that turns coordinate distances into physical distances. For example, in three-dimensional Euclidean space, the physical distance between two points separated by the infinitesimal coordinate distances dx , dy and dz is

$$d\ell^2 = dx^2 + dy^2 + dz^2 = \sum_{i,j=1}^3 \delta_{ij} dx^i dx^j , \quad (1.1.1)$$

5 1. Geometry and Dynamics

where we have introduced the notation $(x^1, x^2, x^3) = (x, y, z)$. In this simple example, the metric is the Kronecker delta $\delta_{ij} = \text{diag}(1, 1, 1)$. However, you also know that if we were to use spherical polar coordinates instead, the square of the physical distance would no longer be the sum of the square of the coordinate distances. Instead, we would get

$$d\ell^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \equiv \sum_{i,j=1}^3 g_{ij} dx^i dx^j, \quad (1.1.2)$$

where $(x^1, x^2, x^3) = (r, \theta, \phi)$. In this case, the metric has taken a less trivial form, namely $g_{ij} = \text{diag}(1, r^2, r^2 \sin^2 \theta)$. This illustrates that observers using different coordinate systems won't necessarily agree on the coordinate distances between two points, but they will always agree on the physical distance, $d\ell$. We say that $d\ell$ is an *invariant*. Hence, the metric turns observer-dependent coordinates into invariants.

A fundamental object in relativity is the spacetime metric. It turns observer-dependent spacetime coordinates $X^\mu = (t, x^i)$ into the invariant line element¹

$$ds^2 = \sum_{\mu,\nu=0}^3 g_{\mu\nu} dX^\mu dX^\nu \equiv g_{\mu\nu} dX^\mu dX^\nu. \quad (1.1.3)$$

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

$$g_{\mu\nu} = \text{diag}(1, -1, -1, -1). \quad (1.1.4)$$

In general relativity, on the other hand, the metric will depend on the spacetime location,

$$g_{\mu\nu}(t, \mathbf{x}). \quad (1.1.5)$$

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. The large degree of symmetry of the homogeneous universe means that the metric of the expanding universe take a rather simple form.

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). The four-dimensional line element can then be written as

$$ds^2 = dt^2 - a^2(t) d\ell^2, \quad (1.1.6)$$

where $d\ell^2 \equiv \gamma_{ij} dx^i dx^j$ is the line element of a maximally symmetric 3-space and the *scale factor* $a(t)$ describes the expansion of the universe.

1.1.2 Symmetric Three-Spaces

We start with a classification of maximally symmetric 3-spaces. First, we note that homogeneous and isotropic 3-spaces have constant 3-curvature.² There are only three options: zero

¹Throughout the course, we will use the Einstein summation convention where repeated indices are summed over. We will also use natural units with $c \equiv 1$, so that $dX^0 = dt$. Our metric signature will be mostly minus, $(+, -, -, -)$.

²We give a precise definition of Riemann curvature below.

6 1. Geometry and Dynamics

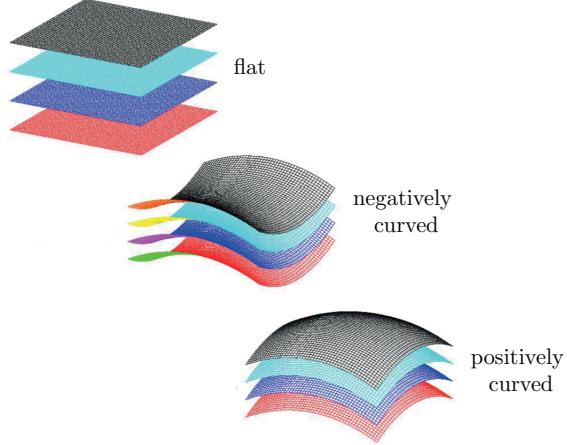


Figure 1.2: The spacetime of the universe can be foliated into flat, positively curved or negatively curved spatial hypersurfaces.

curvature (E^3), positive curvature (S^3) and negative curvature (H^3). The corresponding line elements are

$$d\ell^2 = \frac{dr^2}{1 - kr^2} + r^2 \underbrace{(d\theta^2 + \sin^2 \theta d\phi^2)}_{\equiv d\Omega^2}, \quad \text{where } k \begin{cases} = 0 & E^3 \\ > 0 & S^3 \\ < 0 & H^3 \end{cases}. \quad (1.1.7)$$

*Derivation**.—Let us determine the metric for each case:

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

$$d\ell^2 = d\mathbf{x}^2 = \delta_{ij} dx^i dx^j. \quad (1.1.8)$$

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

- *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 ,

$$d\ell^2 = d\mathbf{x}^2 + du^2, \quad \mathbf{x}^2 + u^2 = a^2, \quad (1.1.9)$$

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 $\mathbb{R}^{1,3}$,

$$d\ell^2 = d\mathbf{x}^2 - du^2, \quad \mathbf{x}^2 - u^2 = -a^2, \quad (1.1.10)$$

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, $\mathbf{x} \rightarrow a\mathbf{x}$ and $u \rightarrow au$. The line elements of the spherical and hyperbolic cases then are

$$d\ell^2 = a^2 [d\mathbf{x}^2 \pm du^2], \quad \mathbf{x}^2 \pm u^2 = \pm 1. \quad (1.1.11)$$

7 1. Geometry and Dynamics

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

$$d\ell^2 = a^2 \left[d\mathbf{x}^2 \pm \frac{(\mathbf{x} \cdot d\mathbf{x})^2}{1 \mp \mathbf{x}^2} \right]. \quad (1.1.12)$$

We can unify (1.1.12) with the Euclidean line element (1.1.8) by writing

$$d\ell^2 = a^2 \left[d\mathbf{x}^2 + k \frac{(\mathbf{x} \cdot d\mathbf{x})^2}{1 - k\mathbf{x}^2} \right], \quad \text{for } k \equiv \begin{cases} 0 & E^3 \\ +1 & S^3 \\ -1 & H^3 \end{cases}. \quad (1.1.13)$$

Note that we must take $a^2 > 0$ in order to have $d\ell^2$ positive at $\mathbf{x} = 0$, and hence everywhere. It is convenient to use spherical polar coordinates, (r, θ, ϕ) , because it makes the symmetries of the space manifest. Using

$$d\mathbf{x}^2 = dr^2 + r^2(d\theta^2 + \sin^2 \theta d\phi^2), \quad (1.1.14)$$

$$\mathbf{x} \cdot d\mathbf{x} = r dr, \quad (1.1.15)$$

the metric in (1.1.13) becomes diagonal

$$d\ell^2 = a^2 \left[\frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right], \quad (1.1.16)$$

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

Note that while I kept $k = 0, \pm 1$ and r dimensionless during this derivation (which makes the discrete three choices for the symmetric 3-space manifest), I adopt the convention that $a(t)$ is dimensionless throughout most of the lecture. Both conventions are connected via the rescaling symmetry in Eq. (1.1.18) below.

Exercise.—Show that despite appearance $r = 0$ is not a special point in (1.1.7).

1.1.3 Robertson-Walker Metric

Substituting (1.1.7) into (1.1.6), we obtain the *Robertson-Walker metric*³ in polar coordinates:

$$ds^2 = dt^2 - a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right]. \quad (1.1.17)$$

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 line element (1.1.17) has a rescaling symmetry

$$a \rightarrow \lambda a, \quad r \rightarrow r/\lambda, \quad k \rightarrow \lambda^2 k. \quad (1.1.18)$$

This means that the geometry of the spacetime stays the same if we simultaneously rescale a , r and k as in (1.1.18). We can use this freedom to set the scale factor to unity today:⁴

³Sometimes this is called the Friedmann-Robertson-Walker (FRW) metric.

⁴Quantities that are evaluated at the present time t_0 will have a subscript ‘0’.

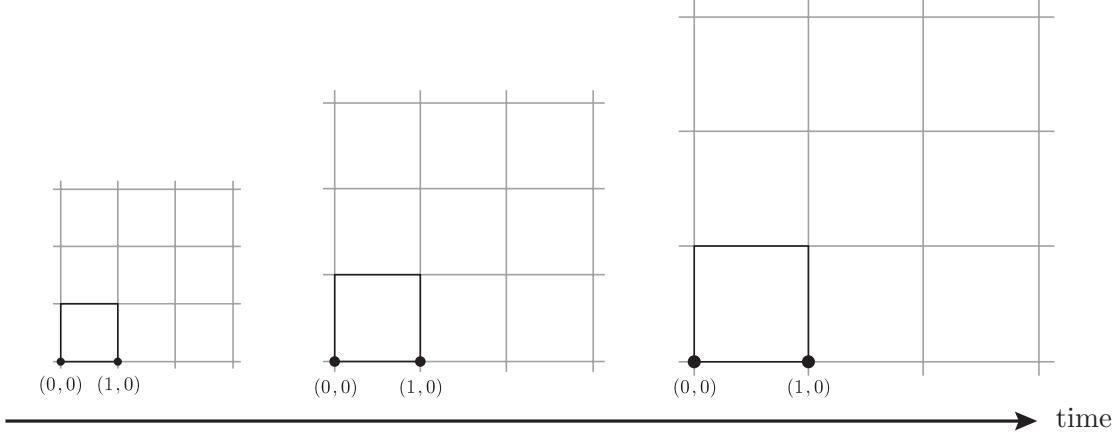


Figure 1.3: Expansion of the universe. The comoving distance between points on an imaginary coordinate grid remains constant as the universe expands. The physical distance is proportional to the comoving distance times the scale factor $a(t)$ and hence gets larger as time evolves.

$a(t_0) \equiv 1$. In this case, $a(t)$ becomes dimensionless, and r and $k^{-1/2}$ inherit the dimension of length.

- The complicated g_{rr} component of (1.1.17) can be inconvenient. In that case, we may define the *comoving distance*, $d\chi \equiv dr/\sqrt{1 - kr^2}$, such that

$$ds^2 = dt^2 - a^2(t) [d\chi^2 + S_k^2(\chi) d\Omega^2], \quad (1.1.19)$$

where

$$S_k(\chi) \equiv \frac{1}{\sqrt{k}} \begin{cases} \sinh(\sqrt{k}\chi) & k < 0 \\ \sqrt{k}\chi & k = 0 \\ \sin(\sqrt{k}\chi) & k > 0 \end{cases}. \quad (1.1.20)$$

- Per definition, χ is a *comoving* coordinate. Physical results depend only on *proper distances* $d_\chi = a(t)\chi$ (see Fig. 1.3). The physical velocity of an object is

$$v_{\text{phys}} \equiv \frac{dd_\chi}{dt} = a \frac{d\chi}{dt} + \frac{da}{dt} \chi \equiv v_{\text{pec}} + H d_\chi. \quad (1.1.21)$$

We see that this has two contributions: the so-called *peculiar velocity*, $v_{\text{pec}} \equiv a(t)\dot{\chi}$, and the *Hubble flow*, $H d_\chi$, where we have defined the *Hubble parameter* as⁵

$$H \equiv \frac{\dot{a}}{a}. \quad (1.1.22)$$

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

- It is also often useful to introduce *conformal time*,

$$d\eta = \frac{dt}{a(t)}, \quad (1.1.23)$$

so that (1.1.19) becomes

$$ds^2 = a^2(\eta) [d\eta^2 - (d\chi^2 + S_k^2(\chi)d\Omega^2)]. \quad (1.1.24)$$

⁵Here, and in the following, an overdot denotes a time derivative, i.e. $\dot{a} \equiv da/dt$.

9 1. Geometry and Dynamics

We see that the metric has factorized into a static metric multiplied by a time-dependent conformal factor $a(\eta)$. This form of the metric is particularly convenient for studying the propagation of light.

1.2 Kinematics

1.2.1 Geodesics

In this section, we will study how particles evolve in the FRW spacetime. Let us first, however, look at the simpler problem of the Newtonian dynamics of a free particle. In Cartesian coordinates, we would simply have

$$\frac{d^2x^i}{dt^2} = 0. \quad (1.2.25)$$

We want to know what this equation turns into for an arbitrary coordinate system in which the three-dimensional metric is $g_{ij} \neq \delta_{ij}$. To derive the equation of motion in these coordinates, we start from the Lagrangian of the free particle

$$L = \frac{m}{2} g_{ij}(x^k) \dot{x}^i \dot{x}^j. \quad (1.2.26)$$

Substituting this into the Euler-Lagrange equation (see below), we find

$$\boxed{\frac{d^2x^i}{dt^2} = -\Gamma_{ab}^i \frac{dx^a}{dt} \frac{dx^b}{dt}}, \quad (1.2.27)$$

where we have introduced the *Christoffel symbol*

$$\Gamma_{ab}^i \equiv \frac{1}{2} g^{ij} (\partial_a g_{jb} + \partial_b g_{aj} - \partial_j g_{ab}), \quad \text{with } \partial_j \equiv \partial/\partial x^j. \quad (1.2.28)$$

Derivation.*—The Euler-Lagrange equation is

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}^k} \right) = \frac{\partial L}{\partial x^k}. \quad (1.2.29)$$

Substituting (1.2.26) into the derivatives, we have

$$\frac{\partial L}{\partial x^k} = \frac{1}{2} \partial_k g_{ij} \dot{x}^i \dot{x}^j, \quad (1.2.30)$$

$$\frac{\partial L}{\partial \dot{x}^k} = g_{ik} \dot{x}^i, \quad (1.2.31)$$

where we have set $m \equiv 1$ since it will cancel on both sides. The l.h.s. of (1.2.29) then becomes

$$\begin{aligned} \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}^k} \right) &= \frac{d}{dt} (g_{ik} \dot{x}^i) = g_{ik} \ddot{x}^i + \frac{dx^j}{dt} \frac{\partial g_{ik}}{\partial x^j} \dot{x}^i \\ &= g_{ik} \ddot{x}^i + \partial_j g_{ik} \dot{x}^i \dot{x}^j \\ &= g_{ik} \ddot{x}^i + \frac{1}{2} (\partial_i g_{jk} + \partial_j g_{ik}) \dot{x}^i \dot{x}^j. \end{aligned} \quad (1.2.32)$$

Equation (1.2.29) then implies

$$g_{ki} \ddot{x}^i = -\frac{1}{2} (\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}) \dot{x}^i \dot{x}^j. \quad (1.2.33)$$

Multiplying both sides by g^{lk} , we get

$$\ddot{x}^l = -\frac{1}{2} g^{lk} (\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}) \dot{x}^i \dot{x}^j \equiv -\Gamma_{ij}^l \dot{x}^i \dot{x}^j, \quad (1.2.34)$$

which is the desired result (1.2.27).

10 1. Geometry and Dynamics

The equation of motion of a massive particle in general relativity will take a similar form as eq. (1.2.27). However, in this case, the term involving the Christoffel symbol cannot be removed by going to Cartesian coordinates, but is a physical manifestation of the curvature of the spacetime.

Geodesic Equation

In the absence of additional non-gravitational forces, freely-falling particles in a curved spacetime move along special trajectories called *geodesics*. For massive particles, a geodesic is the timelike curve $X^\mu(\tau)$ which extremises the proper time $\Delta\tau$ between two points in the spacetime. In Appendix A, I show that this extremal path satisfies the *geodesic equation*

$$\boxed{\frac{d^2X^\mu}{d\tau^2} = -\Gamma_{\alpha\beta}^\mu \frac{dX^\alpha}{d\tau} \frac{dX^\beta}{d\tau}}, \quad (1.2.35)$$

where

$$\Gamma_{\alpha\beta}^\mu \equiv \frac{1}{2}g^{\mu\lambda}(\partial_\alpha g_{\beta\lambda} + \partial_\beta g_{\alpha\lambda} - \partial_\lambda g_{\alpha\beta}). \quad (1.2.36)$$

Notice the similarity between eqs. (1.2.35) and (1.2.27). It will be convenient to write the geodesic equation in a few different ways:

Introducing the *four-velocity* of the particle, $U^\mu \equiv dX^\mu/d\tau$, we get

$$\frac{dU^\mu}{d\tau} = -\Gamma_{\alpha\beta}^\mu U^\alpha U^\beta. \quad (1.2.37)$$

For massive particles, we can furthermore define the *four-momentum* $P^\mu \equiv mU^\mu$, and write

$$m\frac{dP^\mu}{d\tau} = -\Gamma_{\alpha\beta}^\mu P^\alpha P^\beta. \quad (1.2.38)$$

Using the above definitions, and $X^0 = t$ and $P^0 = E$, we can then replace the derivative w.r.t. proper time τ by a derivative w.r.t. coordinate time t ,

$$\frac{d}{d\tau} = \frac{dX^0}{d\tau} \frac{d}{dX^0} = \frac{E}{m} \frac{d}{dt}. \quad (1.2.39)$$

With this, the geodesic equation becomes

$$\boxed{E\dot{P}^\mu = -\Gamma_{\alpha\beta}^\mu P^\alpha P^\beta}, \quad (1.2.40)$$

which also holds for *massless* particles.

I will now show you how to apply the geodesic equation (1.2.40) to particles in the FRW universe.

Particles in the Expanding Universe

To evaluate the r.h.s. of (1.2.40) we need to compute the Christoffel symbols for the FRW metric,

$$ds^2 = dt^2 - a^2(t)\gamma_{ij}dx^i dx^j. \quad (1.2.41)$$

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

$$\boxed{\Gamma_{ij}^0 = a\dot{a}\gamma_{ij}, \quad \Gamma_{0j}^i = \frac{\dot{a}}{a}\delta_j^i, \quad \Gamma_{jk}^i = \frac{1}{2}\gamma^{il}(\partial_j\gamma_{kl} + \partial_k\gamma_{jl} - \partial_l\gamma_{jk})}, \quad (1.2.42)$$

11 1. Geometry and Dynamics

or are related to these by symmetry (note that $\Gamma_{\alpha\beta}^\mu = \Gamma_{\beta\alpha}^\mu$). I will derive Γ_{ij}^0 as an example and leave Γ_{0j}^i as an exercise.

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

$$\Gamma_{\alpha\beta}^0 = \frac{1}{2}g^{0\lambda}(\partial_\alpha g_{\beta\lambda} + \partial_\beta g_{\alpha\lambda} - \partial_\lambda g_{\alpha\beta}). \quad (1.2.43)$$

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

$$\Gamma_{\alpha\beta}^0 = \frac{1}{2}(\partial_\alpha g_{\beta 0} + \partial_\beta g_{\alpha 0} - \partial_0 g_{\alpha\beta}). \quad (1.2.44)$$

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

$$\Gamma_{\alpha\beta}^0 = -\frac{1}{2}\partial_0 g_{\alpha\beta}. \quad (1.2.45)$$

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

$$\Gamma_{ij}^0 = a\dot{a}\gamma_{ij}. \quad (1.2.46)$$

The geodesic equation (1.2.40) can be written as

$$\begin{aligned} P^0 \frac{dP^\mu}{dt} &= -\Gamma_{\alpha\beta}^\mu P^\alpha P^\beta, \\ &= -\left(2\Gamma_{0j}^\mu P^0 + \Gamma_{ij}^\mu P^i\right) P^j, \end{aligned} \quad (1.2.47)$$

where I have used (1.2.42) in the second line.

- The first thing to notice from (1.2.47) is that massive particles at rest in the comoving frame, $P^j = 0$, will stay at rest because the r.h.s. then vanishes,

$$P^j = 0 \quad \Rightarrow \quad \frac{dP^i}{dt} = 0. \quad (1.2.48)$$

- Next, we consider the $\mu = 0$ component of (1.2.47), but don't require the particles to be at rest. The first term on the r.h.s. vanishes because $\Gamma_{0j}^0 = 0$. Using (1.2.42), we then find

$$E \frac{dE}{dt} = -\Gamma_{ij}^0 P^i P^j = -\frac{\dot{a}}{a} p^2, \quad (1.2.49)$$

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

$$p^2 \equiv -g_{ij} P^i P^j = a^2 \gamma_{ij} P^i P^j. \quad (1.2.50)$$

Notice the appearance of the scale factor in (1.2.50) 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 $E dE = pdp$, so that (1.2.49) can be written as

$$\frac{\dot{p}}{p} = -\frac{\dot{a}}{a} \quad \Rightarrow \quad p \propto \frac{1}{a}. \quad (1.2.51)$$

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. (1.2.51) implies

$$p = E \propto \frac{1}{a} \quad (\text{massless particles}), \quad (1.2.52)$$

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

- For massive particles, eq. (1.2.51) implies

$$p = \frac{mv}{\sqrt{1-v^2}} \propto \frac{1}{a} \quad (\text{massive particles}), \quad (1.2.53)$$

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. (1.1.21). To get the first equality in (1.2.53), I have used

$$P^i = mU^i = m \frac{dX^i}{d\tau} = m \frac{dt}{d\tau} v^i = \frac{mv^i}{\sqrt{1-a^2\gamma_{ij}v^iv^j}} = \frac{mv^i}{\sqrt{1-v^2}}. \quad (1.2.54)$$

Equation (1.2.53) 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.

Photons.—In the quantum mechanical description, the wavelength of light is inversely proportional to the photon momentum, $\lambda = h/p$. Since according to (1.2.52) 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

$$\lambda_0 = \frac{a(t_0)}{a(t_1)} \lambda_1. \quad (1.2.55)$$

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

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 η_1 , the galaxy emits a signal of short conformal duration $\Delta\eta$. The light arrives at our telescopes at time $\eta_0 = \eta_1 + d$. 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,

$$\Delta t_1 = a(\eta_1)\Delta\eta \quad \text{and} \quad \Delta t_0 = a(\eta_0)\Delta\eta. \quad (1.2.56)$$

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

$$\frac{\lambda_0}{\lambda_1} = \frac{a(\eta_0)}{a(\eta_1)}. \quad (1.2.57)$$

13 1. Geometry and Dynamics

Redshift.—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,

$$z \equiv \frac{\lambda_0 - \lambda_1}{\lambda_1}. \quad (1.2.58)$$

We then find

$$1 + z = \frac{a(t_0)}{a(t_1)}. \quad (1.2.59)$$

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

$$\boxed{1 + z = \frac{1}{a(t_1)}}. \quad (1.2.60)$$

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

$$a(t_1) = a(t_0)[1 + (t_1 - t_0)H_0 + \dots], \quad (1.2.61)$$

where H_0 is the *Hubble constant*

$$H_0 \equiv \frac{\dot{a}(t_0)}{a(t_0)}. \quad (1.2.62)$$

Equation (1.2.59) then gives $z = H_0(t_0 - t_1) + \dots$. 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

$$z \simeq H_0 d. \quad (1.2.63)$$

The slope in a redshift-distance diagram (cf. fig. 1.6) 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 define⁶

$$H_0 \equiv 100 h \text{ km s}^{-1} \text{Mpc}^{-1}, \quad (1.2.64)$$

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,⁷

$$h \approx 0.67 \pm 0.01. \quad (1.2.65)$$

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 (1.1.19),

$$ds^2 = dt^2 - a^2(t)[d\chi^2 + S_k^2(\chi)d\Omega^2], \quad (1.2.66)$$

where⁸

$$S_k(\chi) \equiv \begin{cases} R_0 \sinh(\chi/R_0) & k = -1 \\ \chi & k = 0 \\ R_0 \sin(\chi/R_0) & k = +1 \end{cases}. \quad (1.2.67)$$

⁶A parsec (pc) is 3.26 light-years. Blame astronomers for the funny units in (6.3.29).

⁷Planck 2015 Results: *Cosmological Parameters* [[arXiv:1502.01589](#)].

⁸Notice 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 $a \rightarrow \lambda a$, $\chi \rightarrow \chi/\lambda$, and $S_k^2 \rightarrow S_k^2/\lambda$.

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

$$d_m = S_k(\chi). \quad (1.2.68)$$

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

$$\chi(z) = \int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_0^z \frac{dz}{H(z)}, \quad (1.2.69)$$

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

$$F = \frac{L}{4\pi\chi^2}. \quad (1.2.70)$$

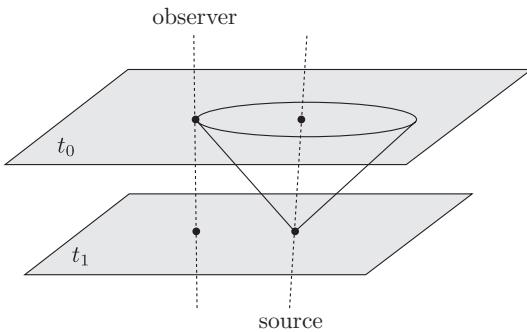


Figure 1.4: Geometry associated with the definition of luminosity distance.

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

$$F = \frac{L}{4\pi d_m^2(1+z)^2} \equiv \frac{L}{4\pi d_L^2}, \quad (1.2.71)$$

where we have defined the *luminosity distance*, d_L , so that the relation between luminosity, flux and luminosity distance is the same as in (1.2.70). Hence, we find

$$d_L = d_m(1 + z). \quad (1.2.72)$$

- *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,⁹

$$d_A = \frac{D}{\delta\theta}. \quad (1.2.73)$$

This quantity is called the *angular diameter distance*. The FRW metric (1.1.24) implies

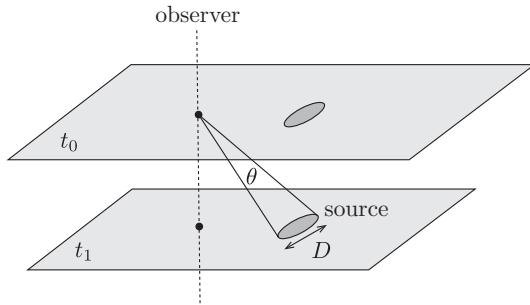


Figure 1.5: Geometry associated with the definition of angular diameter distance.

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

$$D = a(t_1)S_k(\chi)\delta\theta = \frac{d_m}{1 + z}\delta\theta. \quad (1.2.74)$$

Hence, we get

$$d_A = \frac{d_m}{1 + z}. \quad (1.2.75)$$

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

$$d_A = \frac{d_L}{(1 + z)^2}. \quad (1.2.76)$$

Figure 1.6 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. 1.7 in §1.3.3).

⁹This formula assumes $\delta\theta \ll 1$ (in radians) which is true for all cosmological objects.

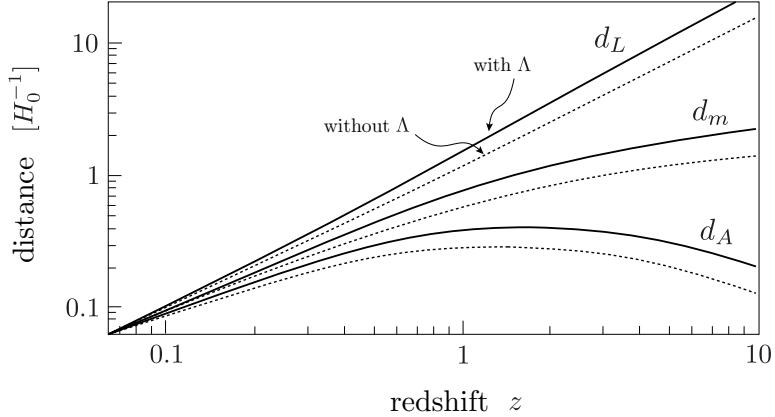


Figure 1.6: Distance measures in a flat universe, with matter only (dotted lines) and with 70% dark energy (solid lines). In a dark energy dominated universe, distances out to a fixed redshift are larger than in a matter-dominated universe.

1.3 Dynamics

The dynamics of the universe is determined by the Einstein equation

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}. \quad (1.3.77)$$

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 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*,

$$T_{\mu\nu} = (\rho + P) U_\mu U_\nu - P g_{\mu\nu}, \quad (1.3.78)$$

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-scalar¹⁰, 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

$$N^0 = n(t), \quad N^i = 0, \quad (1.3.79)$$

¹⁰A 3-scalar is a quantity that is invariant under purely spatial coordinate transformations.

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

$$N^\mu = nU^\mu, \quad (1.3.80)$$

where $U^\mu \equiv dX^\mu/d\tau$ is the relative four-velocity between the particles and the observer. Of course, we recover the previous result (1.3.79) for a comoving observer, $U^\mu = (1, 0, 0, 0)$. For $U^\mu = \gamma(1, v^i)$, eq. (1.3.80) 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

$$\dot{N}^0 = -\partial_i N^i, \quad (1.3.81)$$

or, in relativistic notation,

$$\partial_\mu N^\mu = 0. \quad (1.3.82)$$

Equation (1.3.82) is generalised to curved spacetimes by replacing the partial derivative ∂_μ with a *covariant derivative* ∇_μ ,¹¹

$$\nabla_\mu N^\mu = \partial_\mu N^\mu + \Gamma_{\mu\lambda}^\mu N^\lambda = 0. \quad (1.3.83)$$

Using (1.3.79), this becomes

$$\frac{dn}{dt} + \Gamma_{i0}^i n = 0, \quad (1.3.84)$$

and substituting (1.2.42), we find

$$\frac{\dot{n}}{n} = -3\frac{\dot{a}}{a} \Rightarrow n(t) \propto a^{-3}. \quad (1.3.85)$$

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 $\mathbf{x} = 0$ 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 $\mathbf{x} = 0$,

$$T_{ij}(\mathbf{x} = 0) \propto \delta_{ij} \propto g_{ij}(\mathbf{x} = 0). \quad (1.3.86)$$

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

¹¹The 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.

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

$$T_{00} = \rho(t), \quad \pi_i \equiv T_{i0} = 0, \quad T_{ij} = -P(t)g_{ij}(t, \mathbf{x}). \quad (1.3.87)$$

It looks even nicer with mixed upper and lower indices

$$T^{\mu}_{\nu} = g^{\mu\lambda}T_{\lambda\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & -P & 0 & 0 \\ 0 & 0 & -P & 0 \\ 0 & 0 & 0 & -P \end{pmatrix}. \quad (1.3.88)$$

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

$$T^{\mu}_{\nu} = (\rho + P)U^{\mu}U_{\nu} - P\delta^{\mu}_{\nu}, \quad (1.3.89)$$

where $U^{\mu} \equiv dX^{\mu}/d\tau$ 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 (1.3.88) 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, $\dot{\pi}_i = \partial_iP$. These conservation laws can be combined into a four-component conservation equation for the stress-energy tensor

$$\partial_{\mu}T^{\mu}_{\nu} = 0. \quad (1.3.90)$$

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

$$\nabla_{\mu}T^{\mu}_{\nu} = \partial_{\mu}T^{\mu}_{\nu} + \Gamma^{\mu}_{\mu\lambda}T^{\lambda}_{\nu} - \Gamma^{\lambda}_{\mu\nu}T^{\mu}_{\lambda} = 0. \quad (1.3.91)$$

This corresponds to four separate equations (one for each ν). The evolution of the energy density is determined by the $\nu = 0$ equation

$$\partial_{\mu}T^{\mu}_{0} + \Gamma^{\mu}_{\mu\lambda}T^{\lambda}_{0} - \Gamma^{\lambda}_{\mu 0}T^{\mu}_{\lambda} = 0. \quad (1.3.92)$$

Since T^i_0 vanishes by isotropy, this reduces to

$$\frac{d\rho}{dt} + \Gamma^{\mu}_{\mu 0}\rho - \Gamma^{\lambda}_{\mu 0}T^{\mu}_{\lambda} = 0. \quad (1.3.93)$$

From eq. (1.2.42) 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 (1.3.93) therefore reads

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

(1.3.94)

Exercise.—Show that (1.3.94) can be written as, $dU = -PdV$, 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 (1.3.94) gives

$$\rho \propto a^{-3}. \quad (1.3.95)$$

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.¹²

- **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. (1.3.94) implies

$$\rho \propto a^{-4}. \quad (1.3.96)$$

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. (1.3.94), we find that the energy *density* is constant,

$$\rho \propto a^0. \quad (1.3.97)$$

¹²Of 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”.

Since the energy density doesn't dilute, energy has to be created as the universe expands.¹³

- *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

$$T_{\mu\nu}^{\text{vac}} = \rho_{\text{vac}} g_{\mu\nu}. \quad (1.3.98)$$

Comparison with eq. (1.3.89), show that this indeed implies $P_{\text{vac}} = -\rho_{\text{vac}}$. Unfortunately, the predicted size of ρ_{vac} is completely off,

$$\frac{\rho_{\text{vac}}}{\rho_{\text{obs}}} \sim 10^{120}. \quad (1.3.99)$$

This so-called “cosmological constant problem” is the biggest crisis in modern theoretical physics.

- *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.

Cosmological constant.—The left-hand side of the Einstein equation (1.3.77) 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

$$G_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}. \quad (1.3.100)$$

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

$$T_{\mu\nu}^{(\Lambda)} = \frac{\Lambda}{8\pi G} g_{\mu\nu} \equiv \rho_\Lambda g_{\mu\nu}. \quad (1.3.101)$$

This is of the same form as the stress-energy tensor from vacuum energy, eq. (1.3.98).

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 (1.3.94) scale as

$$\rho \propto a^{-3(1+w)} = \begin{cases} a^{-3} & \text{matter} \\ a^{-4} & \text{radiation} \\ a^0 & \text{vacuum} \end{cases}. \quad (1.3.102)$$

1.3.2 Spacetime Curvature

We want to relate these matter sources to the evolution of the scale factor in the FRW metric. To do this we have to compute the Einstein tensor on the l.h.s. of the Einstein equation (1.3.77),

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}. \quad (1.3.103)$$

¹³In a gravitational system this doesn't have to violate the conservation of energy. It is the conservation equation (1.3.94) that counts.

21 1. Geometry and Dynamics

We will need the Ricci tensor

$$R_{\mu\nu} \equiv \partial_\lambda \Gamma_{\mu\nu}^\lambda - \partial_\nu \Gamma_{\mu\lambda}^\lambda + \Gamma_{\lambda\rho}^\lambda \Gamma_{\mu\nu}^\rho - \Gamma_{\mu\lambda}^\rho \Gamma_{\nu\rho}^\lambda, \quad (1.3.104)$$

and the Ricci scalar

$$R = R^\mu_\mu = g^{\mu\nu} R_{\mu\nu}. \quad (1.3.105)$$

Again, there is a lot of beautiful geometry behind these definitions. We will simply keep plugging-and-playing: given the Christoffel symbols (1.2.42) nothing stops us from computing (1.3.104).

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

$$R_{00} = -3 \frac{\ddot{a}}{a}, \quad (1.3.106)$$

$$R_{ij} = - \left[\frac{\ddot{a}}{a} + 2 \left(\frac{\dot{a}}{a} \right)^2 + 2 \frac{k}{a^2} \right] g_{ij}. \quad (1.3.107)$$

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 (1.3.104), we have

$$R_{00} = \partial_\lambda \Gamma_{00}^\lambda - \partial_0 \Gamma_{0\lambda}^\lambda + \Gamma_{\lambda\rho}^\lambda \Gamma_{00}^\rho - \Gamma_{0\lambda}^\rho \Gamma_{0\rho}^\lambda, \quad (1.3.108)$$

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

$$R_{00} = -\partial_0 \Gamma_{0i}^i - \Gamma_{0j}^i \Gamma_{0i}^j. \quad (1.3.109)$$

Using $\Gamma_{0j}^i = (\dot{a}/a)\delta_j^i$, we find

$$R_{00} = -\frac{d}{dt} \left(3 \frac{\dot{a}}{a} \right) - 3 \left(\frac{\dot{a}}{a} \right)^2 = -3 \frac{\ddot{a}}{a}. \quad (1.3.110)$$

The Ricci scalar is

$$\begin{aligned} R &= g^{\mu\nu} R_{\mu\nu} \\ &= R_{00} - \frac{1}{a^2} R_{ii} = -6 \left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} \right]. \end{aligned} \quad (1.3.111)$$

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

$$G^0_0 = 3 \left[\left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} \right], \quad (1.3.112)$$

$$G^i_j = \left[2 \frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} \right] \delta_j^i. \quad (1.3.113)$$

Exercise.—Verify eqs. (1.3.112) and (1.3.113).

1.3.3 Friedmann Equations

Combining eqs. (1.3.112) and (1.3.113) with the stress-tensor (1.3.88), we get the *Friedmann equations*,

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3} \rho - \frac{k}{a^2}, \quad (1.3.114)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P), \quad (1.3.115)$$

where 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, $H \equiv \dot{a}/a$,

$$H^2 = \frac{8\pi G}{3} \rho - \frac{k}{a^2}. \quad (1.3.116)$$

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

$$\begin{aligned} \rho_{\text{crit},0} &= \frac{3H_0^2}{8\pi G} = 1.9 \times 10^{-29} h^2 \text{ grams cm}^{-3} \\ &= 2.8 \times 10^{11} h^2 M_\odot \text{ Mpc}^{-3} \\ &= 1.1 \times 10^{-5} h^2 \text{ protons cm}^{-3}. \end{aligned} \quad (1.3.117)$$

We use the critical density to define dimensionless density parameters

$$\Omega_{a,0} \equiv \frac{\rho_{a,0}}{\rho_{\text{crit},0}}, \quad a = r, m, \Lambda, \dots \quad (1.3.118)$$

The Friedmann equation (1.3.116) can then be written as

$$H^2(a) = H_0^2 \left[\Omega_{r,0} \left(\frac{a_0}{a} \right)^4 + \Omega_{m,0} \left(\frac{a_0}{a} \right)^3 + \Omega_{k,0} \left(\frac{a_0}{a} \right)^2 + \Omega_{\Lambda,0} \right], \quad (1.3.119)$$

where we have defined a “curvature” density parameter, $\Omega_{k,0} \equiv -k/(a_0 H_0)^2$. 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, $a_0 \equiv 1$. Equation (1.3.119) then becomes

$$\frac{H^2}{H_0^2} = \Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda. \quad (1.3.120)$$

Λ CDM

Observations (see Figs. 1.7 and 1.8) show that the universe is filled with radiation (‘ r ’), matter (‘ m ’) and dark energy (‘ Λ ’):

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

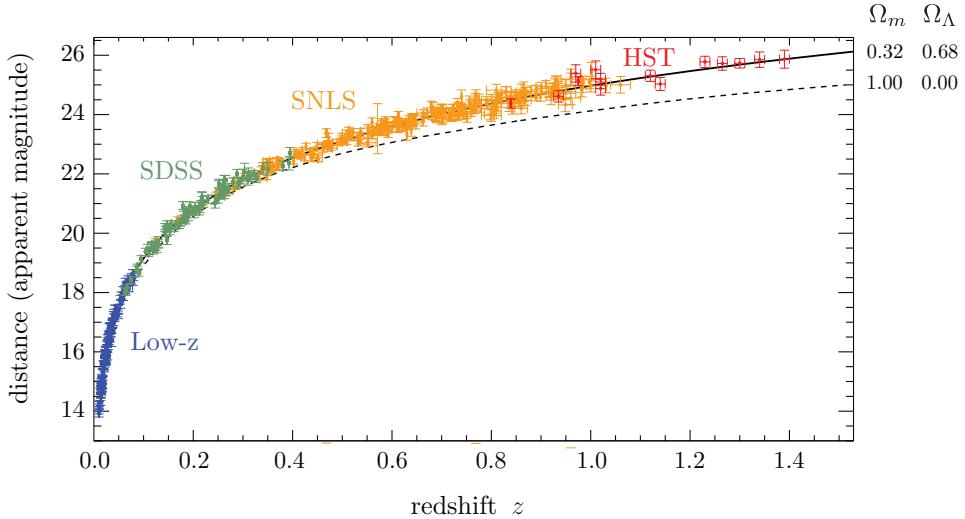


Figure 1.7: Type IA supernovae and the discovery dark energy. If we assume a flat universe, then the supernovae clearly appear fainter (or more distant) than predicted in a matter-only universe ($\Omega_m = 1.0$). (SDSS = Sloan Digital Sky Survey; SNLS = SuperNova Legacy Survey; HST = Hubble Space Telescope.)

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.

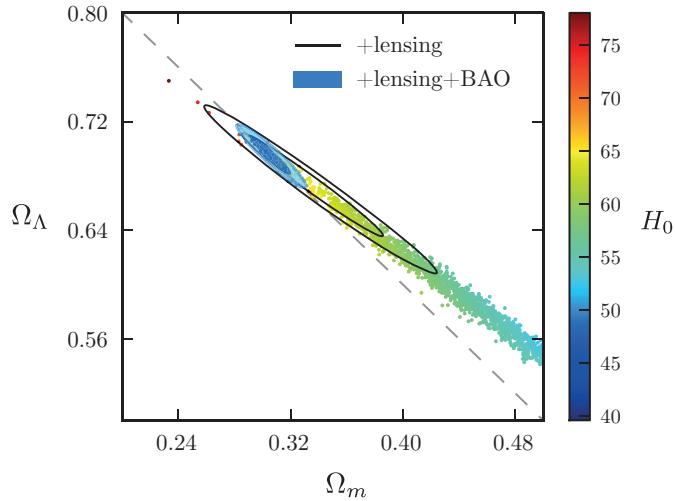


Figure 1.8: A combination CMB and LSS observations indicate that the spatial geometry of the universe is flat. The energy density of the universe is dominated by a cosmological constant. Notice that the CMB data alone cannot exclude a matter-only universe with large spatial curvature. The evidence for 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. 1.9). Parameterising this component by its equation of state w_a captures all cases of interest. For a flat, single-component universe, the Friedmann equation (1.3.120) reduces to

$$\frac{\dot{a}}{a} = H_0 \sqrt{\Omega_a} a^{-\frac{3}{2}(1+w_a)}. \quad (1.3.121)$$

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

$$a(t) \propto \begin{cases} t^{2/3(1+w_a)} & w_a \neq -1 \\ e^{Ht} & w_a = -1 \end{cases} \quad \begin{matrix} t^{2/3} & \text{MD} \\ t^{1/2} & \text{RD} \\ & \Lambda\text{D} \end{matrix} \quad (1.3.122)$$

or, in conformal time,

$$a(\eta) \propto \begin{cases} \eta^{2/(1+3w_a)} & w_a \neq -1 \\ (-\eta)^{-1} & w_a = -1 \end{cases} \quad \begin{matrix} \eta^2 & \text{MD} \\ \eta & \text{RD} \\ & \Lambda\text{D} \end{matrix} \quad (1.3.123)$$

Exercise.—Derive eq. (1.3.123) from eq. (1.3.122).

Two-Component Universe*

Matter and radiation were equally important at $a_{\text{eq}} \equiv \Omega_r/\Omega_m \approx 3 \times 10^{-4}$, which was shortly before the cosmic microwave background was released (in §3.3.3, we will show that this happened at $a_{\text{rec}} \approx 9 \times 10^{-4}$). It will be useful to have an exact solution describing the transition era. Let

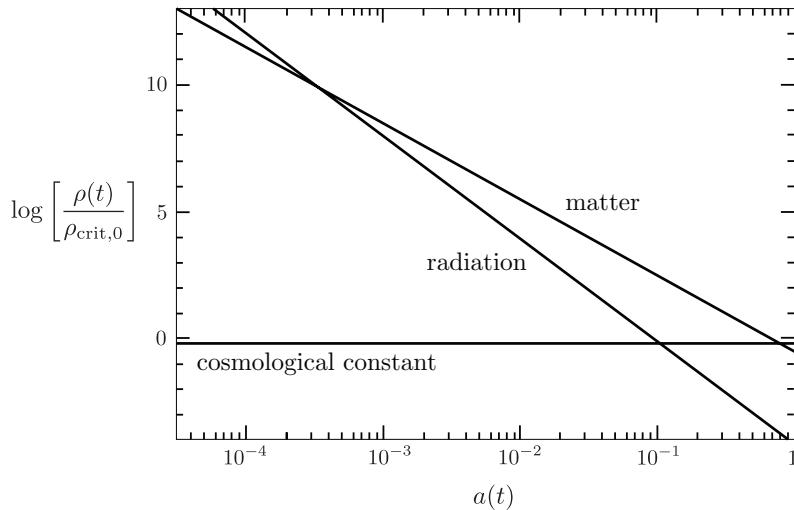


Figure 1.9: Evolution of the energy densities in the universe.

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 (1.3.114) and (1.3.115) then are

$$(a')^2 = \frac{8\pi G}{3}\rho a^4, \quad (1.3.124)$$

$$a'' = \frac{4\pi G}{3}(\rho - 3P)a^3, \quad (1.3.125)$$

where primes denote derivatives with respect to conformal time and

$$\rho \equiv \rho_m + \rho_r = \frac{\rho_{\text{eq}}}{2} \left[\left(\frac{a_{\text{eq}}}{a} \right)^3 + \left(\frac{a_{\text{eq}}}{a} \right)^4 \right]. \quad (1.3.126)$$

Exercise.—Derive eqs. (1.3.124) and (1.3.125). 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. (1.3.125), $\rho_r - 3P_r = 0$. Moreover, since $\rho_m a^3 = \text{const.} = \frac{1}{2}\rho_{\text{eq}} a_{\text{eq}}^3$, we can write eq. (1.3.125) as

$$a'' = \frac{2\pi G}{3}\rho_{\text{eq}} a_{\text{eq}}^3. \quad (1.3.127)$$

This equation has the following solution

$$a(\eta) = \frac{\pi G}{3}\rho_{\text{eq}} a_{\text{eq}}^3 \eta^2 + C\eta + D. \quad (1.3.128)$$

Imposing $a(\eta = 0) \equiv 0$, fixes one integration constant, $D = 0$. We find the second integration constant by substituting (1.3.128) and (1.3.126) into (1.3.124),

$$C = \left(\frac{4\pi G}{3}\rho_{\text{eq}} a_{\text{eq}}^4 \right)^{1/2}. \quad (1.3.129)$$

Eq. (1.3.128) can then be written as

$$a(\eta) = a_{\text{eq}} \left[\left(\frac{\eta}{\eta_*} \right)^2 + 2 \left(\frac{\eta}{\eta_*} \right) \right], \quad (1.3.130)$$

where

$$\eta_* \equiv \left(\frac{\pi G}{3}\rho_{\text{eq}} a_{\text{eq}}^2 \right)^{-1/2} = \frac{\eta_{\text{eq}}}{\sqrt{2} - 1}. \quad (1.3.131)$$

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

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 G for the (reduced) Planck mass,

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

so that the Friedmann equation (1.3.116) is written as $H^2 = \rho/(3M_{\text{pl}}^2)$.

2.1 The Horizon Problem

2.1.1 Particle Horizon

The size of a causally connected patch of space is 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. $\theta = \phi = \text{const.}$). The evolution is then determined by a two-dimensional line element¹

$$ds^2 = a^2(\eta) [d\eta^2 - d\chi^2]. \quad (2.1.1)$$

Since photons travel along null geodesics, $ds^2 = 0$, their path is defined by

$$\Delta\chi(\eta) = \pm\Delta\eta, \quad (2.1.2)$$

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 - η coordinates. If instead we had used physical time t , then the light cones for curved spacetimes would be curved.

Equation (2.1.2) tells us that the maximal comoving distance that light can travel between two times η_1 and $\eta_2 > \eta_1$ is simply $\Delta\eta = \eta_2 - \eta_1$ (recall that $c \equiv 1$). Hence, if the Big Bang

¹For the radial coordinate χ we have used the parameterisation of (1.1.24), 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.

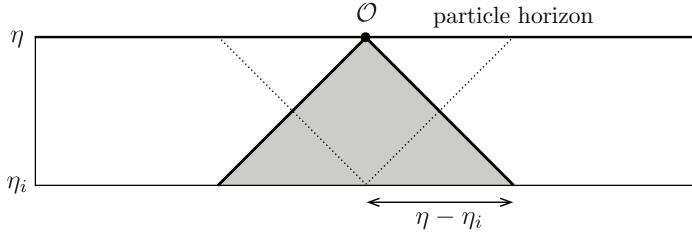


Figure 2.1: Spacetime diagram illustrating the concept of the particle horizon, i.e. the maximal distance from which we can receive signals.

‘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

$$\chi_p(\eta) = \eta - \eta_i = \int_{t_i}^t \frac{dt}{a(t)}. \quad (2.1.3)$$

This is called the (comoving) *particle horizon*. The size of the particle horizon at time η may be visualised by the intersection of the past light cone of an observer \mathcal{O} with the spacelike surface $\eta = \eta_i$ (see Fig. 2.1). Causal influences have to come from within this region.

2.1.2 Hubble Radius

Equation (2.1.3) can be written in the following illuminating way

$$\chi_p(\eta) = \int_{t_i}^t \frac{dt}{a} = \int_{a_i}^a \frac{da}{a\dot{a}} = \int_{\ln a_i}^{\ln a} \chi_H(a) d\ln a, \quad (2.1.4)$$

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

$$\chi_H = H_0^{-1} a^{\frac{1}{2}(1+3w)}. \quad (2.1.5)$$

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 (2.1.4) 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 (2.1.5) in (2.1.4), we find

$$\chi_p(a) = \frac{2H_0^{-1}}{(1+3w)} \left[a^{\frac{1}{2}(1+3w)} - a_i^{\frac{1}{2}(1+3w)} \right] \equiv \eta - \eta_i. \quad (2.1.6)$$

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

$$\eta_i \equiv \frac{2H_0^{-1}}{(1+3w)} a_i^{\frac{1}{2}(1+3w)} \xrightarrow{a_i \rightarrow 0, w > -\frac{1}{3}} 0. \quad (2.1.7)$$

The comoving horizon is finite,

$$\chi_p(t) = \frac{2H_0^{-1}}{(1+3w)} a(t)^{\frac{1}{2}(1+3w)} = \frac{2}{(1+3w)} \chi_H(t). \quad (2.1.8)$$

We see that in the standard cosmology $\chi_p \sim \chi_H$. This has led to the confusing practice of referring to both the particle horizon and the Hubble radius as the “horizon”.

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

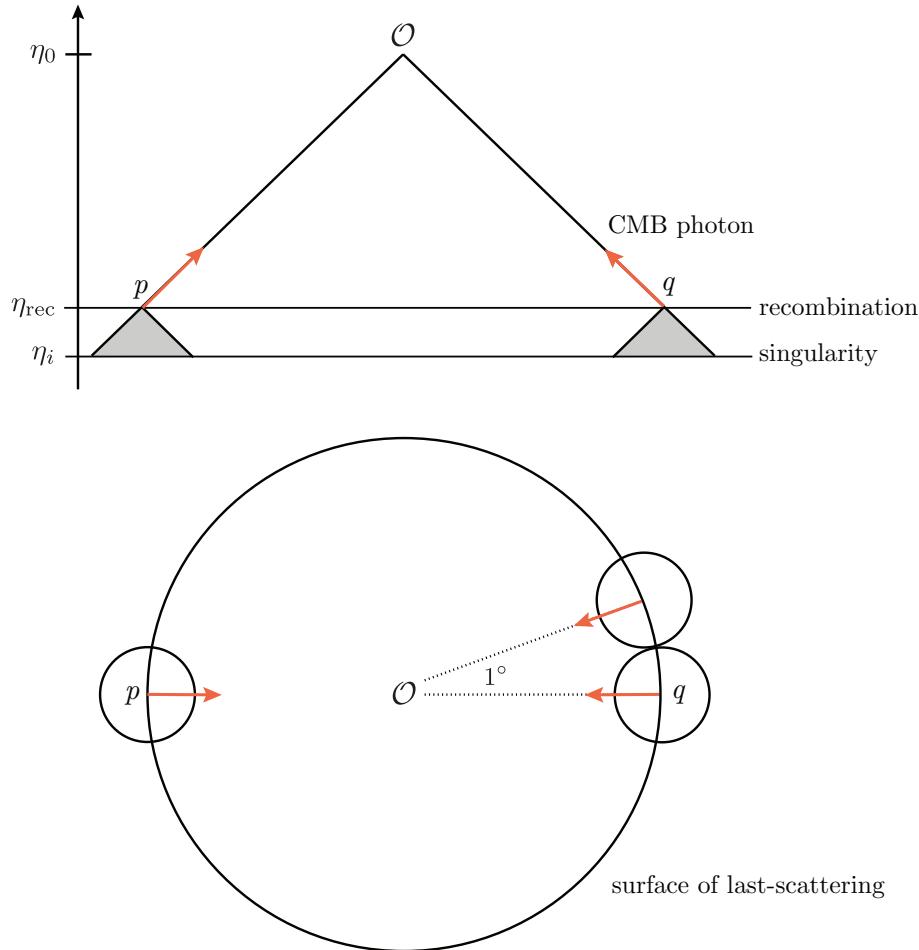


Figure 2.2: The horizon problem in the conventional Big Bang model. All events that we currently observe are on our past light cone. The intersection of our past light cone with the spacelike slice labelled “recombination” corresponds to two opposite points in the observed CMB. Their past light cones don’t overlap before they hit the singularity, $a_i = 0$, 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.

2.1.3 Why is the CMB so uniform?

About 380 000 years after the Big Bang, the universe had cooled enough to allow the formation of the first hydrogen atoms. In this process, photons decoupled 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 parts of 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 . This leads to the following puzzle: 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*.

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,

$$\frac{d}{dt}\chi_H = \frac{d}{dt}(aH)^{-1} < 0. \quad (2.2.9)$$

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

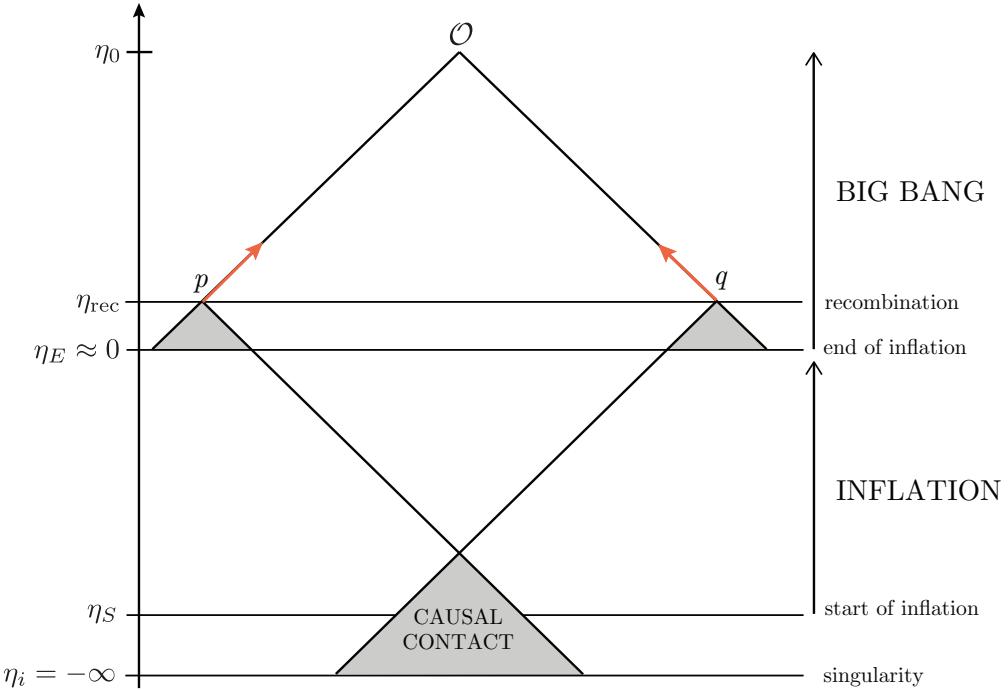


Figure 2.3: Inflationary solution to the horizon problem. Conformal time during inflation is negative. The spacelike singularity of the standard Big Bang is replaced by the reheating surface, i.e. rather than marking the beginning of time it now corresponds simply to the transition from inflation to 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.

2.2.1 Solution of the Horizon Problem

For a shrinking Hubble sphere, the integral in (2.1.4) is dominated by the lower limit. The Big Bang singularity is now pushed to *negative conformal time*,

$$\eta_i = \frac{2H_0^{-1}}{(1+3w)} a_i^{\frac{1}{2}(1+3w)} \xrightarrow{a_i \rightarrow 0, w < -\frac{1}{3}} -\infty. \quad (2.2.10)$$

This implies that there was “much more conformal time between the singularity and decoupling than we had thought”! Figure 2.3 shows the new spacetime diagram. If $|\eta_S| > e^{60}|\eta_E|$, then the past light cones of widely separated points in the CMB now had enough time to intersect. The uniformity of the CMB is not a mystery anymore. In inflationary cosmology, $\eta = 0$ 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 $\eta = 0$.

2.2.2 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 a key feature of 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

$$\frac{d}{dt}(aH)^{-1} = \frac{d}{dt}(\dot{a})^{-1} = -\frac{\ddot{a}}{(\dot{a})^2}, \quad (2.2.11)$$

we see that a shrinking comoving Hubble radius implies accelerated expansion

$$\ddot{a} > 0. \quad (2.2.12)$$

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

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

$$\frac{d\chi_H}{dt} = -\frac{\dot{a}H + a\dot{H}}{(aH)^2} = -\frac{1}{a}(1 - \varepsilon), \quad \text{where } \varepsilon \equiv -\frac{\dot{H}}{H^2}. \quad (2.2.13)$$

The shrinking Hubble sphere therefore also corresponds to

$$\varepsilon = -\frac{\dot{H}}{H^2} < 1. \quad (2.2.14)$$

- *Quasi-de Sitter expansion*.—In the limit $\varepsilon \rightarrow 0$, the spacetime becomes de Sitter space

$$ds^2 = dt^2 - e^{2Ht} d\mathbf{x}^2, \quad (2.2.15)$$

where $H = \partial_t \ln a = \text{const}$. 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 (2.2.15) 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, $H^2 = \rho/(3M_{\text{pl}}^2)$, and the continuity equation, $\dot{\rho} = -3H(\rho + P)$, together imply

$$\dot{H} + H^2 = -\frac{1}{6M_{\text{pl}}^2}(\rho + 3P) = -\frac{H^2}{2} \left(1 + \frac{3P}{\rho}\right). \quad (2.2.16)$$

We rearrange this to find that

$$\varepsilon = -\frac{\dot{H}}{H^2} = \frac{3}{2} \left(1 + \frac{P}{\rho} \right) < 1 \quad \Leftrightarrow \quad w \equiv \frac{P}{\rho} < -\frac{1}{3}, \quad (2.2.17)$$

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 continuity equation, $\dot{\rho} = -3H(\rho+P)$, with eq. (2.2.16), we find

$$\left| \frac{d \ln \rho}{d \ln a} \right| = 2\varepsilon < 1. \quad (2.2.18)$$

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

$$\varepsilon \equiv -\frac{\dot{H}}{H^2} = -\frac{d \ln H}{dN} < 1. \quad (2.3.19)$$

Here, we have defined $dN \equiv d \ln a = H dt$, which measures the number of *e-folds* N of inflationary expansion. Equation (2.3.19) 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

$$\eta \equiv \frac{d \ln \varepsilon}{dN} = \frac{\dot{\varepsilon}}{H\varepsilon}. \quad (2.3.20)$$

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* $\phi(t, \mathbf{x})$. As indicated by the notation, the value of the field can depend on time t and the position in space \mathbf{x} . Associated with each field value is a potential energy density $V(\phi)$ (see Fig. 2.4). 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.

The stress-energy tensor of the scalar field is

$$T_{\mu\nu} = \partial_\mu \phi \partial_\nu \phi - g_{\mu\nu} \left(\frac{1}{2} g^{\alpha\beta} \partial_\alpha \phi \partial_\beta \phi - V(\phi) \right). \quad (2.3.21)$$

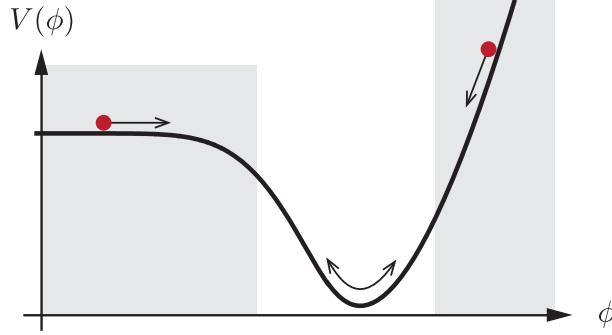


Figure 2.4: Example of a slow-roll potential. Inflation occurs in the shaded parts of the potential.

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

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi). \quad (2.3.22)$$

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,

$$P_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi). \quad (2.3.23)$$

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 (2.3.22) into the *Friedmann equation*, $H^2 = \rho_\phi/(3M_{\text{pl}}^2)$, we get

$$H^2 = \frac{1}{3M_{\text{pl}}^2} \left[\frac{1}{2}\dot{\phi}^2 + V \right]. \quad (2.3.24)$$

Taking a time derivative, we find

$$2H\dot{H} = \frac{1}{3M_{\text{pl}}^2} \left[\ddot{\phi}\dot{\phi} + V_{,\phi}\dot{\phi} \right], \quad (2.3.25)$$

where $V_{,\phi} \equiv dV/d\phi$. Substituting ρ_ϕ and P_ϕ into the second Friedmann equation (2.2.16), $\dot{H} = -(\rho_\phi + P_\phi)/(2M_{\text{pl}}^2)$, we get

$$\dot{H} = -\frac{1}{2}\frac{\dot{\phi}^2}{M_{\text{pl}}^2}. \quad (2.3.26)$$

Notice that \dot{H} is sourced by the kinetic energy density. Combining (2.3.26) with (2.3.25) leads to the *Klein-Gordon equation*

$$\boxed{\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} = 0}. \quad (2.3.27)$$

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

2.3.2 Slow-Roll Inflation

Substituting eq. (2.3.26) into the definition of ε , eq. (2.3.19), we find

$$\varepsilon = \frac{\frac{1}{2}\dot{\phi}^2}{M_{\text{pl}}^2 H^2}. \quad (2.3.28)$$

Inflation ($\varepsilon < 1$) therefore occurs if the kinetic energy, $\frac{1}{2}\dot{\phi}^2$, only makes a small contribution to the total energy, $\rho_\phi = 3M_{\text{pl}}^2 H^2$. 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

$$\delta \equiv -\frac{\ddot{\phi}}{H\dot{\phi}}. \quad (2.3.29)$$

Taking the time-derivative of (2.3.28),

$$\dot{\varepsilon} = \frac{\dot{\phi}\ddot{\phi}}{M_{\text{pl}}^2 H^2} - \frac{\dot{\phi}^2 \dot{H}}{M_{\text{pl}}^2 H^3}, \quad (2.3.30)$$

and comparing to (2.3.20), we find

$$\eta = \frac{\dot{\varepsilon}}{H\varepsilon} = 2\frac{\ddot{\phi}}{H\dot{\phi}} - 2\frac{\dot{H}}{H^2} = 2(\varepsilon - \delta). \quad (2.3.31)$$

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 (2.3.24),

$$H^2 \approx \frac{V}{3M_{\text{pl}}^2}. \quad (2.3.32)$$

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 (2.3.27) to

$$3H\dot{\phi} \approx -V_{,\phi}. \quad (2.3.33)$$

This provides a simple relationship between the gradient of the potential and the speed of the inflaton. Substituting (2.3.32) and (2.3.33) into (2.3.28) gives

$$\varepsilon = \frac{\frac{1}{2}\dot{\phi}^2}{M_{\text{pl}}^2 H^2} \approx \frac{M_{\text{pl}}^2}{2} \left(\frac{V_{,\phi}}{V} \right)^2. \quad (2.3.34)$$

Furthermore, taking the time-derivative of (2.3.33),

$$3\dot{H}\dot{\phi} + 3H\ddot{\phi} = -V_{,\phi\phi}\dot{\phi}, \quad (2.3.35)$$

leads to

$$\delta + \varepsilon = -\frac{\ddot{\phi}}{H\dot{\phi}} - \frac{\dot{H}}{H^2} \approx M_{\text{pl}}^2 \frac{V_{,\phi\phi}}{V}. \quad (2.3.36)$$

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*³

$$\epsilon_v \equiv \frac{M_{\text{pl}}^2}{2} \left(\frac{V_{,\phi}}{V} \right)^2, \quad |\eta_v| \equiv M_{\text{pl}}^2 \frac{|V_{,\phi\phi}|}{V}. \quad (2.3.37)$$

Successful slow-roll inflation occurs when these parameters are small, $\{\epsilon_v, |\eta_v|\} \ll 1$.

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

$$N_{\text{tot}} \equiv \int_{a_S}^{a_E} d \ln a = \int_{t_S}^{t_E} H(t) dt, \quad (2.3.38)$$

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

$$H dt = \frac{H}{\dot{\phi}} d\phi = \frac{1}{\sqrt{2\varepsilon}} \frac{|\dot{\phi}|}{M_{\text{pl}}} \approx \frac{1}{\sqrt{2\epsilon_v}} \frac{|\dot{\phi}|}{M_{\text{pl}}} \quad (2.3.39)$$

to write (2.3.38) as an integral in the field space of the inflaton⁴

$$N_{\text{tot}} = \int_{\phi_S}^{\phi_E} \frac{1}{\sqrt{2\epsilon_v}} \frac{|\dot{\phi}|}{M_{\text{pl}}}, \quad (2.3.40)$$

where ϕ_S and ϕ_E are defined as the boundaries of the interval where $\epsilon_v < 1$. The largest scales observed in the CMB are produced about 60 e-folds before the end of inflation

$$N_* = \int_{\phi_*}^{\phi_E} \frac{1}{\sqrt{2\epsilon_v}} \frac{|\dot{\phi}|}{M_{\text{pl}}} \approx 60. \quad (2.3.41)$$

A successful solution to the horizon problem requires $N_{\text{tot}} > N_* \approx 60$.

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

$$V(\phi) = \frac{1}{2} m^2 \phi^2. \quad (2.3.42)$$

The slow-roll parameters are

$$\epsilon_v(\phi) = \eta_v(\phi) = 2 \left(\frac{M_{\text{pl}}}{\phi} \right)^2. \quad (2.3.43)$$

To satisfy the slow-roll conditions $\epsilon_v, |\eta_v| < 1$, we therefore need to consider super-Planckian values for the inflaton

$$\phi > \sqrt{2} M_{\text{pl}} \equiv \phi_E. \quad (2.3.44)$$

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

$$N(\phi) = \int_{\phi_E}^{\phi} \frac{d\phi}{M_{\text{pl}}} \frac{1}{\sqrt{2\epsilon_v}} = \frac{\phi^2}{4M_{\text{pl}}^2} - \frac{1}{2}. \quad (2.3.45)$$

Fluctuations observed in the CMB are created at

$$\phi_* = 2\sqrt{N_*} M_{\text{pl}} \sim 15 M_{\text{pl}}. \quad (2.3.46)$$

³In contrast, the parameters ε and η are often called the *Hubble slow-roll parameters*. During slow-roll, the parameters are related as follows: $\epsilon_v \approx \varepsilon$ and $\eta_v \approx 2\varepsilon - \frac{1}{2}\eta$.

⁴The absolute value around the integration measure indicates that we pick the overall sign of the integral in such a way as to make $N_{\text{tot}} > 0$.

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.

Scalar field oscillations.—After inflation, the inflaton field ϕ begins to oscillate at the bottom of the potential $V(\phi)$, see Fig. 2.4. 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

$$\ddot{\phi} + 3H\dot{\phi} = -m^2\phi. \quad (2.3.47)$$

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

$$\dot{\rho}_\phi + 3H\rho_\phi = -3HP_\phi = -\frac{3}{2}H(m^2\phi^2 - \dot{\phi}^2). \quad (2.3.48)$$

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

$$\dot{\rho}_\phi + 3H\rho_\phi = -\Gamma_\phi\rho_\phi, \quad (2.3.49)$$

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 equilibrium with some temperature T_{rh} . This reheating temperature is determined by the energy density ρ_{rh} at the end of the reheating epoch. Necessarily, $\rho_{\text{rh}} < \rho_{\phi,E}$ (where $\rho_{\phi,E}$ is the inflaton energy density at the end of inflation). If reheating takes a long time, we may have $\rho_{\text{rh}} \ll \rho_{\phi,E}$. 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 minutes¹ 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, we 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, $k_B \equiv 1$, 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:

$$t_c \equiv \frac{1}{\Gamma} \ll t_H \equiv \frac{1}{H}. \quad (3.1.1)$$

Local thermal equilibrium is then reached before the effect of the expansion becomes relevant. As the universe cools, the rate of interactions typically decreases 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 as²

$$\Gamma \equiv n\sigma v, \quad (3.1.2)$$

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,

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

²For 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 $v \simeq 1$. 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 analysis³

$$\sigma \sim \left| \begin{array}{c} \diagup \\ \diagdown \end{array} \right. \sim \frac{\alpha^2}{T^2}, \quad (3.1.3)$$

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

$$\Gamma = n\sigma v \sim T^3 \times \frac{\alpha^2}{T^2} = \alpha^2 T. \quad (3.1.4)$$

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

$$H \sim \frac{T^2}{M_{\text{pl}}}. \quad (3.1.5)$$

The ratio of (3.1.4) and (3.1.5) is

$$\frac{\Gamma}{H} \sim \frac{\alpha^2 M_{\text{pl}}}{T} \sim \frac{10^{16} \text{ GeV}}{T}, \quad (3.1.6)$$

where we have used $\alpha \sim 0.01$ in the numerical estimate. Below $T \sim 10^{16}$ GeV, but above 100 GeV, the condition (3.1.1) is therefore satisfied and all particles of the Standard Model are in thermal equilibrium.

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 form⁴

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

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, $f \rightarrow e^{-m/T}$. 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, $\rho_r \propto \sum_i \int d^3p f_i(p) E_i(p)$. The result can be written as (see below)

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

where $g_*(T)$ is the number of relativistic degrees of freedom. Figure 3.1 shows the evolution of $g_*(T)$ assuming the particle content of the Standard Model. At early times, all particles are relativistic and $g_* = 106.75$. The value of g_* 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 $g_* = 3.38$.

³Shown in eq. (3.1.3) is the Feynman diagram associated with a $2 \rightarrow 2$ 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. $\sigma \propto (\sqrt{\alpha} \times \sqrt{\alpha})^2 = \alpha^2$.

⁴The precise formula will include the chemical potential (see below).

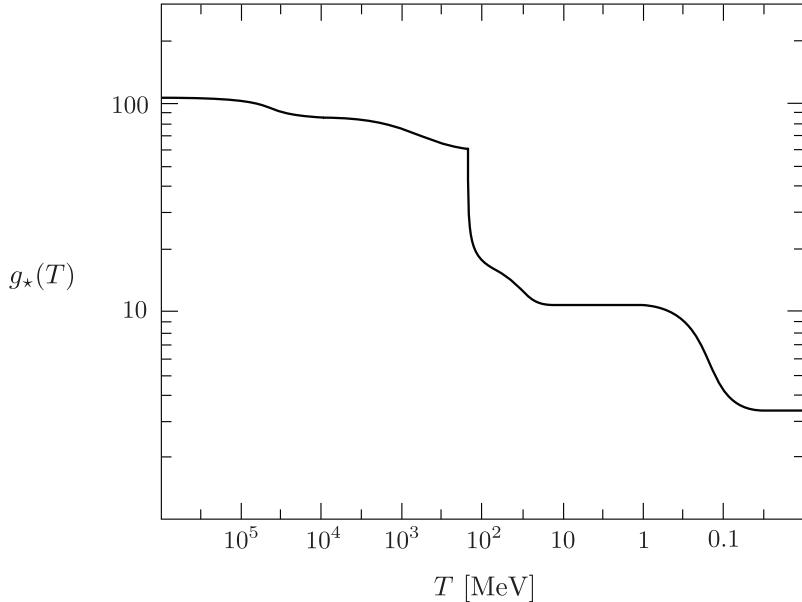


Figure 3.1: Evolution of the number of relativistic degrees of freedom assuming the Standard Model.

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).

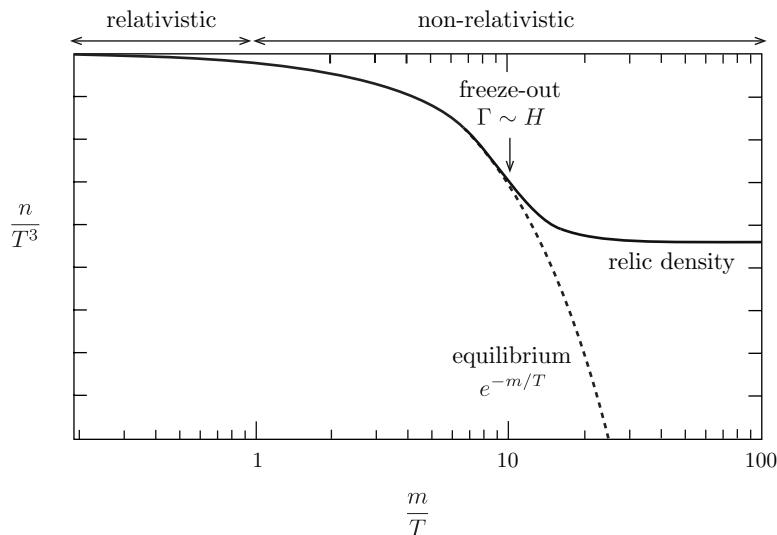


Figure 3.2: A schematic illustration of particle freeze-out. At high temperatures, $T \gg m$, the particle abundance tracks its equilibrium value. At low temperatures, $T \ll m$, the particles freeze out and maintain a density that is much larger than the Boltzmann-suppressed equilibrium abundance.

Below the scale of electroweak symmetry breaking, $T \lesssim 100$ GeV, the gauge bosons of the weak interactions, W^\pm and Z , receive masses, $M_W \approx 80$ GeV and $M_Z \approx 90$ GeV. The cross

⁵This 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.

section associated with processes mediated by the weak force becomes

$$\sigma \sim \left| \begin{array}{c} \diagup \\ \diagdown \end{array} \right|^2 \sim G_F^2 T^2 , \quad (3.1.9)$$

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

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

which drops below unity at $T_{dec} \sim 1 \text{ MeV}$. 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 key events in the thermal history of the universe:

- **Baryogenesis.*** Relativistic quantum field theory requires the existence of anti-particles. This poses a slight puzzle. Particles and anti-particles annihilate through processes such as $e^+ + e^- \rightarrow \gamma + \gamma$. 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

$$\eta_b \equiv \frac{n_b}{n_\gamma} \sim 10^{-9} , \quad (3.1.11)$$

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 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.

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

Event	time t	redshift z	temperature T
Singularity	0	∞	∞
Quantum gravity	$\sim 10^{-43}$ s	–	$\sim 10^{18}$ GeV
Inflation	$\gtrsim 10^{-34}$ s	–	–
Baryogenesis	$\lesssim 20$ ps	$> 10^{15}$	> 100 GeV
EW phase transition	20 ps	10^{15}	100 GeV
QCD phase transition	$20 \mu\text{s}$	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	1100	0.26 eV
Reionization	100–400 Myr	10–30	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.

- **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 $e^- + p^+ \rightarrow H + \gamma$ 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, $e^- + \gamma \rightarrow e^- + \gamma$. 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.

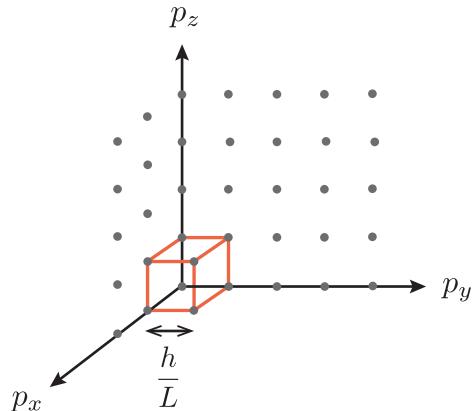
3.2 Equilibrium

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 $\{\mathbf{p}\}$ then is $L^3/h^3 = V/h^3$, and the state density in phase space $\{\mathbf{x}, \mathbf{p}\}$ is

$$\frac{1}{h^3}. \quad (3.2.12)$$

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

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

⁷Strictly speaking, the universe can never truly be in equilibrium since the FRW spacetime doesn't possess a time-like Killing vector. But this is physics and 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* $f(\mathbf{x}, \mathbf{p}, t)$. Because of homogeneity, the distribution function should, in fact, be independent of the position \mathbf{x} . Moreover, isotropy requires that the momentum dependence is only in terms of the magnitude of the momentum $p \equiv |\mathbf{p}|$. 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

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

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

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

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

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

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

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

Similarly, we define the *pressure* as

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

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

$$dN = \frac{g}{(2\pi)^3} f(E) \times R^2 |\mathbf{v}| dt d\Omega . \quad (3.2.19)$$

Not all particles in dV 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 $dA \hat{\mathbf{n}}$ with velocity $\mathbf{v} = |\mathbf{v}| \hat{\mathbf{v}}$ is

$$dN_A = \frac{|\hat{\mathbf{v}} \cdot \hat{\mathbf{n}}| dA}{4\pi R^2} \times dN = \frac{g}{(2\pi)^3} f(E) \times \frac{|\mathbf{v} \cdot \hat{\mathbf{n}}|}{4\pi} dA dt d\Omega , \quad (3.2.20)$$

43 3. Thermal History

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

$$dP(|\mathbf{v}|) = \int \frac{2|\mathbf{p} \cdot \hat{\mathbf{n}}|}{dA dt} dN_A = \frac{g}{(2\pi)^3} f(E) \times \frac{p^2}{2\pi E} \int \cos^2 \theta \sin \theta d\theta d\phi = \frac{g}{(2\pi)^3} \times f(E) \frac{p^2}{3E}, \quad (3.2.21)$$

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

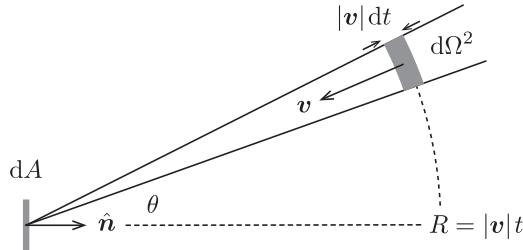


Figure 3.3: Pressure in a weakly interacting gas of particles.

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

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

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 , 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,

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

The change in entropy of a system therefore is

$$dS = \frac{dU + PdV - \mu dN}{T}, \quad (3.2.25)$$

where μdN 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

$$\mu_1 + \mu_2 = \mu_3 + \mu_4. \quad (3.2.26)$$

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

$$\mu_\gamma = 0. \quad (3.2.27)$$

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

$$\mu_{\bar{X}} = -\mu_X, \quad (3.2.28)$$

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 $T_i = T$.⁸

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.⁹ Setting the chemical potential to zero, we get

$$n = \frac{g}{2\pi^2} \int_0^\infty dp \frac{p^2}{\exp[\sqrt{p^2 + m^2}/T] \pm 1}, \quad (3.2.29)$$

$$\rho = \frac{g}{2\pi^2} \int_0^\infty dp \frac{p^2 \sqrt{p^2 + m^2}}{\exp[\sqrt{p^2 + m^2}/T] \pm 1}. \quad (3.2.30)$$

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

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

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

⁸This temperature is often identified with the photon temperature T_γ — the “temperature of the universe”.

⁹For electrons and protons this is a fact, while for neutrinos it is likely true, but not proven.

45 3. Thermal History

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

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

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

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

Relativistic Limit

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

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

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

$$I_-(0) = 2\zeta(3), \quad (3.2.36)$$

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

$$\frac{1}{e^\xi + 1} = \frac{1}{e^\xi - 1} - \frac{2}{e^{2\xi} - 1}, \quad (3.2.37)$$

so that

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

Hence, we get

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

A similar computation for the energy density gives

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

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

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

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

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’)

$$P = \frac{1}{3} \rho. \quad (3.2.43)$$

*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

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

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

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

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 ,

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

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

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

which leads to

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

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

$$\rho \approx mn. \quad (3.2.49)$$

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

$$\rho = mn + \frac{3}{2} nT. \quad (3.2.50)$$

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

$$P = nT \ll \rho = mn. \quad (3.2.51)$$

Exercise.—Derive (4.4.112). Notice that this is nothing but the ideal gas law, $PV = Nk_B T$.

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

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

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

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

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

where $g_*(T)$ 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, $T_i = T \gg m_i$,

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

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, $T_i \neq T \gg m_i$,

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

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).

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	$8 \cdot 2 = 16$
leptons	τ^\pm	1777 MeV	$\frac{1}{2}$	$2 \cdot 2 = 4$
	μ^\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

Figure 3.4 shows the evolution of $g_*(T)$ 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 their internal degrees of freedom we get:¹⁰

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

and hence

$$g_* = g_b + \frac{7}{8} g_f = 106.75. \quad (3.2.57)$$

As the temperature drops, various particle species become non-relativistic and annihilate. To estimate g_* 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 $T \sim$

¹⁰Here, 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, μ^\pm, τ^\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.

$\frac{1}{6}m_t \sim 30 \text{ GeV}$,¹¹ the effective number of relativistic species is reduced to $g_* = 106.75 - \frac{7}{8} \times 12 = 96.25$. The Higgs boson and the gauge bosons W^\pm, Z^0 annihilate next. This happens roughly at the same time. At $T \sim 10 \text{ GeV}$, we have $g_* = 96.26 - (1+3 \cdot 3) = 86.25$. Next, the bottom quarks annihilate ($g_* = 86.25 - \frac{7}{8} \times 12 = 75.75$), followed by the charm quarks and the tau leptons ($g_* = 75.75 - \frac{7}{8} \times (12+4) = 61.75$). Before the strange quarks had time to annihilate, something else happens: matter undergoes the QCD phase transition. At $T \sim 150 \text{ MeV}$, the quarks combine into baryons (protons, neutrons, ...) and mesons (pions, ...). There are many different species of baryons and mesons, but all except the pions (π^\pm, π^0) are non-relativistic below the temperature of the QCD phase transition. Thus, the only particle species left in large numbers are pions, electrons, muons, neutrinos, and photons. The three pions (spin-0) correspond to $g = 3 \cdot 1 = 3$ internal degrees of freedom. We therefore get $g_* = 2 + 3 + \frac{7}{8} \times (4+4+6) = 17.25$. Next, electrons and positrons annihilate. However, to understand this process we first need to talk about entropy.

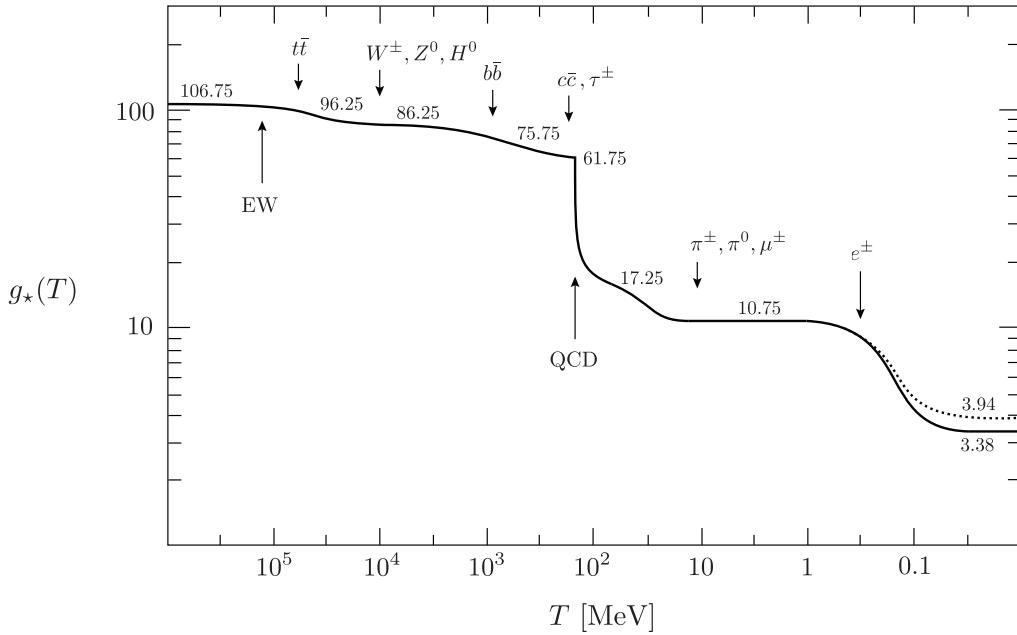


Figure 3.4: Evolution of relativistic degrees of freedom $g_*(T)$ assuming the Standard Model particle content. The dotted line stands for the number of effective degrees of freedom in entropy $g_{*S}(T)$.

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

¹¹The transition from relativistic to non-relativistic behaviour isn't instantaneous. About 80% of the particle-antiparticle annihilations takes place in the interval $T = m \rightarrow \frac{1}{6}m$.

50 3. Thermal History

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 $\mu = 0$:

$$\boxed{\frac{\partial P}{\partial T} = \frac{\rho + P}{T}}. \quad (3.2.58)$$

Consider the second law of thermodynamics: $TdS = dU + PdV$. Using $U = \rho V$, we get

$$\begin{aligned} dS &= \frac{1}{T} \left(d[(\rho + P)V] - VdP \right) \\ &= \frac{1}{T} d[(\rho + P)V] - \frac{V}{T^2}(\rho + P) dT \\ &= d \left[\frac{\rho + P}{T} V \right], \end{aligned} \quad (3.2.59)$$

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

$$\begin{aligned} \frac{dS}{dt} &= \frac{d}{dt} \left[\frac{\rho + P}{T} V \right] \\ &= \frac{V}{T} \left[\frac{d\rho}{dt} + \frac{1}{V} \frac{dV}{dt}(\rho + P) \right] + \frac{V}{T} \left[\frac{dP}{dt} - \frac{\rho + P}{T} \frac{dT}{dt} \right]. \end{aligned} \quad (3.2.60)$$

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 establishes 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

$$s = \frac{\rho + P}{T}. \quad (3.2.61)$$

Using (3.2.40) and (4.4.112), the total entropy density for a collection of different particle species is

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

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

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

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

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

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).

51 3. Thermal History

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

$$N_i \equiv \frac{n_i}{s}. \quad (3.2.65)$$

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

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

Away from particle mass thresholds g_{*S} is approximately constant and $T \propto a^{-1}$, as expected. The factor of $g_{*S}^{-1/3}$ 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 $T \propto g_{*S}^{-1/3} a^{-1}$ into the Friedmann equation

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

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_{*S} changes. For $T \propto t^{-1/2}$, we can integrate the Friedmann equation and get the temperature as a function of time

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

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

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

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 than the Hubble rate $H \sim T^2/M_{\text{pl}}$:

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

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

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

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 temperature¹² 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

$$e^+ + e^- \leftrightarrow \gamma + \gamma. \quad (3.2.72)$$

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,

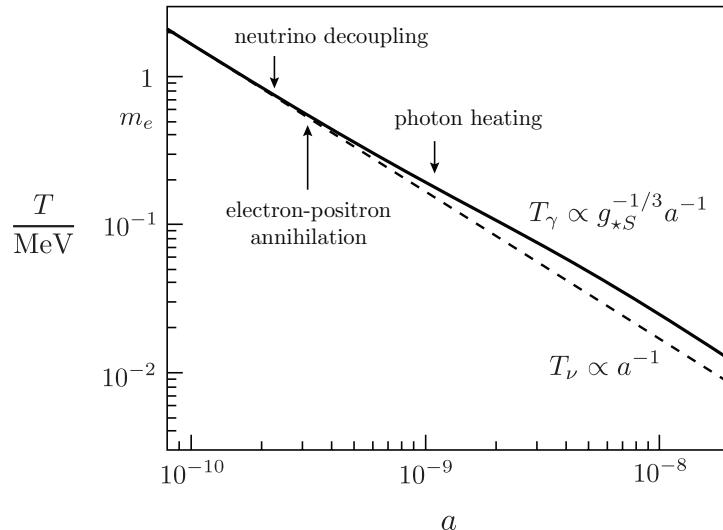


Figure 3.5: Thermal history through electron-positron annihilation. Neutrinos are decoupled and their temperature redshifts simply as $T_\nu \propto a^{-1}$. The energy density of the electron-positron pairs is transferred to the photon gas whose temperature therefore redshifts more slowly, $T_\gamma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

we consider the change in the effective number of degrees of freedom in entropy. If we neglect neutrinos and other decoupled species,¹³ we have

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

¹²For the moment we will restore the subscript on the photon temperature to highlight the difference with the neutrino temperature.

¹³Obviously, entropy is separately conserved for the thermal bath and the decoupled species.

53 3. Thermal History

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

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

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

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

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

where we have introduced the parameter N_{eff} as the *effective* number of neutrino species in the universe. If neutrinos decoupling was instantaneous then we have $N_{\text{eff}} = 3$. 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 account¹⁴ raises the effective number of neutrinos to $N_{\text{eff}} = 3.046$.¹⁵ 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

$$n_\nu = \frac{3}{4} N_{\text{eff}} \times \frac{4}{11} n_\gamma. \quad (3.2.77)$$

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 believed that neutrinos were massless in which case we would have

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

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 universe¹⁶, and as matter-like particles in the late universe (see fig. 3.6). One can show that energy density of massive neutrinos, $\rho_\nu = \sum m_{\nu,i} n_{\nu,i}$, corresponds to

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

¹⁴To get the precise value of N_{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.

¹⁵The Planck constraint on N_{eff} is 3.36 ± 0.34 . This still leaves room for discovering that $N_{\text{eff}} \neq 3.046$, which is one of the avenues in which cosmology could discover new physics beyond the Standard Model.

¹⁶For $m_\nu < 0.2$ eV, neutrinos are relativistic at recombination.

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 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, $0.001 < \Omega_\nu < 0.02$.

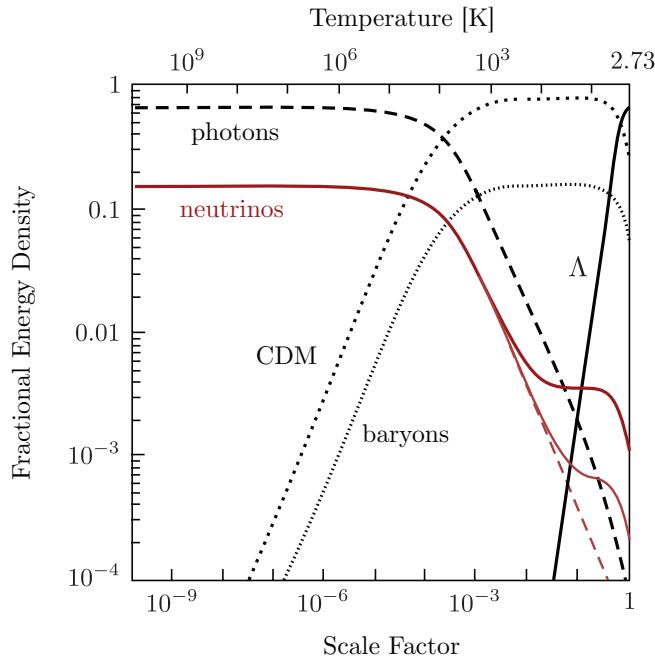


Figure 3.6: Evolution of the fractional energy densities of photons, three neutrino species (one massless and two massive – 0.05 and 0.01 eV), cold dark matter (CDM), baryons, and a cosmological constant (Λ). Notice the change in the behaviour of the two massive neutrinos when they become non-relativistic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

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

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. (1.3.85). To include the effects of interactions we add a collision term to the r.h.s. of (3.3.80),

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

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



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,

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

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*.¹⁷ The second parameter β can be related to α by noting that the collision term has to vanish in (chemical) equilibrium

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

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

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

¹⁷You 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 .

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

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

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*, $[1 - \dots]$.

For $\Gamma_1 \gg H$, the natural state of the system is chemical equilibrium. Imagine that we start with $N_1 \gg N_1^{\text{eq}}$ (while $N_i \sim N_i^{\text{eq}}$, $i = 2, 3, 4$). 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^{eq} . Similarly, if $N_1 \ll N_1^{\text{eq}}$, the r.h.s. of (3.3.86) is positive and N_1 is driven towards N_1^{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. $N_1 = \text{const}$. 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).

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 $\bar{\ell}$,

$$X + \bar{X} \leftrightarrow \ell + \bar{\ell}. \quad (3.3.87)$$

Moreover, we assume that the light particles are tightly coupled to the cosmic plasma,¹⁸ so that throughout they maintain their equilibrium densities, $n_\ell = n_\ell^{\text{eq}}$. 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

$$\frac{dN_X}{dt} = -s \langle \sigma v \rangle \left[N_X^2 - (N_X^{\text{eq}})^2 \right], \quad (3.3.88)$$

¹⁸This would be case case, for instance, if ℓ and $\bar{\ell}$ were electrically charged.

57 3. Thermal History

where $N_X^{\text{eq}} \equiv n_X^{\text{eq}}/s$. 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,

$$x \equiv \frac{M_X}{T}. \quad (3.3.89)$$

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

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

where we have assumed that $T \propto a^{-1}$ (i.e. $g_{*S} \approx \text{const.} \equiv g_{*S}(M_X)$) 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*,

$$\boxed{\frac{dN_X}{dx} = -\frac{\lambda}{x^2} \left[N_X^2 - (N_X^{\text{eq}})^2 \right]}, \quad (3.3.91)$$

where we have defined

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

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,

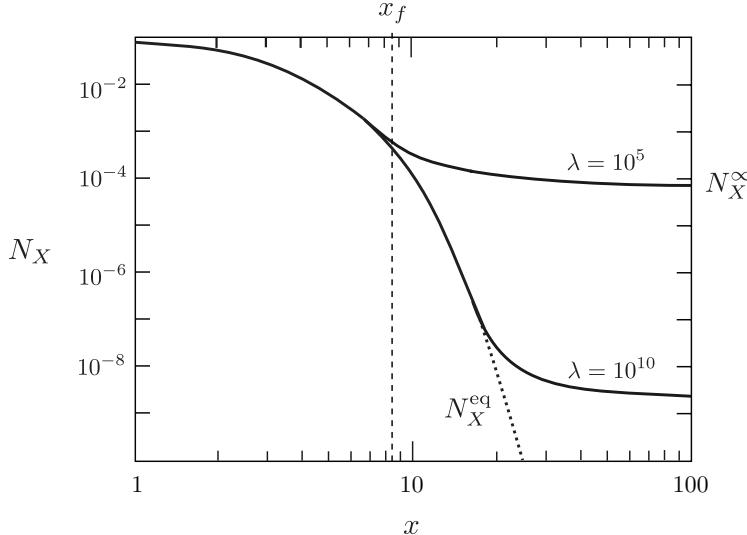


Figure 3.7: Abundance of dark matter particles as the temperature drops below the mass.

at very high temperatures, $x < 1$, we have $N_X \approx N_X^{\text{eq}} \simeq 1$. However, at low temperatures, $x \gg 1$, the equilibrium abundance becomes exponentially suppressed, $N_X^{\text{eq}} \sim e^{-x}$. 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

58 3. Thermal History

much larger than N_X^{eq} (see Fig. 3.7). Thus at late times, we can drop N_X^{eq} from the Boltzmann equation,

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

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

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

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

$$N_X^\infty \simeq \frac{x_f}{\lambda}. \quad (3.3.95)$$

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 $\Gamma(x_f) = H(x_f)$.

Equation (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:

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

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

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

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

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

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

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

This reproduces the observed dark matter density if

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

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*¹⁹), 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



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

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

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

$$\left(\frac{n_H}{n_e n_p} \right)_{\text{eq}} = \frac{g_H}{g_e g_p} \left(\frac{m_H}{m_e m_p} \frac{2\pi}{T} \right)^{3/2} e^{(m_p + m_e - m_H)/T}. \quad (3.3.102)$$

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

$$B_H \equiv m_p + m_e - m_H = 13.6 \text{ eV}. \quad (3.3.103)$$

The number of internal degrees of freedom are $g_p = g_e = 2$ and $g_H = 4$.²⁰ Since, as far as we know, the universe isn't electrically charged, we have $n_e = n_p$. Eq. (3.3.102) therefore becomes

$$\left(\frac{n_H}{n_e^2} \right)_{\text{eq}} = \left(\frac{2\pi}{m_e T} \right)^{3/2} e^{B_H/T}. \quad (3.3.104)$$

¹⁹Don't ask me why this is called *recombination*; this is the first time electrons and nuclei combined.

²⁰The 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 $g_H = 1 + 3 = 4$.

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

$$X_e \equiv \frac{n_e}{n_b} , \quad (3.3.105)$$

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

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

where $\eta_b = 5.5 \times 10^{-10} (\Omega_b h^2 / 0.020)$ 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 $n_b \approx n_p + n_H = n_e + n_H$ and hence

$$\frac{1 - X_e}{X_e^2} = \frac{n_H}{n_e^2} n_b . \quad (3.3.107)$$

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

$$\left(\frac{1 - X_e}{X_e^2} \right)_{\text{eq}} = \frac{2\zeta(3)}{\pi^2} \eta_b \left(\frac{2\pi T}{m_e} \right)^{3/2} e^{B_H/T} . \quad (3.3.108)$$

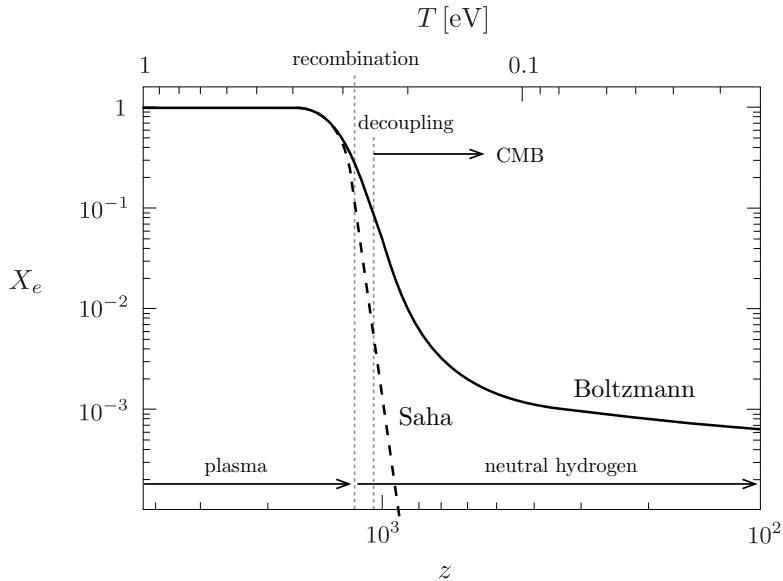


Figure 3.8: Free electron fraction as a function of redshift.

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.

Hydrogen Recombination

Let us define the recombination temperature T_{rec} as the temperature where²¹ $X_e = 10^{-1}$ in (3.3.108), i.e. when 90% of the electrons have combined with protons to form hydrogen. We find

$$T_{rec} \approx 0.3 \text{ eV} \simeq 3600 \text{ K} . \quad (3.3.109)$$

²¹There is nothing deep about the choice $X_e(T_{rec}) = 10^{-1}$. It is as arbitrary as it looks.

61 3. Thermal History

The reason that $T_{rec} \ll B_H = 13.6$ eV is that there are very many photons for each hydrogen atom, $\eta_b \sim 10^{-9} \ll 1$. Even when $T < B_H$, the high-energy tail of the photon distribution contains photons with energy $E > B_H$ 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 $T_0 = 2.7$ K, gives the redshift of recombination,

$$z_{rec} \approx 1320. \quad (3.3.110)$$

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

$$t_{rec} = \frac{t_0}{(1 + z_{rec})^{3/2}} \sim 290\,000 \text{ yrs}. \quad (3.3.111)$$

Photon Decoupling

Photons are most strongly coupled to the primordial plasma through their interactions with electrons

$$e^- + \gamma \leftrightarrow e^- + \gamma, \quad (3.3.112)$$

with an interaction rate given by

$$\Gamma_\gamma \approx n_e \sigma_T, \quad (3.3.113)$$

where $\sigma_T \approx 2 \times 10^{-3} \text{ MeV}^{-2}$ 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,

$$\Gamma_\gamma(T_{dec}) \sim H(T_{dec}). \quad (3.3.114)$$

Writing

$$\Gamma_\gamma(T_{dec}) = n_b X_e(T_{dec}) \sigma_T = \frac{2\zeta(3)}{\pi^2} \eta_b \sigma_T X_e(T_{dec}) T_{dec}^3, \quad (3.3.115)$$

$$H(T_{dec}) = H_0 \sqrt{\Omega_m} \left(\frac{T_{dec}}{T_0} \right)^{3/2}. \quad (3.3.116)$$

we get

$$X_e(T_{dec}) T_{dec}^{3/2} \sim \frac{\pi^2}{2\zeta(3)} \frac{H_0 \sqrt{\Omega_m}}{\eta_b \sigma_T T_0^{3/2}}. \quad (3.3.117)$$

Using the Saha equation for $X_e(T_{dec})$, we find

$$T_{dec} \sim 0.27 \text{ eV}. \quad (3.3.118)$$

Notice that although T_{dec} isn't far from T_{rec} , the ionization fraction decreases significantly between recombination and decoupling, $X_e(T_{rec}) \simeq 0.1 \rightarrow X_e(T_{dec}) \simeq 0.01$. 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

$$z_{dec} \sim 1100, \quad (3.3.119)$$

$$t_{dec} \sim 380\,000 \text{ yrs}. \quad (3.3.120)$$

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, $n_H \approx n_H^{\text{eq}}$. The Boltzmann equation for the electron density can then be written as

$$\frac{1}{a^3} \frac{d(n_e a^3)}{dt} = -\langle \sigma v \rangle [n_e^2 - (n_e^{\text{eq}})^2]. \quad (3.3.121)$$

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

$$\langle \sigma v \rangle \simeq \sigma_T \left(\frac{B_H}{T} \right)^{1/2}. \quad (3.3.122)$$

Writing $n_e = n_b X_e$ and using that $n_b a^3 = \text{const.}$, we find

$$\frac{dX_e}{dx} = -\frac{\lambda}{x^2} [X_e^2 - (X_e^{\text{eq}})^2], \quad (3.3.123)$$

where $x \equiv B_H/T$. We have used the fact that the universe is matter-dominated at recombination and defined

$$\lambda \equiv \left[\frac{n_b \langle \sigma v \rangle}{x H} \right]_{x=1} = 3.9 \times 10^3 \left(\frac{\Omega_b h}{0.03} \right). \quad (3.3.124)$$

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),

$$X_e^\infty \simeq \frac{x_f}{\lambda} = 0.9 \times 10^{-3} \left(\frac{x_f}{x_{rec}} \right) \left(\frac{0.03}{\Omega_b h} \right). \quad (3.3.125)$$

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

$$X_e(T_f) T_f = \frac{\pi^2}{2\zeta(3)} \frac{H_0 \sqrt{\Omega_m}}{\eta_b \sigma_T T_0^{3/2} B_H^{1/2}}. \quad (3.3.126)$$

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,

$$\frac{n_{\text{He}}}{n_{\text{H}}} \sim \frac{1}{16}. \quad (3.3.127)$$

Figure 3.9 summarizes the four steps that will lead us from protons and neutrons to helium.

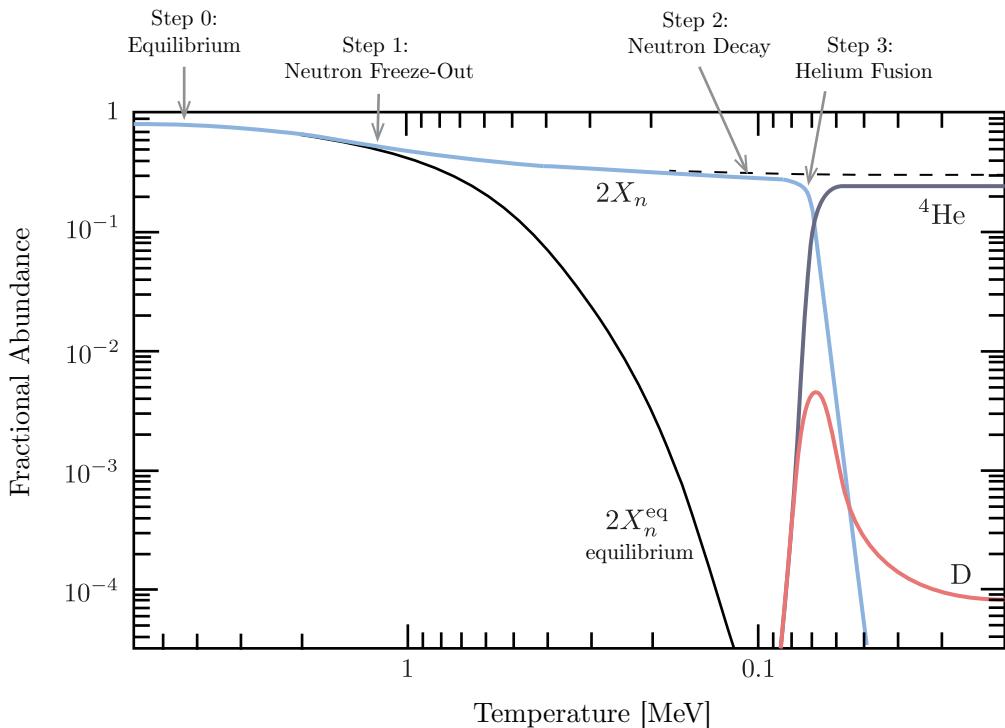


Figure 3.9: Numerical results for helium production in the early universe.

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 ^3He .

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

$$\begin{aligned} n + \nu_e &\leftrightarrow p^+ + e^- , \\ n + e^+ &\leftrightarrow p^+ + \bar{\nu}_e . \end{aligned} \quad (3.3.128)$$

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^{eq} , we then have

$$\left(\frac{n_n}{n_p} \right)_{\text{eq}} = \left(\frac{m_n}{m_p} \right)^{3/2} e^{-(m_n - m_p)/T} . \quad (3.3.129)$$

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

$$\left(\frac{n_n}{n_p} \right)_{\text{eq}} = e^{-Q/T} , \quad (3.3.130)$$

where $Q \equiv m_n - m_p = 1.30$ 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



Since $\mu_\gamma = 0$, we have $\mu_n + \mu_p = \mu_D$. To remove the dependence on the chemical potentials we consider

$$\left(\frac{n_D}{n_n n_p} \right)_{\text{eq}} = \frac{3}{4} \left(\frac{m_D}{m_n m_p} \frac{2\pi}{T} \right)^{3/2} e^{-(m_D - m_n - m_p)/T} , \quad (3.3.132)$$

65 3. Thermal History

where, as before, we have used (3.3.101) for n_i^{eq} (with $g_D = 3$ and $g_p = g_n = 2$). In the prefactor, m_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_D is crucial: it is the binding energy of deuterium

$$B_D \equiv m_n + m_p - m_D = 2.22 \text{ MeV}. \quad (3.3.133)$$

Therefore, as long as chemical equilibrium holds the deuterium-to-proton ratio is

$$\left(\frac{n_D}{n_p}\right)_{\text{eq}} = \frac{3}{4} n_n^{\text{eq}} \left(\frac{4\pi}{m_p T}\right)^{3/2} e^{B_D/T}. \quad (3.3.134)$$

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,

$$n_n \sim n_b = \eta_b n_\gamma = \eta_b \times \frac{2\zeta(3)}{\pi^2} T^3. \quad (3.3.135)$$

Equation (3.3.134) then becomes

$$\left(\frac{n_D}{n_p}\right)_{\text{eq}} \approx \eta_b \left(\frac{T}{m_p}\right)^{3/2} e^{B_D/T}. \quad (3.3.136)$$

The smallness of the baryon-to-photon ratio η_b inhibits the production of deuterium until the temperature drops well beneath the binding energy B_D . The temperature has to drop enough so that $e^{B_D/T}$ can compete with $\eta_b \sim 10^{-9}$. 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 ${}^4\text{He}$. As we have seen, weak interactions keep neutrons and protons in equilibrium until $T \sim \text{MeV}$. 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

$$X_n \equiv \frac{n_n}{n_n + n_p}. \quad (3.3.137)$$

From the equilibrium ratio of neutrons to protons (3.3.130), we then get

$$X_n^{\text{eq}}(T) = \frac{e^{-Q/T}}{1 + e^{-Q/T}}. \quad (3.3.138)$$

Neutrons follows this equilibrium abundance until neutrinos decouple at²² $T_f \sim T_{dec} \sim 0.8$ 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

$$X_n^{\text{eq}}(0.8 \text{ MeV}) = 0.17. \quad (3.3.139)$$

²²If is fortunate that $T_f \sim Q$. This seems to be a coincidence: 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 $T_f \ll Q$!

We will take this as a rough estimate for the final freeze-out abundance,

$$X_n^\infty \sim X_n^{\text{eq}}(0.8 \text{ MeV}) \sim \frac{1}{6}. \quad (3.3.140)$$

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 $n_\ell = n_\ell^{\text{eq}}$), we find

$$\frac{1}{a^3} \frac{d(n_n a^3)}{dt} = -\Gamma_n \left[n_n - \left(\frac{n_n}{n_p} \right)_{\text{eq}} n_p \right], \quad (3.3.141)$$

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

$$\frac{dX_n}{dt} = -\Gamma_n \left[X_n - (1 - X_n) e^{-Q/T} \right]. \quad (3.3.142)$$

Instead of trying to solve this for X_n as a function of time, we introduce a new evolution variable

$$x \equiv \frac{Q}{T}. \quad (3.3.143)$$

We write the l.h.s. of (3.3.142) as

$$\frac{dX_n}{dt} = \frac{dx}{dt} \frac{dX_n}{dx} = -\frac{x}{T} \frac{dT}{dt} \frac{dX_n}{dx} = xH \frac{dX_n}{dx}, \quad (3.3.144)$$

where in the last equality we used that $T \propto a^{-1}$. During BBN, we have

$$H = \sqrt{\frac{\rho}{3M_{\text{pl}}^2}} = \underbrace{\frac{\pi}{3} \sqrt{\frac{g_*}{10}} \frac{Q^2}{M_{\text{pl}}} \frac{1}{x^2}}_{\equiv H_1 \approx 1.13 \text{ s}^{-1}}, \quad \text{with } g_* = 10.75. \quad (3.3.145)$$

Eq. (3.3.142) then becomes

$$\frac{dX_n}{dx} = \frac{\Gamma_n}{H_1} x [e^{-x} - X_n(1 + e^{-x})]. \quad (3.3.146)$$

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

$$\Gamma_n(x) = \frac{255}{\tau_n} \cdot \frac{12 + 6x + x^2}{x^5}, \quad (3.3.147)$$

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)

$$X_n^\infty \equiv X_n(x = \infty) = 0.15. \quad (3.3.148)$$

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

$$X_n(t) = X_n^\infty e^{-t/\tau_n} = \frac{1}{6} e^{-t/\tau_n}, \quad (3.3.149)$$

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,



Only when deuterium is available can helium be formed,



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_{nuc} when the deuterium fraction in equilibrium would be of order one, i.e. $(n_D/n_p)_{\text{eq}} \sim 1$. Using (3.3.136), I find

$$T_{\text{nuc}} \sim 0.06 \text{ MeV}, \quad (3.3.153)$$

which via (3.2.68) with $g_* = 3.38$ translates into

$$t_{\text{nuc}} = 120 \text{ sec} \left(\frac{0.1 \text{ MeV}}{T_{\text{nuc}}} \right)^2 \sim 330 \text{ sec}. \quad (3.3.154)$$

Comment.—From fig. 3.9, we see that a better estimate would be $n_D^{\text{eq}}(T_{\text{nuc}}) \simeq 10^{-3} n_p^{\text{eq}}(T_{\text{nuc}})$. This gives $T_{\text{nuc}} \simeq 0.07$ MeV and $t_{\text{nuc}} \simeq 250$ sec. Notice that $t_{\text{nuc}} \ll \tau_n$, so eq. (3.3.149) won't be very sensitive to the estimate for t_{nuc} .

Substituting $t_{\text{nuc}} \sim 330$ sec into (3.3.149), we find

$$X_n(t_{\text{nuc}}) \sim \frac{1}{8}. \quad (3.3.155)$$

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 $t \sim t_{\text{nuc}}$ then are processed into ${}^4\text{He}$. Since two neutrons go into one nucleus of ${}^4\text{He}$, the final ${}^4\text{He}$ abundance is equal to half of the neutron abundance at t_{nuc} , i.e. $n_{\text{He}} = \frac{1}{2}n_n(t_{\text{nuc}})$, or

$$\frac{n_{\text{He}}}{n_{\text{H}}} = \frac{n_{\text{He}}}{n_p} \simeq \frac{\frac{1}{2}X_n(t_{\text{nuc}})}{1 - X_n(t_{\text{nuc}})} \sim \frac{1}{2}X_n(t_{\text{nuc}}) \sim \boxed{\frac{1}{16}}, \quad (3.3.156)$$

as we wished to show. Sometimes, the result is expressed as the mass fraction of helium,

$$\boxed{\frac{4n_{\text{He}}}{n_{\text{H}}} \sim \frac{1}{4}}. \quad (3.3.157)$$

This prediction is consistent with the observed helium in the universe (see Fig. 3.10).

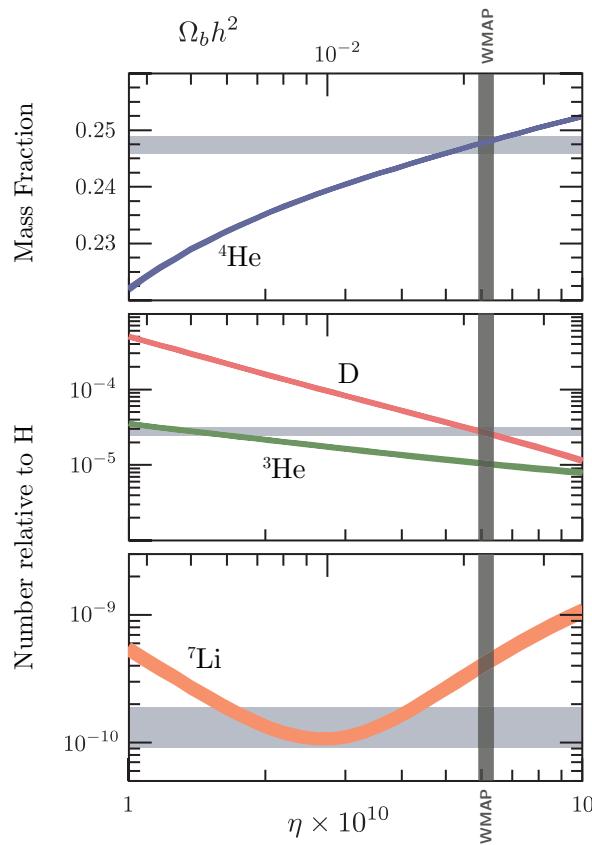


Figure 3.10: Theoretical predictions (colored bands) and observational constraints (grey bands).

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_* : the number of relativistic degrees of freedom determines the Hubble parameter during the radiation era, $H \propto g_*^{1/2}$, and hence affects the freeze-out temperature

$$G_F^2 T_f^5 \sim \sqrt{G_N g_*} T_f^2 \rightarrow T_f \propto g_*^{1/6}. \quad (3.3.158)$$

Increasing g_* increases T_f , which increases the n/p ratio at freeze-out and hence increases the final helium abundance.

- τ_n : a large neutron lifetime would reduce the amount of neutron decay after freeze-out and therefore would increase the final helium abundance.
- 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_b : the amount of helium increases with increasing η_b as nucleosynthesis 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). Figure 3.10 shows that theoretical predictions for the light element abundances as a function of η_b (or Ω_b).²³ The fact that we find reasonably good quantitative agreement with observations is one of the great triumphs of the Big Bang model.

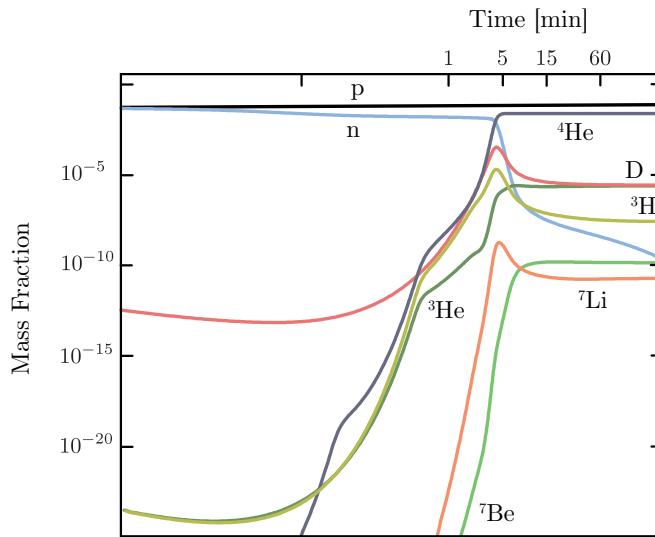


Figure 3.11: Numerical results for the evolution of light element abundances.

²³The shape of the curves in Fig. 3.11 is relatively easy to understand: The abundance of ^4He increases with increasing η_b as nucleosynthesis starts earlier for larger baryon density. D and ^3He are burnt by fusion, thus their abundances decrease as η_b increases. Finally, ^7Li is destroyed by protons at low η_b with an efficiency that increases with η_b . On the other hand, its precursor ^7Be is produced more efficiently as η_b increases. This explains the valley in the curve for ^7Li .

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 this chapter, we will develop the formalism of cosmological perturbation theory.

The Einstein equations couple perturbations in the stress-energy tensor to those in the metric, so the two need to be studied simultaneously. We write the small perturbations of the metric and the stress-energy tensor as

$$\begin{aligned} g_{\mu\nu}(\eta, \mathbf{x}) &= \bar{g}_{\mu\nu}(\eta) + \delta g_{\mu\nu}(\eta, \mathbf{x}), \\ T_{\mu\nu}(\eta, \mathbf{x}) &= \bar{T}_{\mu\nu}(\eta) + \delta T_{\mu\nu}(\eta, \mathbf{x}). \end{aligned}$$

To avoid clutter we will often drop the argument (η, \mathbf{x}) on the perturbations.

4.1 Metric Perturbations

To avoid unnecessary technical distractions, we will take the background metric $\bar{g}_{\mu\nu}$ to be the flat FRW metric. The perturbed spacetime can then be written as

$$ds^2 = a^2(\eta) \left[(1 + 2A)d\eta^2 - 2B_i dx^i d\eta - (\delta_{ij} + h_{ij})dx^i dx^j \right], \quad (4.1.1)$$

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. $h^i{}_i = \delta^{ij}h_{ij}$.

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

$$B_i = \underbrace{\partial_i B}_{\text{scalar}} + \underbrace{\hat{B}_i}_{\text{vector}}, \quad (4.1.2)$$

with $\partial^i \hat{B}_i = 0$. Similarly, any rank-2 symmetric tensor can be written

$$h_{ij} = \underbrace{2C\delta_{ij} + 2\partial_{\langle i}\partial_{j\rangle} E}_{\text{scalar}} + \underbrace{2\partial_{(i}\hat{E}_{j)}}_{\text{vector}} + \underbrace{2\hat{E}_{ij}}_{\text{tensor}}, \quad (4.1.3)$$

where

$$\partial_{\langle i}\partial_{j\rangle} E \equiv \left(\partial_i \partial_j - \frac{1}{3} \delta_{ij} \nabla^2 \right) E, \quad (4.1.4)$$

$$\partial_{(i}\hat{E}_{j)} \equiv \frac{1}{2} \left(\partial_i \hat{E}_j + \partial_j \hat{E}_i \right). \quad (4.1.5)$$

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.

*Gauge problem.**—The metric perturbations in (4.1.1) 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 specific 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.

Fictitious perturbations.—For example, consider a flat FRW spacetime and make the following change of the spatial coordinates, $x^i \mapsto \tilde{x}^i = x^i + \xi^i(\eta, \mathbf{x})$. We assume that ξ^i is small, so that it can also be treated as a perturbation. Using $dx^i = d\tilde{x}^i - \partial_\eta \xi^i d\eta - \partial_k \xi^i d\tilde{x}^k$, the line element becomes

$$ds^2 = a^2(\eta) [d\eta^2 - 2\xi'_i d\tilde{x}^i d\eta - (\delta_{ij} + 2\partial_{(i}\xi_{j)}) d\tilde{x}^i d\tilde{x}^j], \quad (4.1.6)$$

where we have dropped terms that are quadratic in ξ^i and defined $\xi'_i \equiv \partial_\eta \xi_i$. 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.

As another example, consider a change in the time slicing, $\eta \mapsto \eta + \xi^0(\eta, \mathbf{x})$. The homogeneous density of the universe then gets perturbed, $\rho(\eta) \mapsto \rho(\eta + \xi^0(\eta, \mathbf{x})) = \bar{\rho}(\eta) + \bar{\rho}'\xi^0$. Even in an unperturbed universe, a change of the time coordinate can therefore introduce a fictitious density perturbation

$$\delta\rho = \bar{\rho}'\xi^0. \quad (4.1.7)$$

Conversely, we can also remove a real perturbation in the energy density by choosing the hypersurface of constant time to coincide with the hypersurface of constant energy density. We then have $\delta\rho = 0$ 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.

We need to face the fact that the metric perturbations can be changed by a change of coordinates. Consider the following transformation

$$X^\mu \mapsto \tilde{X}^\mu \equiv X^\mu + \xi^\mu(\eta, \mathbf{x}), \quad \text{where} \quad \xi^0 \equiv T, \quad \xi^i \equiv L^i = \partial^i L + \hat{L}^i. \quad (4.1.8)$$

We have split the spatial shift L^i into a scalar, L , and a divergenceless vector, \hat{L}^i . In the next insert, I will show how the metric transforms under this change of coordinates. In terms of the

73 4. Cosmological Perturbation Theory

SVT-decomposition, we get

$$A \mapsto A - T' - \mathcal{H}T, \quad (4.1.9)$$

$$B \mapsto B + T - L', \quad \hat{B}_i \mapsto \hat{B}_i - \hat{L}'_i, \quad (4.1.10)$$

$$C \mapsto C - \mathcal{H}T - \frac{1}{3}\nabla^2L, \quad (4.1.11)$$

$$E \mapsto E - L, \quad \hat{E}_i \mapsto \hat{E}_i - \hat{L}_i, \quad \hat{E}_{ij} \mapsto \hat{E}_{ij}. \quad (4.1.12)$$

Derivation.*—The trick is to exploit the invariance of the spacetime interval,

$$ds^2 = g_{\mu\nu}(X)dX^\mu dX^\nu = \tilde{g}_{\alpha\beta}(\tilde{X})d\tilde{X}^\alpha d\tilde{X}^\beta, \quad (4.1.13)$$

where I have used a different set of dummy indices on both sides to make the next few lines clearer. Writing $d\tilde{X}^\alpha = (\partial\tilde{X}^\alpha/\partial X^\mu)dX^\mu$ (and similarly for dX^β), we find

$$g_{\mu\nu}(X) = \frac{\partial\tilde{X}^\alpha}{\partial X^\mu} \frac{\partial\tilde{X}^\beta}{\partial X^\nu} \tilde{g}_{\alpha\beta}(\tilde{X}). \quad (4.1.14)$$

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 (4.1.14) implies for the transformation of the metric perturbations in (4.1.1). I will work out the 00-component as an example and leave the rest as an exercise. Consider $\mu = \nu = 0$ in (4.1.14):

$$g_{00}(X) = \frac{\partial\tilde{X}^\alpha}{\partial\eta} \frac{\partial\tilde{X}^\beta}{\partial\eta} \tilde{g}_{\alpha\beta}(\tilde{X}). \quad (4.1.15)$$

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 $\partial\tilde{X}^i/\partial\eta$ 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. (4.1.15) therefore reduces to

$$g_{00}(X) = \left(\frac{\partial\tilde{\eta}}{\partial\eta} \right)^2 \tilde{g}_{00}(\tilde{X}). \quad (4.1.16)$$

Substituting (4.1.8) and (4.1.1), we get

$$\begin{aligned} a^2(\eta)(1+2A) &= (1+T')^2 a^2(\eta+T)(1+2\tilde{A}) \\ &= (1+2T'+\dots)(a(\eta)+a'T+\dots)^2(1+2\tilde{A}) \\ &= a^2(\eta)(1+2\mathcal{H}T+2T'+2\tilde{A}+\dots), \end{aligned} \quad (4.1.17)$$

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

$$A \mapsto \tilde{A} = A - T' - \mathcal{H}T. \quad (4.1.18)$$

I leave it to you to repeat the argument for the other metric components and show that

$$B_i \mapsto \tilde{B}_i = B_i + \partial_i T - L'_i, \quad (4.1.19)$$

$$h_{ij} \mapsto \tilde{h}_{ij} = h_{ij} - 2\partial_{(i} L_{j)} - 2\mathcal{H}T\delta_{ij}. \quad (4.1.20)$$

In terms of the SVT-decomposition this leads to (4.1.9)–(4.1.12).

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*

$$\Psi \equiv A + \mathcal{H}(B - E') + (B - E')' , \quad \hat{\Phi}_i \equiv \hat{E}'_i - \hat{B}_i , \quad \hat{E}_{ij} , \quad (4.1.21)$$

$$\Phi \equiv -C - \mathcal{H}(B - E') + \frac{1}{3}\nabla^2 E . \quad (4.1.22)$$

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.

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 (4.1.8) to set two of the four scalar metric perturbations to zero:

- *Newtonian gauge*.—The choice

$$B = E = 0 , \quad (4.1.23)$$

gives the metric

$$ds^2 = a^2(\eta) [(1 + 2\Psi)d\eta^2 - (1 - 2\Phi)\delta_{ij}dx^i dx^j] . \quad (4.1.24)$$

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 (4.1.21) and (4.1.22). For perturbations that decay at spatial infinity, the Newtonian gauge is unique (i.e. the gauge is fixed completely). 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 (see Chapter 5).

- *Spatially-flat gauge*.—A convenient gauge for computing inflationary perturbations is

$$C = E = 0 . \quad (4.1.25)$$

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 Matter Perturbations

We write the perturbed stress-energy tensor as

$$T^0{}_0 = \bar{\rho}(\eta) + \delta\rho , \quad (4.2.26)$$

$$T^i{}_0 = [\bar{\rho}(\eta) + \bar{P}(\eta)] v^i , \quad (4.2.27)$$

$$T^i{}_j = -[\bar{P}(\eta) + \delta P] \delta^i_j - \Pi^i{}_j , \quad (4.2.28)$$

where v_i is the *bulk velocity* and Π^i_j is a transverse and traceless tensor describing *anisotropic stress*. We will use q^i for the *momentum density* $(\bar{\rho} + \bar{P})v^i$. In case there are several contributions to the stress-energy tensor (e.g. photons, baryons, dark matter, etc.), they are added: $T_{\mu\nu} = \sum_a T_{\mu\nu}^{(a)}$. This implies

$$\delta\rho = \sum_a \delta\rho_a, \quad \delta P = \sum_a \delta P_a, \quad q^i = \sum_a q_{(a)}^i, \quad \Pi^{ij} = \sum_a \Pi_{(a)}^{ij}. \quad (4.2.29)$$

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,

$$q_i = \partial_i q + \hat{q}_i, \quad (4.2.30)$$

and Π_{ij} has scalar, vector and tensor parts,

$$\Pi_{ij} = \partial_{\langle i} \partial_{j\rangle} \Pi + \partial_{(i} \hat{\Pi}_{j)} + \hat{\Pi}_{ij}. \quad (4.2.31)$$

It is also convenient to write the density perturbations in terms of the dimensionless density contrast $\delta \equiv \delta\rho/\rho$. In summary, scalar perturbations of the total matter are described by four perturbation variables, $(\delta, \delta P, v, \Pi)$. Similarly, the perturbations of distinct species $a = \gamma, \nu, c, b, \dots$ are represented by $(\delta_a, \delta P_a, v_a, \Pi_a)$.

Under the coordinate transformation (4.1.8), the stress-energy tensor transform as

$$T^\mu_\nu(X) = \frac{\partial X^\mu}{\partial \tilde{X}^\alpha} \frac{\partial \tilde{X}^\beta}{\partial X^\nu} \tilde{T}^\alpha_\beta(\tilde{X}). \quad (4.2.32)$$

Evaluating this for the different components, we find

$$\delta\rho \mapsto \delta\rho - T\bar{\rho}', \quad (4.2.33)$$

$$\delta P \mapsto \delta P - T\bar{P}', \quad (4.2.34)$$

$$q_i \mapsto q_i + (\bar{\rho} + \bar{P})L'_i, \quad (4.2.35)$$

$$v_i \mapsto v_i + L'_i, \quad (4.2.36)$$

$$\Pi_{ij} \mapsto \Pi_{ij}. \quad (4.2.37)$$

Exercise.—Confirm eqs. (4.2.33)–(4.2.37). [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 $\boldsymbol{\varepsilon}$.]

There are various gauge-invariant quantities that can be formed from metric and matter variables. One useful combination is

$$\bar{\rho}\Delta \equiv \delta\rho + \bar{\rho}'(v + B), \quad (4.2.38)$$

where $v_i = \partial_i v$. The quantity Δ is called the *comoving-gauge density perturbation*.

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

$$\delta\rho = 0. \quad (4.2.39)$$

- *Comoving gauge.*—Similarly, we can ask for the scalar momentum density to vanish,

$$q = 0. \quad (4.2.40)$$

Fluctuations in comoving gauge are most naturally connected to the inflationary initial conditions. This will be explained in §4.4 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 $B = 0$.

4.3 Equations of Motion

In this section, we will determine the linearized equations of motion for the matter and metric perturbations. We will work in *Newtonian gauge* where the metric takes the following form

$$g_{\mu\nu} = a^2 \begin{pmatrix} 1 + 2\Psi & 0 \\ 0 & -(1 - 2\Phi)\delta_{ij} \end{pmatrix}. \quad (4.3.41)$$

Exercise.—Show that the connection coefficients associated with the metric (4.3.41) are

$$\Gamma_{00}^0 = \mathcal{H} + \Psi', \quad (4.3.42)$$

$$\Gamma_{i0}^0 = \partial_i \Psi, \quad (4.3.43)$$

$$\Gamma_{00}^i = \delta^{ij} \partial_j \Psi, \quad (4.3.44)$$

$$\Gamma_{ij}^0 = \mathcal{H}\delta_{ij} - [\Phi' + 2\mathcal{H}(\Phi + \Psi)]\delta_{ij}, \quad (4.3.45)$$

$$\Gamma_{j0}^i = [\mathcal{H} - \Phi']\delta_j^i, \quad (4.3.46)$$

$$\Gamma_{jk}^i = -2\delta_{(j}^i \partial_{k)}\Phi + \delta_{jk}\delta^{il}\partial_l\Phi. \quad (4.3.47)$$

4.3.1 Conservation Equations

The equations of motion for the matter perturbations follow from conservation of the stress tensor, $\nabla^\mu T_{\mu\nu} = 0$. If there is no energy and momentum transfer between the different components, then the species are separately conserved and we also have $\nabla^\mu T_{\mu\nu}^{(a)} = 0$. Equipped with the perturbed connection, we can derive the perturbed conservation equations from

$$\begin{aligned} \nabla_\mu T^\mu{}_\nu &= 0 \\ &= \partial_\mu T^\mu{}_\nu + \Gamma_{\mu\alpha}^\mu T^\alpha{}_\nu - \Gamma_{\mu\nu}^\alpha T^\mu{}_\alpha. \end{aligned} \quad (4.3.48)$$

- *Continuity equation.*—For $\nu = 0$, we find

$$\partial_\eta \delta\rho = -3\mathcal{H}(\delta\rho + \delta P) + 3\Phi'(\bar{\rho} + \bar{P}) - \partial_i q^i, \quad (4.3.49)$$

This is the continuity equation describing 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

$\bar{\rho}' = -3\mathcal{H}(\bar{\rho} + \bar{P})$], the $\partial_i q^i$ term accounts for the local fluid flow due to peculiar velocity, and the $\dot{\Phi}$ 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].

- *Euler equation.*—For $\nu = i$, we find

$$\partial_\eta q^i = -4\mathcal{H}q^i - (\bar{\rho} + \bar{P})\partial^i\Psi - \partial^i\delta P - \partial_j\Pi^{ij}. \quad (4.3.50)$$

This is the Euler equation for a viscous fluid, i.e. the “ $F = ma$ ” of the fluid. In §1.2, we showed that peculiar velocities decay as a^{-1} . We therefore expected the momentum density to scale as $q \propto a^{-4}$. This explains the first term on the rhs of (4.3.50). The remaining terms are force terms.

Derivation.*—Consider first the $\nu = 0$ component of (4.3.48),

$$\partial_0 T^0{}_0 + \partial_i T^i{}_0 + \Gamma_{\mu 0}^\mu T^0{}_0 + \underbrace{\Gamma_{\mu i}^\mu T^i{}_0}_{\mathcal{O}(2)} - \Gamma_{00}^0 T^0{}_0 - \underbrace{\Gamma_{i0}^0 T^i{}_0}_{\mathcal{O}(2)} - \underbrace{\Gamma_{00}^i T^0{}_i}_{\mathcal{O}(2)} - \Gamma_{j0}^i T^j{}_i = 0. \quad (4.3.51)$$

Substituting the perturbed stress-energy tensor and the connection coefficients gives

$$\begin{aligned} & \partial_0(\bar{\rho} + \delta\rho) + \partial_i q^i + (\mathcal{H} + \Psi' + 3\mathcal{H} - 3\Phi')(\bar{\rho} + \delta\rho) \\ & - (\mathcal{H} + \Psi')(\bar{\rho} + \delta\rho) - (\mathcal{H} - \Phi')\delta_j^i [-(\bar{P} + \delta P)\delta_i^j] = 0, \end{aligned} \quad (4.3.52)$$

and hence

$$\bar{\rho}' + \partial_0\delta\rho + \partial_i q^i + 3\mathcal{H}(\bar{\rho} + \delta\rho) - 3\bar{\rho}\Phi' + 3\mathcal{H}(\bar{P} + \delta P) - 3\bar{P}\Phi' = 0. \quad (4.3.53)$$

Writing the zeroth-order and first-order parts separately, we get

$$\bar{\rho}' = -3\mathcal{H}(\bar{\rho} + \bar{P}), \quad (4.3.54)$$

$$\partial_\eta\delta\rho = -3\mathcal{H}(\delta\rho + \delta P) + 3\Phi'(\bar{\rho} + \bar{P}) - \nabla \cdot \mathbf{q}. \quad (4.3.55)$$

The zeroth-order equation (4.3.54) is simply the conservation of energy in the homogeneous background. The first-order equation (4.3.55) is the continuity equation for the density perturbation $\delta\rho$.

Next, consider the $\nu = i$ component of (4.3.48), $\partial_\mu T^\mu{}_i + \Gamma_{\mu\rho}^\mu T^\rho{}_i - \Gamma^\rho{}_{\mu i} T^\mu{}_\rho = 0$, and hence

$$\partial_0 T^0{}_i + \partial_j T^j{}_i + \Gamma_{\mu 0}^\mu T^0{}_i + \Gamma_{\mu j}^\mu T^j{}_i - \Gamma_{0i}^0 T^0{}_0 - \Gamma_{ji}^0 T^j{}_0 - \Gamma_{0i}^j T^0{}_j - \Gamma_{ki}^j T^k{}_j = 0. \quad (4.3.56)$$

Substituting the perturbed stress-energy tensor [with $T^0{}_i = -q_i$] and the connection coefficients gives

$$\begin{aligned} & -\partial_0 q_i + \partial_j [-(\bar{P} + \delta P)\delta_i^j - \Pi^j{}_i] - 4\mathcal{H}q_i - (\partial_j\Psi - 3\partial_j\Phi)\bar{P}\delta_i^j - \partial_i\Psi\bar{\rho} \\ & - \mathcal{H}\delta_{ji}q^j + \mathcal{H}\delta_i^j q_j + \underbrace{(-2\delta_{(i}^j\partial_{k)}\Phi + \delta_{ki}\delta^{jl}\partial_l\Phi)\bar{P}\delta_j^k}_{-3\partial_i\Phi\bar{P}} = 0. \end{aligned} \quad (4.3.57)$$

Cleaning this up, we find

$$\partial_\eta q_i = -4\mathcal{H}q_i - (\bar{\rho} + \bar{P})\partial_i\Psi - \partial_i\delta P - \partial^j\Pi_{ij}, \quad (4.3.58)$$

which confirms the form of the Euler equation (4.3.50).

It is instructive to evaluate (4.3.49) and (4.3.50) for a few special cases:

- *Matter*.—For a non-relativistic fluid (i.e. matter), we have $P_m = 0$ and $\Pi_m^{ij} = 0$. The continuity and Euler equations then simplify considerably. Writing $\delta_m \equiv \delta\rho_m/\rho_m$, we obtain

$$\delta'_m = -\nabla \cdot \mathbf{v}_m + 3\Phi', \quad (4.3.59)$$

$$\mathbf{v}'_m = -\mathcal{H}\mathbf{v}_m - \nabla\Psi. \quad (4.3.60)$$

Each term in these equations should be rather intuitive. Combining the time derivative of (4.3.59) with the divergence of (4.3.60), we find

$$\delta''_m + \mathcal{H}\delta'_m = \nabla^2\Psi + 3(\Phi'' + \mathcal{H}\Phi'). \quad (4.3.61)$$

↑ ↑
friction gravity

In Chapter 5, we will apply this equation to the clustering of dark matter perturbations.

- *Radiation*.—For a relativistic fluid (i.e. radiation), we have $P_r = \frac{1}{3}\rho_r$ and $\Pi_r^{ij} = 0$. The continuity and Euler equations become

$$\delta'_r = -\frac{4}{3}\nabla \cdot \mathbf{v}_r + 4\Phi', \quad (4.3.62)$$

$$\mathbf{v}'_r = -\frac{1}{4}\nabla\delta_r - \nabla\Psi. \quad (4.3.63)$$

Combining the time derivative of (4.3.62) with the divergence of (4.3.63), we get

$$\delta''_r - \frac{1}{3}\nabla^2\delta_r = \frac{4}{3}\nabla^2\Psi + 4\Phi''. \quad (4.3.64)$$

↑ ↑
pressure gravity

In Chapter 5, we will show how this equation leads to the oscillations in the observed spectrum of CMB anisotropies.

Exercise.—Show that the most general forms of the continuity and Euler equations are

$$\delta'_a = -\left(1 + \frac{\bar{P}_a}{\bar{\rho}_a}\right)(\partial_i v_a^i - 3\Phi') - 3\mathcal{H}\left(\frac{\delta P_a}{\bar{\rho}_a} - \frac{\bar{P}_a}{\bar{\rho}_a}\delta_a\right), \quad (4.3.65)$$

$$v_a^i' = -\left(\mathcal{H} + \frac{\bar{P}_a'}{\bar{\rho}_a + \bar{P}_a}\right)v_a^i - \frac{1}{\bar{\rho}_a + \bar{P}_a}(\partial^i\delta P_a - \partial_j\Pi_a^{ij}) - \partial^i\Psi. \quad (4.3.66)$$

Confirm that these expressions reduce to the equations for matter and radiation in the appropriate limits.

Comments.—The *two* equations (4.3.65) and (4.3.66) aren't sufficient to completely describe the evolution of the *four* perturbations $(\delta_a, \delta P_a, v_a, \Pi_a)$. To make progress, we either must make further simplifying assumptions or find additional evolution equations. We will do both.

- A perfect fluid is characterized by strong interactions which keep the pressure isotropic, $\Pi_a = 0$. In addition, pressure perturbations satisfy $\delta P_a = c_{s,a}^2 \delta \rho_a$, where $c_{s,a}$ is the adiabatic sound speed of the fluid. The perturbations of a perfect fluid are therefore described by only two independent variables, say δ_a and v_a , and the continuity and Euler equations are sufficient for closing the system.
- Decoupled or weakly interacting species (e.g. neutrinos) cannot be described by a perfect fluid and the above simplifications for the anisotropic stress and the pressure perturbation do not apply. In that case, we can't avoid solving the Boltzmann equation for the evolution of the perturbed distribution function f_a .
- Decoupled cold dark matter is a peculiar case. It is collisionless and has a negligible velocity dispersion. It therefore behaves like a pressureless perfect fluid although it has no interactions and therefore really isn't a fluid.

The different matter components are gravitationally coupled through the metric fluctuations in the continuity and Euler equations. The dynamics to the perturbed spacetime is determined, via the Einstein equations, by the perturbations of the *total* stress-energy tensor.

4.3.2 Einstein Equations

Let us compute the linearised Einstein equation in Newtonian gauge. This is conceptually straightforward, although algebraically a bit tedious.¹ 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 Ricci scalar R .

Ricci tensor.—We recall that the Ricci tensor can be expressed in terms of the connection as

$$R_{\mu\nu} = \partial_\lambda \Gamma_{\mu\nu}^\lambda - \partial_\nu \Gamma_{\mu\lambda}^\lambda + \Gamma_{\lambda\rho}^\lambda \Gamma_{\mu\nu}^\rho - \Gamma_{\mu\lambda}^\rho \Gamma_{\nu\rho}^\lambda. \quad (4.3.67)$$

Substituting the perturbed connection coefficients (4.3.42)–(4.3.47), we find

$$R_{00} = -3\mathcal{H}' + \nabla^2\Psi + 3\mathcal{H}(\Phi' + \Psi') + 3\Phi'', \quad (4.3.68)$$

$$R_{0i} = 2\partial_i(\Phi' + \mathcal{H}\Psi), \quad (4.3.69)$$

$$\begin{aligned} R_{ij} = & [\mathcal{H}' + 2\mathcal{H}^2 - \Phi'' + \nabla^2\Phi - 2(\mathcal{H}' + 2\mathcal{H}^2)(\Phi + \Psi) - \mathcal{H}\Psi' - 5\mathcal{H}\Phi']\delta_{ij} \\ & + \partial_i\partial_j(\Phi - \Psi). \end{aligned} \quad (4.3.70)$$

I will derive R_{00} explicitly and leave the other components as an exercise.

Example.—The 00 component of the Ricci tensor is

$$R_{00} = \partial_\rho \Gamma_{00}^\rho - \partial_0 \Gamma_{0\rho}^\rho + \Gamma_{00}^\alpha \Gamma_{\alpha\rho}^\rho - \Gamma_{0\rho}^\alpha \Gamma_{0\alpha}^\rho. \quad (4.3.71)$$

¹Once in your life you should do this computation by hand. After that you can use Mathematica: an example notebook can be downloaded [here](#).

The terms with $\rho = 0$ cancel in the sum over ρ , so we only need to consider summing over $\rho = i$,

$$\begin{aligned} R_{00} &= \partial_i \Gamma_{00}^i - \partial_0 \Gamma_{0i}^i + \Gamma_{00}^\alpha \Gamma_{\alpha i}^i - \Gamma_{0i}^\alpha \Gamma_{0\alpha}^i \\ &= \partial_i \Gamma_{00}^i - \partial_0 \Gamma_{0i}^i + \underbrace{\Gamma_{00}^0 \Gamma_{0i}^i}_{\mathcal{O}(2)} + \underbrace{\Gamma_{00}^j \Gamma_{ji}^i}_{\mathcal{O}(2)} - \underbrace{\Gamma_{0i}^0 \Gamma_{00}^i}_{\mathcal{O}(2)} - \Gamma_{0i}^j \Gamma_{0j}^i \\ &= \nabla^2 \Psi - 3\partial_0(\mathcal{H} - \Phi') + 3(\mathcal{H} + \Psi')(\mathcal{H} - \Phi') - (\mathcal{H} - \Phi')^2 \delta_i^j \delta_j^i \\ &= -3\mathcal{H}' + \nabla^2 \Psi + 3\mathcal{H}(\Phi' + \Psi') + 3\Phi''. \end{aligned} \quad (4.3.72)$$

This confirms the result (4.3.68).

Exercise.—Derive eqs. (4.3.69) and (4.3.70).

Ricci scalar.—It is now relatively straightforward to compute the Ricci scalar

$$R = g^{00} R_{00} + 2 \underbrace{g^{0i} R_{0i}}_{\mathcal{O}(2)} + g^{ij} R_{ij}. \quad (4.3.73)$$

It follows that

$$\begin{aligned} a^2 R &= (1 - 2\Psi) R_{00} - (1 + 2\Phi) \delta^{ij} R_{ij} \\ &= (1 - 2\Psi) [-3\mathcal{H}' + \nabla^2 \Psi + 3\mathcal{H}(\Phi' + \Psi') + 3\Phi''] \\ &\quad - 3(1 + 2\Phi) [\mathcal{H}' + 2\mathcal{H}^2 - \Phi'' + \nabla^2 \Phi - 2(\mathcal{H}' + 2\mathcal{H}^2)(\Phi + \Psi) - \mathcal{H}\Psi' - 5\mathcal{H}\Phi'] \\ &\quad - (1 + 2\Phi) \nabla^2(\Phi - \Psi). \end{aligned} \quad (4.3.74)$$

Dropping non-linear terms, we find

$$a^2 R = -6(\mathcal{H}' + \mathcal{H}^2) + 2\nabla^2 \Psi - 4\nabla^2 \Phi + 12(\mathcal{H}' + \mathcal{H}^2)\Psi + 6\Phi'' + 6\mathcal{H}(\Psi' + 3\Phi'). \quad (4.3.75)$$

Einstein equations.—We have done all the work to compute the Einstein equation

$$G^\mu_\nu = 8\pi G T^\mu_\nu. \quad (4.3.76)$$

We chose to work with one index raised since that simplifies the form of the stress tensor [see §4.2]. We will first consider the time-time component. The relevant component of the Einstein tensor is

$$\begin{aligned} G^0_0 &= g^{00} \left[R_{00} - \frac{1}{2} g_{00} R \right] \\ &= a^{-2} (1 - 2\Psi) R_{00} - \frac{1}{2} R, \end{aligned} \quad (4.3.77)$$

where we have used that g^{0i} vanishes in Newtonian gauge. Substituting (4.3.68) and (4.3.75), and cleaning up the resulting mess, we find

$$\delta G^0_0 = a^{-2} (2\nabla^2 \Phi - 6\mathcal{H}(\Phi' + \mathcal{H}\Psi')). \quad (4.3.78)$$

The 00-Einstein equation therefore is

$$\boxed{\nabla^2 \Phi - 3\mathcal{H}(\Phi' + \mathcal{H}\Psi) = 4\pi G a^2 \delta\rho}, \quad (4.3.79)$$

81 4. Cosmological Perturbation Theory

where $\delta\rho \equiv \sum_a \delta\rho_a$ is the *total* density perturbation. Equation (4.3.79) is the relativistic generalization of the *Poisson equation*. Inside the Hubble radius, i.e. for Fourier modes with $k \gg \mathcal{H}$, we have $|\nabla^2\Phi| \gg 3\mathcal{H}|\dot{\Phi}| + \mathcal{H}\Psi|$, so that eq. (4.3.79) reduces to $\nabla^2\Phi \approx 4\pi Ga^2\delta\rho$. This is the Poisson equation in the Newtonian limit. The GR corrections in (4.3.79) will be important on scales comparable to the Hubble radius, i.e. for $k \lesssim \mathcal{H}$.

Next, we consider the spatial part of the Einstein equation. The relevant component of the Einstein tensor is

$$\begin{aligned} G^i_j &= g^{ik} \left[R_{kj} - \frac{1}{2} g_{kj} R \right] \\ &= -a^{-2}(1+2\Phi)\delta^{ik}R_{kj} - \frac{1}{2}\delta^i_j R. \end{aligned} \quad (4.3.80)$$

From eq. (4.3.70), we see that most terms in R_{kj} are proportional to δ_{kj} . When contracted with δ^{ik} this leads to a myriad of terms proportional to δ^i_j . We don't want to deal with this mess. Instead we focus on the tracefree part of G^i_j . We can extract this piece by contracting G^i_j with the projection tensor $P^j_i \equiv \partial^j\partial_i - \frac{1}{3}\delta^j_i\nabla^2$. Using (4.3.70), this gives

$$P^j_i G^i_j = -\frac{2}{3}a^{-2}\nabla^4(\Phi - \Psi). \quad (4.3.81)$$

This should be equated to the tracefree part of the stress tensor, which for scalar fluctuations is

$$P^j_i T^i_j = -P^j_i \Pi^i_j = -\frac{2}{3}\nabla^4\Pi. \quad (4.3.82)$$

Setting (4.3.81) and (4.3.82) equal, we get

$$\boxed{\Phi - \Psi = 8\pi Ga^2\Pi}, \quad (4.3.83)$$

where $\Pi \equiv \sum_a \Pi_a$. Dark matter and baryons can be described as perfect fluids and therefore don't contribute to the anisotropic stress in (4.3.83). Photons only start to develop an anisotropic stress component during the matter-dominated era when their energy density is subdominant. The only relevant source in (4.3.83) are therefore free-streaming neutrinos. However, their effect is also relatively small, so to the level of accuracy that we aspire to in these lectures they can be ignored. Equation (4.3.83) then implies $\Psi \approx \Phi$.

Exercise.—By considering the time-space Einstein equation, show that

$$\boxed{\Phi' + \mathcal{H}\Phi = -4\pi Ga^2q}. \quad (4.3.84)$$

With this, the Poisson equation (4.3.79) can be written as

$$\boxed{\nabla^2\Phi = 4\pi Ga^2\bar{\rho}\Delta}, \quad (4.3.85)$$

where Δ is the comoving-gauge density contrast (4.2.38).

By considering the trace of the space-space Einstein equation, we can derive the following evolution equation for the metric potential

$$\boxed{\Phi'' + 3\mathcal{H}\Phi' + (2\mathcal{H}' + \mathcal{H}^2)\Phi = 4\pi Ga^2\delta P}, \quad (4.3.86)$$

where δP is the total pressure perturbation. If we hadn't ignored the anisotropic stress, then (4.3.86) would have an additional source term.

4.3.3 Metric Evolution

It is interesting to consider the evolution of Φ during both the radiation-dominated era and the matter-dominated era:

- During the matter era, we have

$$2\mathcal{H}' + \mathcal{H}^2 = -8\pi G a^2 \bar{P} = 0, \quad (4.3.87)$$

so that eq. (4.3.86) reduces to

$$\Phi'' + 3\mathcal{H}\Phi' = 0, \quad (4.3.88)$$

where $\mathcal{H} = 2/\eta$. This equation has two solutions: $\Phi = \text{const.}$ and $\Phi \propto \eta^{-5} \propto a^{-5/2}$. Notice that the growing mode solution is $\Phi = \text{const.}$ independent of the wavelength of the fluctuation.

- During the radiation era, we have

$$2\mathcal{H}' + \mathcal{H}^2 = -\frac{8\pi G}{3}a^2 \bar{\rho} = -\mathcal{H}^2, \quad (4.3.89)$$

$$4\pi G a^2 \delta P = \frac{8\pi G}{3}a^2 \delta \rho = \frac{1}{3}(\nabla^2 \Phi - 3\mathcal{H}(\Phi' + \mathcal{H}\Phi)), \quad (4.3.90)$$

so that eq. (4.3.86) becomes

$$\Phi'' + 4\mathcal{H}\Phi' = \frac{1}{3}\nabla^2 \Phi, \quad (4.3.91)$$

where $\mathcal{H} = 1/\eta$. Substituting the Fourier expansion of the field,

$$\Phi(\eta, \mathbf{x}) \equiv \int \frac{d^3 k}{(2\pi)^{3/2}} \Phi_{\mathbf{k}}(\eta) e^{i\mathbf{k}\cdot\mathbf{x}}, \quad (4.3.92)$$

we get

$$\Phi_{\mathbf{k}}'' + \frac{4}{\eta}\Phi_{\mathbf{k}}' + \frac{1}{3}k^2\Phi_{\mathbf{k}} = 0, \quad (4.3.93)$$

This equation has the following exact solution

$$\Phi_{\mathbf{k}}(\tau) = A_{\mathbf{k}} \frac{j_1(y)}{y} + B_{\mathbf{k}} \frac{n_1(y)}{y}, \quad y \equiv \frac{1}{\sqrt{3}}k\eta, \quad (4.3.94)$$

where the subscript \mathbf{k} indicates that the solution can have different amplitudes for each value of \mathbf{k} . The size of the initial fluctuations as a function of wavenumber will be a prediction of inflation. The functions $j_1(y)$ and $n_1(y)$ in (4.3.94) are the spherical Bessel and Neumann functions

$$j_1(y) = \frac{\sin y}{y^2} - \frac{\cos y}{y} = \frac{y}{3} + \mathcal{O}(y^3), \quad (4.3.95)$$

$$n_1(y) = -\frac{\cos y}{y^2} - \frac{\sin y}{y} = -\frac{1}{y^2} + \mathcal{O}(y^0). \quad (4.3.96)$$

Since $n_1(y)$ blows up for small y (early times), we reject that solution on the basis of initial conditions, i.e. we set $B_{\mathbf{k}} \equiv 0$. We match the constant $A_{\mathbf{k}}$ to the primordial value of the potential, $\Phi_{\mathbf{k}}(0) = -\frac{2}{3}\mathcal{R}_{\mathbf{k}}(0)$, see eq. (4.4.110) below. Using (4.3.95), we find

$$\Phi_{\mathbf{k}}(\eta) = -2\mathcal{R}_{\mathbf{k}}(0) \left(\frac{\sin y - y \cos y}{y^3} \right). \quad (4.3.97)$$

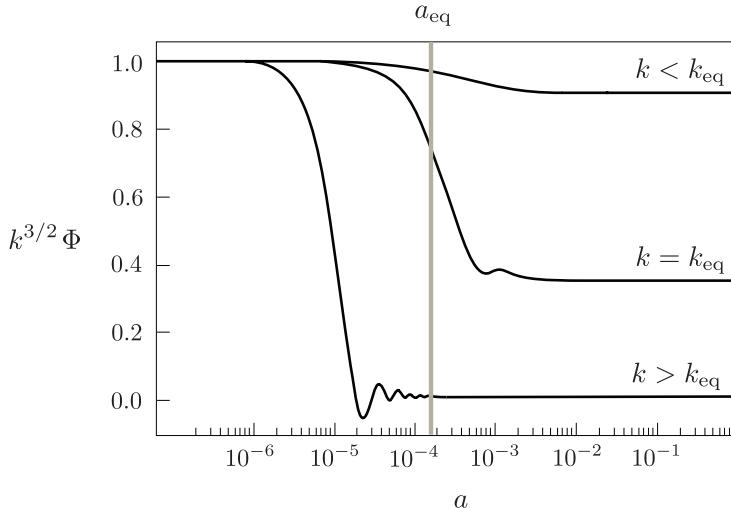


Figure 4.1: Numerical solutions for the linear evolution of the gravitational potential.

Notice that (4.3.97) is valid on all scales. Outside the (sound) horizon, $y = \frac{1}{\sqrt{3}}k\eta \ll 1$, the solution approaches $\Phi = \text{const.}$, while on subhorizon scales, $y \gg 1$, we get

$$\Phi_k(\eta) \approx -6\mathcal{R}_k(0) \frac{\cos(\frac{1}{\sqrt{3}}k\eta)}{(k\eta)^2} \quad (\text{subhorizon}) . \quad (4.3.98)$$

During the radiation era, subhorizon modes of Φ therefore oscillate with frequency $\frac{1}{\sqrt{3}}k$ and an amplitude that decays as $\eta^{-2} \propto a^{-2}$. Remember this.

Figure 4.1 shows the evolution of the gravitational potential for three representative 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 9/10 coming from the radiation-to-matter transition (see §4.4.2).

4.4 Initial Conditions

At sufficiently early times, all scales of interest to current observations were outside the Hubble radius. On super-Hubble scales, the evolution of perturbations becomes very simple, especially for adiabatic initial conditions.

4.4.1 Adiabatic Fluctuations

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 (η, \mathbf{x}) of the perturbed universe is the same as in the *background* universe at some slightly different time $\eta + \delta\eta(\mathbf{x})$. (Notice that the time shift varies with location \mathbf{x} .) We can therefore think of adiabatic perturbations as arising because some parts of the universe are “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

$$\delta\rho_a(\eta, \mathbf{x}) \equiv \bar{\rho}_a(\eta + \delta\eta(\mathbf{x})) - \bar{\rho}_a(\eta) = \bar{\rho}'_a \delta\eta(\mathbf{x}), \quad (4.4.99)$$

where $\delta\eta$ is the same for all species a . This implies

$$\delta\eta = \frac{\delta\rho_a}{\bar{\rho}'_a} = \frac{\delta\rho_b}{\bar{\rho}'_b} \quad \text{for all species } a \text{ and } b. \quad (4.4.100)$$

Using $\bar{\rho}'_a = -3\mathcal{H}(1+w_a)\bar{\rho}_a$, we can write this as

$$\frac{\delta_a}{1+w_a} = \frac{\delta_b}{1+w_b} \quad \text{for all species } a \text{ and } b. \quad (4.4.101)$$

Thus, for adiabatic perturbations, all matter components ($w_m \approx 0$) have the same fractional perturbations, while all radiation perturbations ($w_r = \frac{1}{3}$) obey

$$\delta_r = \frac{4}{3}\delta_m. \quad (4.4.102)$$

It follows that, for adiabatic fluctuations, the total density perturbation, $\delta\rho \equiv \sum_a \bar{\rho}_a \delta_a$, is dominated by the species that carries the dominant energy density $\bar{\rho}_a$, since all the δ_a 's are comparable. At early times, the universe is radiation dominated, so it is natural to set the initial conditions for all super-Hubble Fourier modes then. Equation (4.3.86) implies that $\Phi = \text{const.}$ on super-Hubble scales, while equation (4.3.79) leads to

$$\delta \approx \delta_r = -2\Phi = \text{const.} \quad (4.4.103)$$

Equations (4.4.103) and (4.4.102) show that, for adiabatic initial conditions, all matter perturbations are given in terms of the super-Hubble value of the potential Φ . In these lectures, we will be concerned with the evolution of photons, baryons and cold dark matter (CDM). Their fractional density perturbations will satisfy the relation (4.4.102) on super-Hubble scales, but will start to evolve in distinct ways inside of the horizon.

4.4.2 Curvature Perturbations

The gravitational potential Φ is only constant on super-Hubble scales if the equation of state of the background is constant. Whenever the equation of state evolves (e.g. in the transitions from inflation to radiation domination or from radiation to matter domination), so will the gravitational potential (cf. Fig. 4.1). It will be convenient to identify an alternative perturbation variable that stays constant on large scales even in these more general situations. Such a variable is the *comoving curvature perturbation*:

$$\mathcal{R} = -\Phi + \frac{\mathcal{H}}{\bar{\rho} + \bar{P}} \delta q, \quad (4.4.104)$$

where $T^0_j \equiv -\partial_j \delta q$. Defining the initial conditions in terms of \mathcal{R} will allow us to match the predictions made by inflation to the fluctuations in the primordial plasma most easily.

*Proof.**—The following is a proof that \mathcal{R} is conserved on super-Hubble scales, i.e. for modes with $k \ll \mathcal{H}$. First, it is useful to note that on large scales $k \ll \mathcal{H}$, the Einstein equations imply $\mathcal{H}\delta q = -\frac{1}{3}\delta\rho$. In this limit, the curvature perturbation can be written as

$$\mathcal{R} \xrightarrow{k \ll \mathcal{H}} -\Phi - \frac{\delta\rho}{3(\bar{\rho} + \bar{P})}. \quad (4.4.105)$$

To find the evolution of \mathcal{R} , we consider the continuity equation [eq. (4.3.49)]:

$$\frac{\partial\delta\rho}{\partial\eta} + 3\mathcal{H}(\delta\rho + \delta P) + \partial_i q^i = 3(\bar{\rho} + \bar{P})\frac{\partial\Phi}{\partial\eta}, \quad (4.4.106)$$

On large scales, $\partial_i q^i$ is of order k^2 and can be dropped relative to terms of order \mathcal{H}^2 . Solving (4.4.105) for Φ , and substituting it into (4.4.106), we get

$$\frac{\partial\delta\rho}{\partial\eta} + 3\mathcal{H}(\delta\rho + \delta P) = -3(\bar{\rho} + \bar{P})\frac{\partial\mathcal{R}}{\partial\eta} + (\bar{\rho} + \bar{P})\frac{\partial}{\partial\eta}\left(\frac{\delta\rho}{\bar{\rho} + \bar{P}}\right). \quad (4.4.107)$$

The time derivatives of $\delta\rho$ cancel on both sides and we are left with

$$3(\bar{\rho} + \bar{P})\frac{\partial\mathcal{R}}{\partial\eta} = -3\mathcal{H}(\delta\rho + \delta P) - \frac{\bar{\rho}' + \bar{P}'}{(\bar{\rho} + \bar{P})}\delta\rho. \quad (4.4.108)$$

Using $\bar{\rho}' = -3\mathcal{H}(\bar{\rho} + \bar{P})$, this becomes

$$(\bar{\rho} + \bar{P})\frac{\mathcal{R}'}{\mathcal{H}} = -\left(\delta P - \frac{\bar{P}'}{\bar{\rho}'}\delta\rho\right). \quad (4.4.109)$$

For adiabatic perturbations, the right-hand side vanishes and we have established the conservation of the comoving curvature perturbation, $\mathcal{R}' \xrightarrow{k \ll \mathcal{H}} 0$.

Exercise.—Substituting (4.3.84) into (4.4.104) gives

$$\mathcal{R} = -\Phi - \frac{\mathcal{H}(\Phi' + \mathcal{H}\Phi)}{4\pi G a^2(\bar{\rho} + \bar{P})} \xrightarrow{k \ll \mathcal{H}} -\frac{5+3w}{3+3w}\Phi. \quad (4.4.110)$$

Use this to show that the amplitude of super-Hubble modes of Φ drops by a factor of 9/10 in the radiation-to-matter transition.

Another advantage of the variable \mathcal{R} is that it is *gauge-invariant*, i.e. its value does not depend on the choice of coordinates. Although (4.4.105) was written in terms of variables defined in Newtonian gauge, it applies in an arbitrary gauge if we write $g_{ij} = -a^2[(1-2\Phi)\delta_{ij} + \partial_i \partial_j E]$ for the metric and $T^0{}_j = \partial_j \delta q$ for the momentum density. While we will mostly stick to *Newtonian gauge*, the computation of inflationary fluctuations in Chapter 6 turns out to be simplest in *spatially flat gauge* [$\Phi = E = 0$]. The curvature perturbation \mathcal{R} provides the “bridge” between results obtained in Chapter 6 and the analysis in the rest of the notes.

4.4.3 Statistics

Quantum mechanics during inflation only predicts the statistics of the initial conditions, i.e. it predicts the correlation between the CMB fluctuations in different directions, rather than the specific value of the temperature fluctuation in a specific direction. For Gaussian initial condi-

tions, these correlations are completely specified by the two-point correlation function

$$\langle \mathcal{R}(\mathbf{x})\mathcal{R}(\mathbf{x}') \rangle \equiv \xi_{\mathcal{R}}(\mathbf{x}, \mathbf{x}') = \xi_{\mathcal{R}}(|\mathbf{x}' - \mathbf{x}|), \quad (4.4.111)$$

where the last equality holds as a consequence of statistical homogeneity and isotropy. The Fourier transform of \mathcal{R} then satisfies

$$\langle \mathcal{R}(\mathbf{k})\mathcal{R}^*(\mathbf{k}') \rangle = \frac{2\pi^2}{k^3} \Delta_{\mathcal{R}}^2(k) \delta_D(\mathbf{k} - \mathbf{k}'), \quad (4.4.112)$$

where $\Delta_{\mathcal{R}}^2(k)$ is the (dimensionless) *power spectrum*.

Exercise.—Show that

$$\xi_{\mathcal{R}}(\mathbf{x}, \mathbf{x}') = \int \frac{dk}{k} \Delta_{\mathcal{R}}^2(k) \text{sinc}(k|\mathbf{x} - \mathbf{x}'|). \quad (4.4.113)$$

In Chapter 6, we will compute the form of $\Delta_{\mathcal{R}}^2(k)$ predicted by inflation. We will find that the spectrum takes a power law form

$$\Delta_{\mathcal{R}}^2(k) = A_s \left(\frac{k}{k_*} \right)^{n_s - 1}. \quad (4.4.114)$$

This agrees with observational constraints if $A_s = 2 \times 10^{-9}$ and $n_s = 0.96$, for $k_* = 0.05 \text{ Mpc}^{-1}$.

4.5 Summary

We have derived the linearised evolution equations for scalar perturbations in Newtonian gauge, where the metric has the following form

$$ds^2 = a^2(\eta) [(1 + 2\Psi)d\eta^2 - (1 - 2\Phi)\delta_{ij}dx^i dx^j]. \quad (4.5.115)$$

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

$$\nabla^2\Phi - 3\mathcal{H}(\Phi' + \mathcal{H}\Phi) = 4\pi G a^2 \delta\rho, \quad (4.5.116)$$

$$\Phi' + \mathcal{H}\Phi = -4\pi G a^2 (\bar{\rho} + \bar{P})v, \quad (4.5.117)$$

$$\Phi'' + 3\mathcal{H}\Phi' + (2\mathcal{H}' + \mathcal{H}^2)\Phi = 4\pi G a^2 \delta P. \quad (4.5.118)$$

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

$$\nabla^2\Phi = 4\pi G a^2 \bar{\rho} \Delta. \quad (4.5.119)$$

- From the conservation of the stress-tensor, we derived the relativistic generalisations of the continuity equation and the Euler equation

$$\delta' + 3\mathcal{H} \left(\frac{\delta P}{\delta\rho} - \frac{\bar{P}}{\bar{\rho}} \right) \delta = - \left(1 + \frac{\bar{P}}{\bar{\rho}} \right) (\nabla \cdot \mathbf{v} - 3\Phi'), \quad (4.5.120)$$

$$\mathbf{v}' + 3\mathcal{H} \left(\frac{1}{3} - \frac{\bar{P}'}{\bar{\rho}'} \right) \mathbf{v} = -\frac{\nabla \delta P}{\bar{\rho} + \bar{P}} - \nabla \Phi. \quad (4.5.121)$$

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

$$\mathcal{R} = -\Phi - \frac{\mathcal{H}(\Phi' + \mathcal{H}\Phi)}{4\pi G a^2(\bar{\rho} + \bar{P})}. \quad (4.5.122)$$

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.

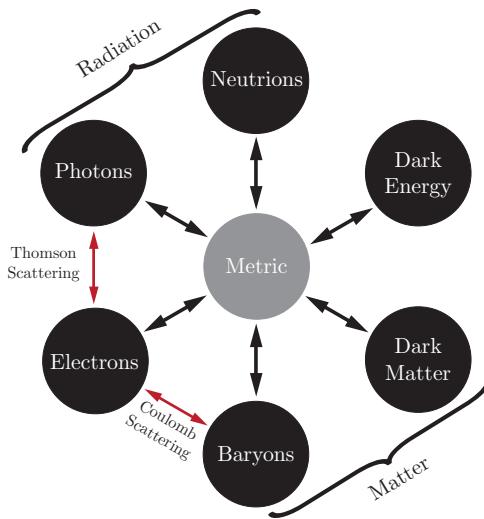


Figure 5.1: Interactions between the different forms of matter in the universe.

5.1 Gravitational Instability

5.1.1 Jeans' Instability

The growth of primordial density perturbations is determined by a competition between gravity and pressure. Gravity will attract matter into overdense regions of space. However, the random thermal motion of particles is also increased in overdensities. If this pressure is large then it will push matter out of dense regions and therefore inhibit the growth of inhomogeneities.

Consider inhomogeneities in a fluid with equation of state $w \equiv \bar{P}/\bar{\rho}$ and speed of sound $c_s^2 \equiv \delta P/\delta\rho$. For adiabatic perturbations, we have $c_s^2 \approx w$. The continuity equation (4.5.120) and the Euler equation (4.5.121) then are

$$\delta' = -(1+w)(\nabla \cdot \mathbf{v} - 3\Phi'), \quad (5.1.1)$$

$$\mathbf{v}' = -\mathcal{H}(1-3w)\mathbf{v} - \frac{c_s^2}{1+w}\nabla\delta - \nabla\Phi. \quad (5.1.2)$$

On subhorizon scales, time derivatives of Φ are subdominant and can be dropped.¹ Combining the time derivative of (5.1.1) with the divergence of (5.1.2), we then find

$$\delta'' + (1 - 3w)\mathcal{H}\delta' - c_s^2 \nabla^2 \delta = (1 + w)\nabla^2 \Phi. \quad (5.1.3)$$

\uparrow \uparrow \uparrow
friction pressure gravity

Using the Poisson equation (4.5.116) to replace $\nabla^2 \Phi$, this becomes

$$\boxed{\delta'' + (1 - 3w)\mathcal{H}\delta' + c_s^2(k_J^2 - k_J^2)\delta = 0}, \quad (5.1.4)$$

where we have defined the *Jeans' scale*,

$$k_J^2 \equiv (1 + w) \frac{4\pi G a^2 \bar{\rho}}{c_s^2}. \quad (5.1.5)$$

For small scales (i.e. large wavenumbers), $k > k_J$, the pressure dominates and the fluctuations oscillate, while on large scales, $k < k_J$, gravity dominates and the fluctuations grow. The Hubble friction 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 characteristic of a gravitational instability in static space.

In §5.1.2, we will study the clustering of dark matter fluctuations. Since the sound speed of CDM is small, the fluctuations grow efficiently over a large range of scales. In §5.2.1, we will look at the evolution of fluctuations in the primordial photon gas. Now, pressure plays an important role: fluctuations only grow on very large scales and oscillate otherwise. These acoustic oscillations are observed both in the microwave background and in the distribution of galaxies.

5.1.2 Clustering of Dark Matter

In this section, we are interested in the evolution of (dark) matter fluctuations from early times (during the radiation era) until late times (when dark energy starts to dominate).

- At early times, the universe was dominated by a mixture of radiation (r) and pressureless matter (m). For now, we ignore baryons. The conformal Hubble parameter is

$$\mathcal{H}^2 = \frac{\mathcal{H}_0^2 \Omega_m^2}{\Omega_r} \left(\frac{1}{y} + \frac{1}{y^2} \right), \quad y \equiv \frac{a}{a_{\text{eq}}}. \quad (5.1.6)$$

We wish to determine how matter fluctuations evolve on subhorizon scales from the radiation era until the matter era. In §4.3.1, we showed that the matter density contrast satisfies

$$\delta_m'' + \mathcal{H}\delta_m' = \nabla^2 \Phi + 3(\Phi'' + \mathcal{H}\Phi'). \quad (5.1.7)$$

In general, the potential Φ is sourced by the total density fluctuation. However, the perturbations in the radiation density oscillate rapidly on small scales (see §5.2). The *time-averaged* gravitational potential is therefore only sourced by the matter fluctuations,

¹This is actually a subtle point. During the radiation era, time derivatives of Φ are only subdominant if averaged over a Hubble time. We are here implicitly assuming such an averaging.

and the fluctuations in the radiation can be neglected (see Weinberg, astro-ph/0207375 for further discussion). Assuming such a time averaging, we have $\nabla^2\Phi \gg \{\Phi'', \mathcal{H}\Phi'\}$ on subhorizon scales ($k \gg \mathcal{H}$). With these simplifications, equation (5.1.7) becomes

$$\delta_m'' + \mathcal{H}\delta_m' - 4\pi G a^2 \bar{\rho}_m \delta_m \approx 0, \quad (5.1.8)$$

where \mathcal{H} given by (5.1.6). One can show that this equation can be written as the *Mészáros equation*

$$\boxed{\frac{d^2\delta_m}{dy^2} + \frac{2+3y}{2y(1+y)} \frac{d\delta_m}{dy} - \frac{3}{2y(1+y)} \delta_m = 0}. \quad (5.1.9)$$

You will also be asked to show that the solutions to this equation take the form

$$\delta_m \propto \begin{cases} 2+3y \\ (2+3y) \ln \left(\frac{\sqrt{1+y}+1}{\sqrt{1+y}-1} \right) - 6\sqrt{1+y} \end{cases} \quad (5.1.10)$$

In the limit $y \ll 1$ (RD), the growing mode solution is $\delta_m \propto \ln y \propto \ln a$, i.e. the matter fluctuations only grow logarithmically in the radiation era. Significant growth of dark matter inhomogeneities only occurs when the universe becomes matter dominated. Indeed, in the limit $y \gg 1$ (MD), the growing mode solution is $\delta_m \propto y \propto a$.

- At late times, the universe is a mixture of pressureless matter (m) and dark energy (Λ). Since dark energy doesn't have fluctuations, equation (5.1.8) still applies

$$\delta_m'' + \mathcal{H}\delta_m' - 4\pi G a^2 \bar{\rho}_m \delta_m \approx 0, \quad (5.1.11)$$

but \mathcal{H} is not the same as before. In the Λ -dominated regime, we have $\mathcal{H}^2 = (-\eta^{-1})^2 \gg 4\pi G a^2 \bar{\rho}_m$. Dropping the last term in (5.1.11), we get

$$\delta_m'' - \frac{1}{\eta} \delta_m' \approx 0, \quad (5.1.12)$$

which has the following solution

$$\delta_m \propto \begin{cases} \text{const.} \\ \eta^2 \propto a^{-2} \end{cases}. \quad (5.1.13)$$

We see that the matter fluctuations stop growing once dark energy comes to dominate.

*Comoving density contrast.**—An elegant way to obtain solutions that are valid on all scales (not just in the subhorizon limit) is to work with the comoving density contrast Δ_m . This has the nice feature that the Poisson equation for the time-average gravitational potential takes a simple form

$$\nabla^2\Phi = 4\pi G a^2 \bar{\rho}_m \Delta_m. \quad (5.1.14)$$

The solution for Δ_m can therefore be obtained directly from the solution for Φ (cf. §4.3.3). Let us see how this reproduces our previous results:

- During the matter era, we have $\Phi \propto \{a^0, a^{-5/2}\}$ and $a^2 \rho_m \propto a^{-1}$, so that

$$\Delta_m = \frac{\nabla^2 \Phi}{4\pi G a^2 \bar{\rho}_m} \propto \begin{cases} a & \\ a^{-3/2} & \end{cases} . \quad (5.1.15)$$

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 .

- Dark energy contributes pressure, but no pressure fluctuations. The Einstein equation (4.5.118) therefore is

$$\Phi'' + 3\mathcal{H}\Phi' + (2\mathcal{H}' + \mathcal{H}^2)\Phi = 0. \quad (5.1.16)$$

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. (5.1.16) implies

$$\partial_\eta^2(\Delta_m/a) + 3\mathcal{H}\partial_\eta(\Delta_m/a) + (2\mathcal{H}' + \mathcal{H}^2)(\Delta_m/a) = 0, \quad (5.1.17)$$

which rearranges to

$$\Delta_m'' + \mathcal{H}\Delta_m' + (\mathcal{H}' - \mathcal{H}^2)\Delta_m = 0. \quad (5.1.18)$$

Using $\mathcal{H}' - \mathcal{H}^2 = -4\pi G a^2 \bar{\rho}_m$, this becomes

$$\Delta_m'' + \mathcal{H}\Delta_m' - 4\pi G a^2 \bar{\rho}_m \Delta_m = 0. \quad (5.1.19)$$

This equation looks similar to (5.1.11), but is now valid on all scales. The solution in Λ -dominated regime is

$$\Delta_m \propto \begin{cases} \text{const.} & \\ \eta^2 \propto a^{-2} & \end{cases} . \quad (5.1.20)$$

This is the same as (5.1.13), but now valid on all scales.

5.1.3 Matter Power Spectrum

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}_k by

$$\Delta_{m,\mathbf{k}}(z) = T(k, z) \mathcal{R}_k. \quad (5.1.21)$$

The transfer function $T(k, z)$ depends only on the magnitude k and not on the direction of \mathbf{k} , because the perturbations are evolving on a homogeneous and isotropic background. The square of the Fourier mode (5.1.21) defines that matter *power spectrum*

$$P_\Delta(k, z) \equiv |\Delta_{m,\mathbf{k}}(z)|^2 = T^2(k, z) |\mathcal{R}_k|^2. \quad (5.1.22)$$

Figure 5.2 shows predicted matter power spectrum for scale-invariant initial conditions, $k^3 |\mathcal{R}_k|^2 = \text{const.}$ (see Chapter 6). The asymptotic scalings of the power spectrum are

$$P_\Delta(k) = \begin{cases} k & k < k_{\text{eq}} \\ k^{-3} & k > k_{\text{eq}} \end{cases}. \quad (5.1.23)$$

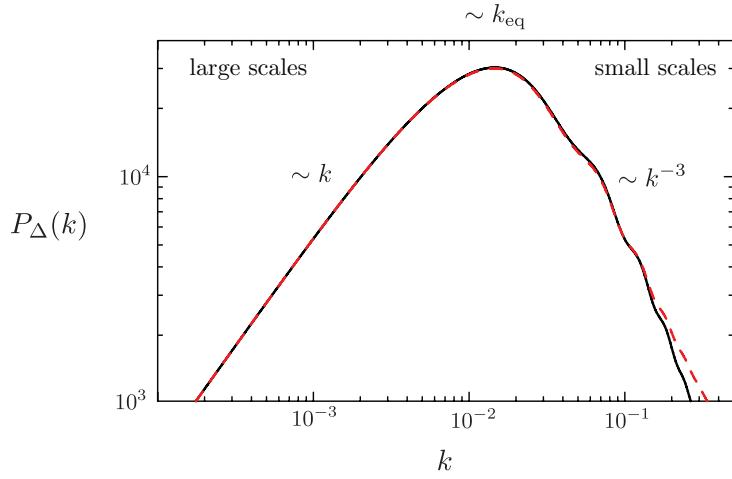


Figure 5.2: The matter power spectrum $P_\Delta(k)$ at $z = 0$ in linear theory (solid) and with non-linear corrections (dashed). On large scales, $P_\Delta(k)$ grows as k . The power spectrum turns over around $k_{\text{eq}} \sim 0.01 \text{ Mpc}^{-1}$ corresponding to the horizon size at matter-radiation equality. Beyond the peak, the power falls as k^{-3} . Visible are small amplitude baryon acoustic oscillations in the spectrum.

These scalings are easy to understand by consulting the Poisson equation, $\nabla^2 \Phi = 4\pi G a^2 \bar{\rho}_m \Delta_m$, which implies $P_\Delta(k) \propto k^4 P_\Phi(k)$. The spectrum of Δ therefore directly related to the time-evolved spectrum of Φ :

- Modes with $k < k_{\text{eq}}$ only enter the horizon during the matter era when their amplitude remains constant. For these modes, the gravitational potential hasn't undergone any evolution and its power spectrum takes the scale-invariant form $P_\Phi(k) \propto k^{-3}$. The matter power spectrum for those scales therefore is

$$P_\Delta(k) \propto k^4 \times k^{-3} = k. \quad (5.1.24)$$

- Modes with $k > k_{\text{eq}}$, on the other hand, have entered the horizon during the radiation era. As we have seen in §4.3.3, for these modes, the gravitational potential decays as $a^{-2} \propto \eta^{-2}$. The amount of suppression of Φ is determined by the amount time that the mode has spent inside the horizon which follows from the horizon crossing condition $k\eta = 1$. Modes with larger k enter the horizon earlier and will be more suppressed, namely by a factor of $(k/k_{\text{eq}})^{-2}$ coming from the $(\eta/\eta_{\text{eq}})^{-2}$ suppression of the potential. The matter power spectrum for $k > k_{\text{eq}}$ therefore is

$$P_\Delta(k) \propto k^4 \times k^{-3} \times (k^{-2})^2 = k^{-3}. \quad (5.1.25)$$

Finally, note that the evolution of $P_\Delta(k, z)$ during the matter era is proportional to $a^2 = (1+z)^{-2}$, cf. eq. (5.1.15).

5.2 Acoustic Oscillations

5.2.1 Radiation Fluctuations

In this section, we wish to determine the evolution of perturbations in the radiation density. We will first ignore the coupling between photons and baryons. In §4.3.1, we showed that the

radiation density contrast then satisfies

$$\boxed{\delta_r'' - \frac{1}{3}\nabla^2\delta_r = \frac{4}{3}\nabla^2\Phi + 4\Phi''} \quad \left\{ \begin{array}{l} \delta_r' = -\frac{4}{3}\nabla \cdot \mathbf{v}_r + 4\Phi' \quad (\text{C}) \\ \mathbf{v}_r' = -\frac{1}{4}\nabla\delta_r - \nabla\Phi \quad (\text{E}) \end{array} \right. \quad (5.2.26)$$

For later convenience, we have also recalled how the equation of motion for δ_r arises from the continuity and Euler equations. Let us discuss the solution on subhorizon scales:

- During the radiation era, the potential decays, $\Phi \propto a^{-2}$, so that

$$\delta_r'' - \frac{1}{3}\nabla^2\delta_r \approx 0. \quad (5.2.27)$$

This shows that fluctuations in the radiation density oscillate around $\delta_r = 0$, with constant amplitude and frequency $\omega = \frac{1}{\sqrt{3}}k$. Matching the solution to the superhorizon initial conditions (see §4.4), we find

$$\delta_r = 4\mathcal{R}(0) \cos\left(\frac{1}{\sqrt{3}}k\eta\right). \quad (5.2.28)$$

- During the matter era, $\Phi = \text{const.}$, so that

$$\delta_r'' - \frac{1}{3}\nabla^2\delta_r = \frac{4}{3}\nabla^2\Phi = \text{const.} \quad (5.2.29)$$

This is the equation of motion of a harmonic oscillator with constant driving force. The subhorizon fluctuations in the radiation density therefore oscillate with constant amplitude around a shifted equilibrium point, $\delta_r = -4\Phi_0(k)$, where $\Phi_0(k)$ is the k -dependent amplitude of the gravitational potential in the matter era; cf. Fig. 4.1.

*Comoving density contrast.**—In the radiation era, perturbations in the radiation density dominate (for adiabatic initial conditions). Given the solution for Φ during the radiation era (cf. §4.3.3), we therefore immediately obtain a solution for the density contrast of radiation (δ_r or Δ_r) via the Poisson equation

$$\delta_r = -\frac{2}{3}(k\eta)^2\Phi - 2\eta\Phi' - 2\Phi, \quad (5.2.30)$$

$$\Delta_r = -\frac{2}{3}(k\eta)^2\Phi. \quad (5.2.31)$$

We see that while δ_r is constant outside the horizon, Δ_r grows as $\eta^2 \propto a^2$. Inside the horizon, equation (4.3.98) implies

$$\delta_r \approx \Delta_r = -\frac{2}{3}(k\eta)^2\Phi = 4\mathcal{R}(0) \cos\left(\frac{1}{\sqrt{3}}k\eta\right), \quad (5.2.32)$$

which agrees with our previous result (5.2.28).

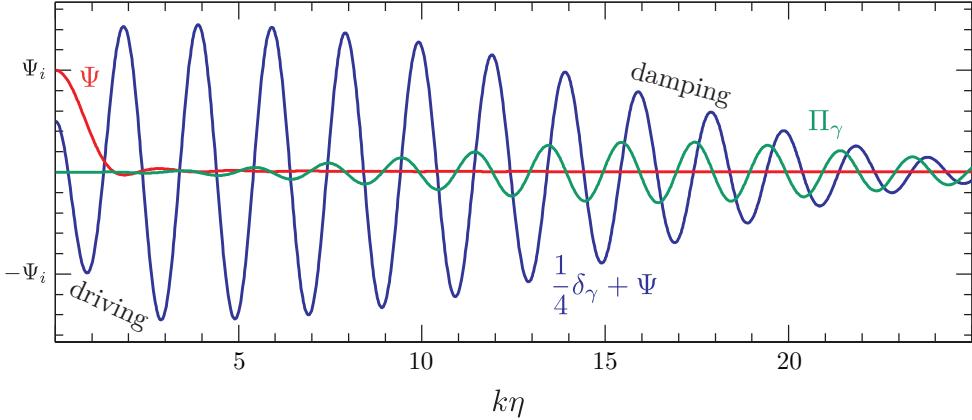


Figure 5.3: Numerical solutions of the gravitational potential Ψ and the photon density contrast δ_γ . The oscillation amplitude first rises (because the gravitational potential decays after horizon crossing) and then decays (when photon anisotropic stress Π_γ becomes significant).

5.2.2 Primordial Sound Waves

Before recombination, photons are tightly coupled to baryons (mostly through the scattering with electrons).² Only the combined momentum density of photons and baryons is now conserved

$$q \equiv q_\gamma + q_b = \frac{4}{3}(1+R)\bar{\rho}_\gamma v_\gamma, \quad (5.2.33)$$

where we have introduced

$$R \equiv \frac{3}{4} \frac{\bar{\rho}_b}{\bar{\rho}_\gamma} = 0.6 \left(\frac{\Omega_b h^2}{0.02} \right) \left(\frac{a}{10^{-3}} \right). \quad (5.2.34)$$

The Euler equation in (5.2.26) gets modified to

$$\begin{aligned} [(1+R)\mathbf{v}_r]' &= -\frac{1}{4}\nabla\delta_r - (1+R)\nabla\Phi, \\ &\quad \uparrow \quad \uparrow \\ &\quad \text{inertial mass} \quad \text{gravitational mass} \end{aligned} \quad (5.2.35)$$

The corrections to the Euler equation are easy to understand: The coupling to baryons adds extra “weight” to the photon-baryon fluid. This increases both the momentum density and the gravitational force term by a factor $(1+R)$. Since the baryons don’t contribute to the pressure, the pressure force term does not receive a factor of $(1+R)$.

The continuity equation in (5.2.26), on the other hand, does not get modified since the coupling to baryons neither creates nor destroys photons. (Thomson scattering conserves photon number, $e + \gamma \leftrightarrow e + \gamma$.) Combining the continuity and Euler equations, as before, we find

$$\boxed{\delta_\gamma'' + \frac{\mathcal{H}R}{1+R} \delta_\gamma' + c_s^2 k^2 \delta_\gamma = -\frac{4}{3} k^2 \Phi + 4\Phi'' + \frac{4R'}{1+R} \Phi'}, \quad (5.2.36)$$

$\uparrow \quad \uparrow$
 pressure gravity

²The treatment in this section is somewhat schematic. A more detailed discussion can be found in D. Baumann, *Advanced Cosmology*.

where we have defined the *sound speed* of the photon-baryon fluid as

$$c_s^2 \equiv \frac{1}{3(1+R)} . \quad (5.2.37)$$

Equation (5.2.36) is the master equation describing the anisotropies in the cosmic microwave background. A numerical solution of the equation is shown in Fig. 5.3.

5.3 CMB Anisotropies

5.3.1 Motion of the Solar System

The first thing one sees when looking at the CMB is the motion of the Solar System with respect to the rest frame of the microwave background (cf. Fig. 5.4).

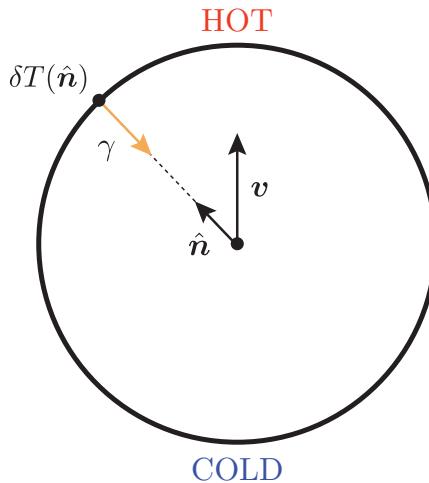


Figure 5.4: The motion of the Solar System relative to the CMB rest frame produces a dipolar pattern in the observed CMB temperature.

Due to the Doppler effect, the observed momentum of the photons coming from a direction³ \hat{n} is

$$p_0(\hat{n}) = \frac{p}{\gamma(1 - \hat{n} \cdot \mathbf{v})} \approx p(1 + \hat{n} \cdot \mathbf{v}), \quad (5.3.38)$$

where \mathbf{v} is our velocity relative to the CMB rest frame, p is the momentum of the photons in the CMB rest frame, and $\gamma = (1 - v^2)^{-1/2}$ is the Lorentz factor. We have also shown an approximation at leading order in $|\mathbf{v}| \ll 1$. As expected, the momentum is higher if we move towards the photon ($\hat{n} \cdot \mathbf{v} = v$) and smaller if we move away from it ($\hat{n} \cdot \mathbf{v} = -v$). Since the CMB is a blackbody, we can relate the change in the observed momentum of the photons to a change in the observed temperature

$$\frac{\delta T(\hat{n})}{T} \equiv \frac{T_0(\hat{n}) - T}{T} = \frac{p_0(\hat{n}) - p}{p} = \hat{n} \cdot \mathbf{v}. \quad (5.3.39)$$

Fitting this dipolar anisotropy to the data, we find that the speed of the Solar System relative to the CMB is

$$v = 368 \text{ km/s}. \quad (5.3.40)$$

After subtracting the dipole, we are left with the *primordial anisotropy*.

³Since photons are coming toward us, the propagation direction of photons, \hat{p} , is opposite of the line-of-sight direction, i.e. $\hat{p} = -\hat{n}$.

5.3.2 Perturbed Photon Geodesics

We start by discussing the effects of perturbations on the evolution of the momentum of photons. For convenience, we recall the perturbed metric in Newtonian gauge

$$ds^2 = a^2(\eta) [(1 + 2\Psi)d\eta^2 - (1 - 2\Phi)\delta_{ij}dx^i dx^j]. \quad (5.3.41)$$

For now, we will allow for the possibility that $\Psi \neq \Phi$. In Chapter 1, we showed that, in a homogeneous universe, the (physical) momentum just redshifts as $p \propto 1/a$, i.e.

$$\frac{1}{p} \frac{dp}{d\eta} = -\frac{1}{a} \frac{da}{d\eta}. \quad (5.3.42)$$

In an inhomogeneous universe, the photon evolution receives additional gravitational corrections. In the next insert, I will derive the following equation

$$\frac{1}{p} \frac{dp}{d\eta} = -\frac{1}{a} \frac{da}{d\eta} - \hat{p}^i \frac{\partial \Psi}{\partial x^i} + \frac{\partial \Phi}{\partial \eta}, \quad (5.3.43)$$

where \hat{p}^i is a unit vector in the direction of the photon 3-momentum (i.e. $\delta_{ij}\hat{p}^i\hat{p}^j = 1$). The inhomogeneous terms in (5.3.43) describe how photons are deflected and lose (gain) energy as they move out of (into) a potential well.⁴

*Derivation.**—We will derive eq. (5.3.43) from the geodesic equation for photons

$$\frac{dP^0}{d\lambda} = -\Gamma_{\alpha\beta}^0 P^\alpha P^\beta, \quad (5.3.44)$$

where $P^\mu = dX^\mu/d\lambda$ is the four-momentum of the photons. We need expressions for the components of the four-momentum in the presence of metric perturbations:

- We first consider the P^0 component. Since photons are massless we have

$$\begin{aligned} P^2 &= g_{\mu\nu} P^\mu P^\nu = 0 \\ &= a^2(1 + 2\Psi)(P^0)^2 - p^2, \end{aligned} \quad (5.3.45)$$

where we have substituted the metric (5.3.41) and defined $p^2 \equiv -g_{ij}P^i P^j$. Solving eq. (5.3.45) for P^0 , we find

$$P^0 = \frac{p}{a}(1 - \Psi). \quad (5.3.46)$$

- The spatial component of the 4-momentum takes the following form

$$P^i \equiv \alpha \hat{p}^i. \quad (5.3.47)$$

To determine the constant of proportionality α , we use

$$\begin{aligned} p^2 &= -g_{ij}P^i P^j = a^2(1 - 2\Phi)\delta_{ij}\hat{p}^i\hat{p}^j\alpha^2 \\ &= a^2(1 - 2\Phi)\alpha^2, \end{aligned} \quad (5.3.48)$$

where the last equality holds because the direction vector is a unit vector. Solving eq. (5.3.48) for α , we get $\alpha = p(1 + \Phi)/a$, or

$$P^i = \frac{p\hat{p}^i}{a}(1 + \Phi). \quad (5.3.49)$$

⁴In our conventions an overdensity corresponds to $\Phi, \Psi < 0$. (To see this consider the Poisson equation.)

Let us substitute these result into the geodesic equation (5.3.44),

$$\frac{p}{a}(1 - \Psi) \frac{d}{d\eta} \left[\frac{p}{a}(1 - \Psi) \right] = -\Gamma_{\alpha\beta}^0 P^\alpha P^\beta. \quad (5.3.50)$$

Here, we have used the standard trick of rewriting the derivative with respect to λ as a derivative with respect to time multiplied by $d\eta/d\lambda = P^0$. We expand out the time derivative to get

$$\frac{dp}{d\eta}(1 - \Psi) = \mathcal{H}p(1 - \Psi) + p \frac{d\Psi}{d\eta} - \Gamma_{\alpha\beta}^0 P^\alpha P^\beta \frac{a^2}{p}(1 + \Psi). \quad (5.3.51)$$

Next, we multiply both sides by $(1 + \Psi)/p$ and drop all quadratic terms in Ψ

$$\frac{1}{p} \frac{dp}{d\eta} = \mathcal{H} + \frac{d\Psi}{d\eta} - \Gamma_{\alpha\beta}^0 P^\alpha P^\beta \frac{a^2}{p^2}(1 + 2\Psi). \quad (5.3.52)$$

Let us consider the terms on the right-hand side in turn:

- Expanding the total time derivative of Ψ , we get

$$\frac{d\Psi}{d\eta} = \frac{\partial\Psi}{\partial\eta} + \frac{dx^i}{d\eta} \frac{\partial\Psi}{\partial x^i} = \frac{\partial\Psi}{\partial\eta} + \hat{p}^i \frac{\partial\Psi}{\partial x^i}. \quad (5.3.53)$$

where we used

$$\frac{dx^i}{d\eta} = \frac{dx^i}{d\lambda} \frac{d\lambda}{d\eta} = \frac{P^i}{P^0} = \hat{p}^i(1 + \Psi + \Phi) = \hat{p}^i + \mathcal{O}(1). \quad (5.3.54)$$

- To evaluate the last term, we need some perturbed Christoffel symbols. The relevant components are

$$\Gamma_{00}^0 = \mathcal{H} + \Psi', \quad \Gamma_{i0}^0 = \partial_i \Psi, \quad \Gamma_{ij}^0 = \mathcal{H}\delta_{ij} - [\Phi' + 2\mathcal{H}(\Phi + \Psi)]\delta_{ij}. \quad (5.3.55)$$

After a bit of algebra, we get

$$-\Gamma_{\alpha\beta}^0 \frac{P^\alpha P^\beta}{p^2}(1 + 2\Psi) = -2\mathcal{H} + \frac{\partial\Phi}{\partial\eta} - \frac{\partial\Psi}{\partial\eta} - 2\hat{p}^i \frac{\partial\Psi}{\partial x^i}. \quad (5.3.56)$$

Substituting eqs. (5.3.53) and (5.3.56) into eq. (5.3.52), we find

$$\frac{1}{p} \frac{dp}{d\eta} = -\mathcal{H} - \hat{p}^i \frac{\partial\Psi}{\partial x^i} + \frac{\partial\Phi}{\partial\eta}, \quad (5.3.57)$$

which confirms eq. (5.3.43).

Consider the gradient term in eq. (5.3.43). Let us add and subtract $\partial\Psi/\partial\eta$,

$$\hat{p}^i \frac{\partial\Psi}{\partial x^i} = \left(\frac{\partial\Psi}{\partial\eta} + \hat{p}^i \frac{\partial\Psi}{\partial x^i} \right) - \frac{\partial\Psi}{\partial\eta}. \quad (5.3.58)$$

By eq. (5.3.53), the terms in brackets simply combine into the total time derivative of Ψ . Hence, we find

$$\hat{p}^i \frac{\partial\Psi}{\partial x^i} = \frac{d\Psi}{d\eta} - \frac{\partial\Psi}{\partial\eta}, \quad (5.3.59)$$

and eq. (5.3.43) becomes

$$\frac{d}{d\eta} \ln(ap) = -\frac{d\Psi}{d\eta} + \frac{\partial(\Psi + \Phi)}{\partial\eta}.$$

(5.3.60)

5.3.3 Line-of-Sight Solution

In Chapter 3, we have seen that recombination occurs over a very short amount of time, $\Delta z \sim 10$. To simplify matters, we will work with the idealised approximation of *instantaneous recombination*. The CMB photons were then emitted at a fixed time $\eta_* = \eta_{\text{rec}}$. This moment is often called *last-scattering*. Equation (5.3.60) can then be integrated from the time of emission η_* to the time of observation η_0 (i.e. today),

$$\ln(ap)_0 = \ln(ap)_* + (\Psi_* - \Psi_0) + \int_{\eta_*}^{\eta_0} d\eta \frac{\partial}{\partial \eta}(\Psi + \Phi). \quad (5.3.61)$$

To relate this to the temperature anisotropy, we note that

$$ap \propto a\bar{T} \left(1 + \frac{\delta T}{\bar{T}} \right), \quad (5.3.62)$$

where $\bar{T}(\eta)$ is the mean temperature. Taylor-expanding the log's in (5.3.61) to first order in $\delta T/\bar{T}$, and keeping in mind that $a_0\bar{T}_0 = a_*\bar{T}_*$, we find

$$\frac{\delta T}{\bar{T}} \Big|_0 = \frac{\delta T}{\bar{T}} \Big|_* + (\Psi_* - \Psi_0) + \int_{\eta_*}^{\eta_0} d\eta \frac{\partial}{\partial \eta}(\Psi + \Phi). \quad (5.3.63)$$

The term Ψ_0 only affects the monopole perturbation, so it is unobservable⁵ and therefore usually dropped from the equation. The fractional temperature perturbation at last scattering can be expressed in terms of the density contrast of photons,⁶

$$\frac{\delta T}{\bar{T}} \Big|_* = \frac{1}{4}(\delta_\gamma)_*. \quad (5.3.64)$$

Equation (5.3.63) then reads

$$\frac{\delta T}{\bar{T}} \Big|_0 = \left(\frac{1}{4}\delta_\gamma + \Psi \right)_* + \int_{\eta_*}^{\eta_0} d\eta \frac{\partial}{\partial \eta}(\Psi + \Phi). \quad (5.3.65)$$

So far, we have ignored the motion of the electrons at the surface of last scattering. Including this leads to an extra Doppler shift in the received energy of photons when referenced to an observer comoving with the electrons at last scattering (see Fig. 5.5),

$$\frac{\delta T}{\bar{T}} \Big|_0 \subset (\hat{n} \cdot \mathbf{v}_e)_*. \quad (5.3.66)$$

Putting everything together, we obtain the following important result

$$\frac{\delta T}{\bar{T}}(\hat{n}) = \left(\frac{1}{4}\delta_\gamma + \Psi + \hat{n} \cdot \mathbf{v}_e \right)_* + \int_{\eta_*}^{\eta_0} d\eta (\Psi' + \Phi'). \quad (5.3.67)$$

Let us summarise the various contributions to the total temperature anisotropy:

- The term $\frac{1}{4}\delta_\gamma$ can be thought of as the intrinsic temperature variation over the background last-scattering surface.

⁵The total monopole is, of course, observable but its perturbation depends on the point identification with the background cosmology, i.e. is gauge-dependent.

⁶Recall that $\rho_\gamma \propto T^4$ for blackbody radiation at temperature T .

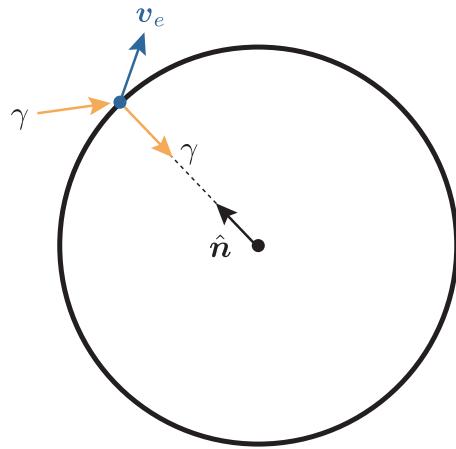


Figure 5.5: The motion of electrons at the surface of last scattering produces an additional temperature anisotropy.

- The term Ψ arises from the gravitational redshift when climbing out of a potential well at last scattering. The combination $\frac{1}{4}\delta_\gamma + \Psi$ is often called the *Sachs-Wolfe* term.
- The *Doppler* term $\hat{n} \cdot \mathbf{v}_e$ describes the blueshift from last scattering off electrons moving towards the observer.
- Finally, the *integrated Sachs-Wolfe* term describes the effect of gravitational redshifting from evolution of the potentials along the line-of-sight.

Figure 5.6 illustrates the contributions that each of the terms in (5.3.67) makes to the power spectrum of the CMB temperature anisotropies (see §5.3.4). We see that the ISW contribution is subdominant and that the shape of the power spectrum is mostly determined by the Sachs-Wolfe and Doppler contributions.

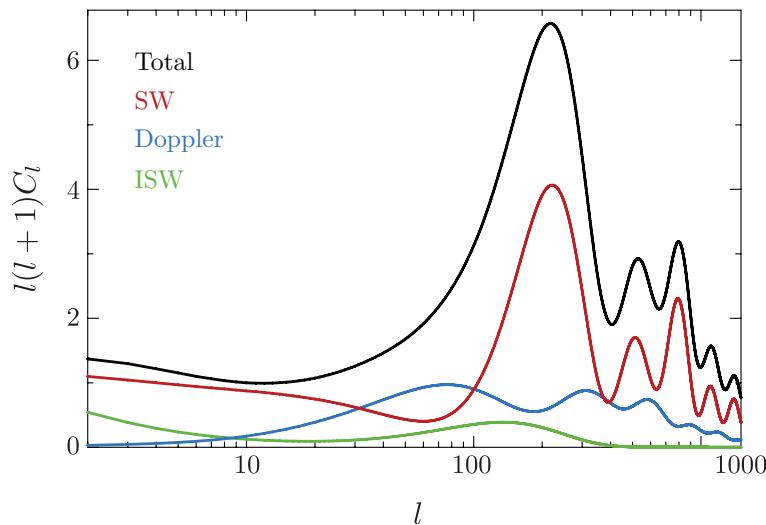


Figure 5.6: Contributions of the various terms in (5.3.67) to the temperature-anisotropy power spectrum.

Comment.—For adiabatic initial conditions (i.e. $\delta_\gamma \approx \frac{4}{3}\delta_m$) and on large scales ($\delta_m \approx -2\Phi$), the Sachs-Wolfe term becomes

$$\frac{1}{4}\delta_\gamma + \Psi = -\frac{2\Phi}{3} + \Psi \approx \frac{1}{3}\Phi. \quad (5.3.68)$$

This shows that, on large scales, an overdense region ($\Psi \approx \Phi < 0$) appears as a cold spot on the sky. While the temperature at the bottom of the potential well is hotter than the average ($-\frac{2}{3}\Phi$), photons lose more energy (Ψ) as they climb out of the potential well, resulting in a cold spot ($\frac{1}{3}\Phi < 0$).

5.3.4 CMB Power Spectrum

A map of the cosmic microwave background radiation describes the variation of the CMB temperature as a function of direction, $\delta T(\hat{n})$. We will be interested in the statistical correlations between temperature fluctuations in two different directions \hat{n} and \hat{n}' (see Fig. 5.7), averaged over the entire sky.

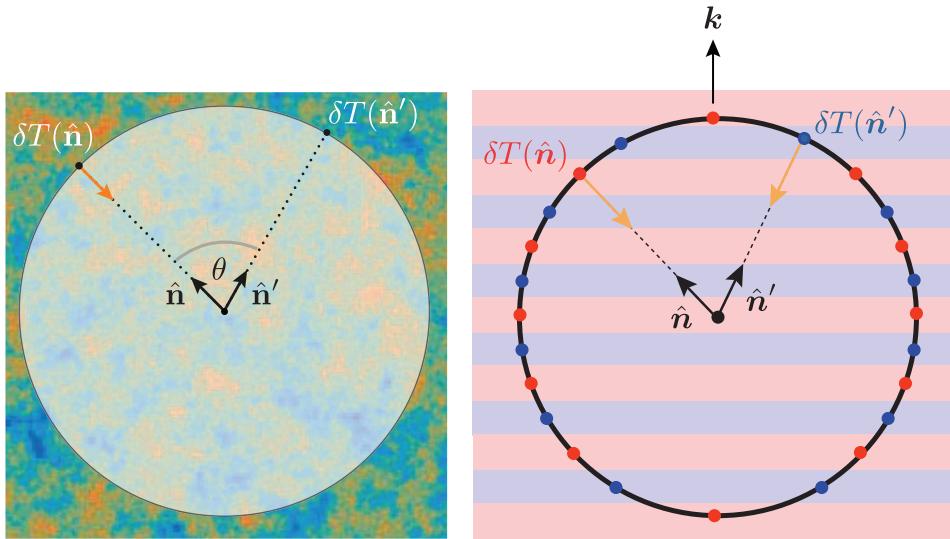


Figure 5.7: Left: Illustration of the two-point correlation function of the temperature anisotropy $\delta T(\hat{n})$. Right: Illustration of the temperature anisotropy created by a single plane wave inhomogeneity are recombination.

If the initial conditions are statistically isotropic, then we expect these correlations only to depend on the relative orientation of \hat{n} and \hat{n}' . In that case, we can write the two-point correlation function as

$$\langle \delta T(\hat{n}) \delta T(\hat{n}') \rangle = \sum_l \frac{2l+1}{4\pi} C_l P_l(\cos \theta), \quad (5.3.69)$$

where $\hat{n} \cdot \hat{n}' \equiv \cos \theta$ and P_l are Legendre polynomials. The expansion coefficients C_l are the *angular power spectrum* (cf. Fig. 5.8). If the fluctuations are Gaussian, then the power spectrum contains all the information of the CMB map.

The right panel in Figure 5.7 illustrates the temperature variations created by a single plane wave inhomogeneity. The CMB anisotropies observed on the sky are a *superposition* of many such plane waves with amplitudes that are weighted by the spectrum of primordial curvature perturbations $\Delta_{\mathcal{R}}^2(k)$. In Chapter 6, we will show that the initial conditions of the primordial

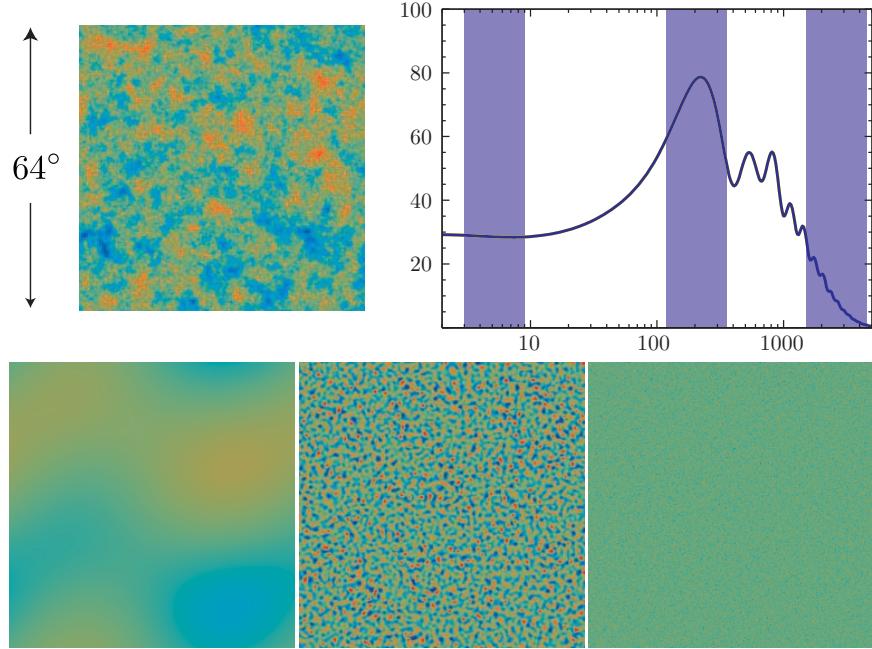


Figure 5.8: From temperature maps to the angular power spectrum. The original temperature fluctuation map (top left) corresponding to a simulation of the power spectrum (top right) can be band-filtered to illustrate the power spectrum in three characteristic regimes: the large-scale regime, the first acoustic peak where most of the power lies, and the damping tail where fluctuations are dissipated.

perturbations are expected to be featureless, $\Delta_{\mathcal{R}}^2(k) = \text{const}$. The observed features in the CMB anisotropy spectrum arise from the subhorizon evolution of the photon density perturbations and the gravitational potential as discussed in §5.2 (cf. Fig. 5.3). These acoustic waves are *captured* at recombination and *projected* onto the sky. The observed oscillations in the CMB power spectrum are therefore a snapshot of primordial sound waves caught at different phases in their evolution at the time when photons last scattered off electrons. The beautiful physics of the CMB fluctuations is described in much detail in D. Baumann, *Advanced Cosmology*.

6

Quantum Initial Conditions*

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 $\delta\phi(t, \mathbf{x}) \equiv \phi(t, \mathbf{x}) - \bar{\phi}(t)$. There will hence be local differences in the time when inflation ends, $\delta t(\mathbf{x})$, so that different regions of space inflate by different amounts. These differences in the

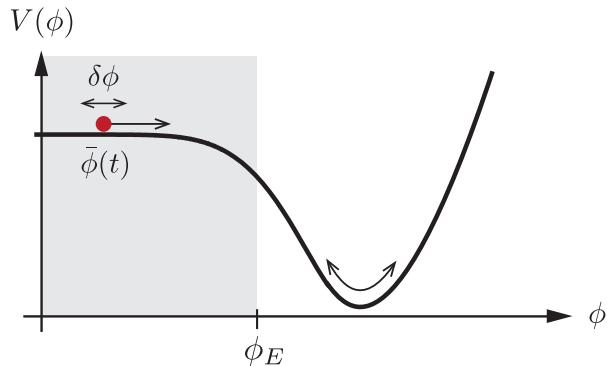


Figure 6.1: Quantum fluctuations $\delta\phi(t, \mathbf{x})$ around the classical background evolution $\bar{\phi}(t)$. Regions acquiring a negative fluctuation $\delta\phi$ remain potential-dominated longer than regions with positive $\delta\phi$. Different parts of the universe therefore undergo slightly different evolutions. After inflation, this induces density fluctuations $\delta\rho(t, \mathbf{x})$.

local expansion histories lead to differences in the local densities after inflation, $\delta\rho(t, \mathbf{x})$, and ultimately in the CMB temperature, $\delta T(\mathbf{x})$. 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

$$\langle |\delta\phi_k|^2 \rangle \equiv \langle 0 || \delta\phi_k |^2 | 0 \rangle . \quad (6.1.1)$$

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.)

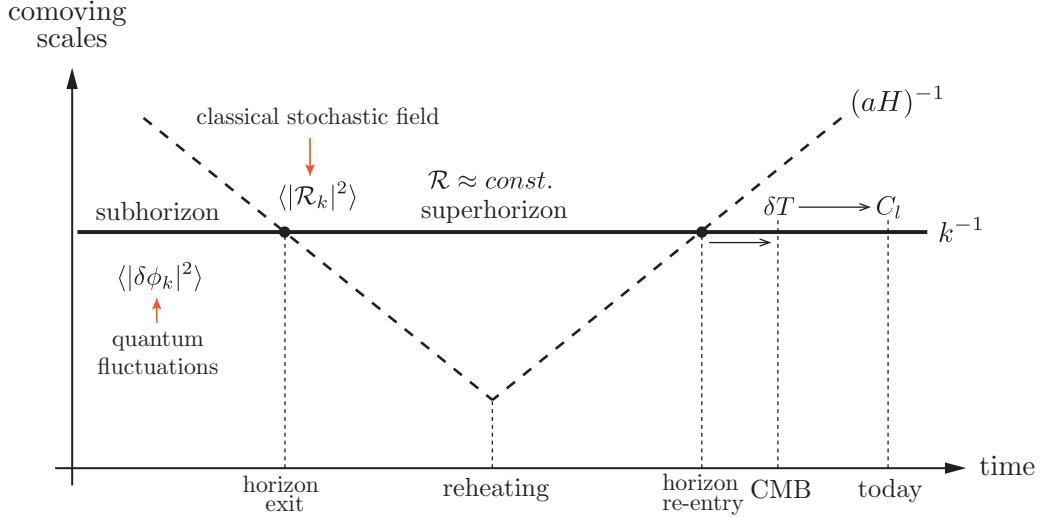


Figure 6.2: Curvature perturbations during and after inflation: The comoving horizon $(aH)^{-1}$ shrinks during inflation and grows in the subsequent FRW evolution. This implies that comoving scales k^{-1} exit the horizon at early times and re-enter the horizon at late times. While the curvature perturbations \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*:

$$\boxed{\mathcal{R} = -\frac{\mathcal{H}}{\bar{\phi}'} \delta\phi} . \quad (6.1.2)$$

$\delta\phi \rightarrow \mathcal{R}$.—From the gauge-invariant definition of \mathcal{R} , eq. (??), we get

$$\mathcal{R} = C - \frac{1}{3} \nabla^2 E + \mathcal{H}(B + v) \xrightarrow{\text{spatially flat}} \mathcal{H}(B + v) . \quad (6.1.3)$$

We recall that the combination $B + v$ appeared in the off-diagonal component of the perturbed stress tensor, cf. eq. (??),

$$\delta T^0_j = -(\bar{\rho} + \bar{P}) \partial_j(B + v) . \quad (6.1.4)$$

We compare this to the first-order perturbation of the stress tensor of a scalar field, cf. eq. (2.3.21),

$$\delta T^0_j = g^{0\mu} \partial_\mu \phi \partial_j \delta\phi = \bar{g}^{00} \partial_0 \bar{\phi} \partial_j \delta\phi = \frac{\bar{\phi}'}{a^2} \partial_j \delta\phi , \quad (6.1.5)$$

to get

$$B + v = -\frac{\delta\phi}{\bar{\phi}'} . \quad (6.1.6)$$

Substituting (6.1.6) into (6.1.3) we obtain (6.1.2).

The variance of curvature perturbations therefore is

$$\langle |\mathcal{R}_k|^2 \rangle = \left(\frac{\mathcal{H}}{\bar{\phi}'} \right)^2 \langle |\delta\phi_k|^2 \rangle , \quad (6.1.7)$$

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

$$S = \int d\tau d^3x \sqrt{-g} \left[\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) \right] , \quad (6.2.8)$$

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 the coupled dynamics of inflaton and metric perturbations.)

Evaluating (6.2.8) for the *unperturbed* FRW metric, we find

$$S = \int d\tau d^3x \left[\frac{1}{2} a^2 ((\phi')^2 - (\nabla\phi)^2) - a^4 V(\phi) \right] . \quad (6.2.9)$$

It is convenient to write the perturbed inflaton field as

$$\phi(\tau, \mathbf{x}) = \bar{\phi}(\tau) + \frac{f(\tau, \mathbf{x})}{a(\tau)} . \quad (6.2.10)$$

To get the linearised equation of motion for $f(\tau, \mathbf{x})$, 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

$$S^{(1)} = \int d\tau d^3x \left[a\bar{\phi}'f' - a'\bar{\phi}'f - a^3V_{,\phi}f \right], \quad (6.2.11)$$

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

$$\begin{aligned} S^{(1)} &= - \int d\tau d^3x \left[\partial_\tau(a\bar{\phi}') + a'\bar{\phi}' + a^3V_{,\phi} \right] f, \\ &= - \int d\tau d^3x a \left[\bar{\phi}'' + 2\mathcal{H}\bar{\phi}' + a^2V_{,\phi} \right] f. \end{aligned} \quad (6.2.12)$$

Requiring that $S^{(1)} = 0$, for all f , gives the Klein-Gordon equation for the background field,

$$\bar{\phi}'' + 2\mathcal{H}\bar{\phi}' + a^2V_{,\phi} = 0. \quad (6.2.13)$$

- Isolating all terms with two factors of f , we get the quadratic action

$$S^{(2)} = \frac{1}{2} \int d\tau d^3x \left[(f')^2 - (\nabla f)^2 - 2\mathcal{H}ff' + (\mathcal{H}^2 - a^2V_{,\phi\phi})f^2 \right]. \quad (6.2.14)$$

Integrating the $ff' = \frac{1}{2}(f^2)'$ term by parts, gives

$$\begin{aligned} S^{(2)} &= \frac{1}{2} \int d\tau d^3x \left[(f')^2 - (\nabla f)^2 + (\mathcal{H}' + \mathcal{H}^2 - a^2V_{,\phi\phi})f^2 \right], \\ &= \frac{1}{2} \int d\tau d^3x \left[(f')^2 - (\nabla f)^2 + \left(\frac{a''}{a} - a^2V_{,\phi\phi} \right) f^2 \right]. \end{aligned} \quad (6.2.15)$$

During slow-roll inflation, we have

$$\frac{V_{,\phi\phi}}{H^2} \approx \frac{3M_{\text{pl}}^2 V_{,\phi\phi}}{V} = 3\eta_v \ll 1. \quad (6.2.16)$$

Since $a' = a^2H$, with $H \approx \text{const.}$, we also have

$$\frac{a''}{a} \approx 2a'H = 2a^2H^2 \gg a^2V_{,\phi\phi}. \quad (6.2.17)$$

Hence, we can drop the $V_{,\phi\phi}$ term in (6.2.15),

$$S^{(2)} \approx \int d\tau d^3x \frac{1}{2} \left[(f')^2 - (\nabla f)^2 + \frac{a''}{a} f^2 \right].$$

(6.2.18)

Applying the Euler-Lagrange equation to (6.2.18) gives the *Mukhanov-Sasaki equation*

$$f'' - \nabla^2 f - \frac{a''}{a} f = 0, \quad (6.2.19)$$

or, for each Fourier mode,

$$f''_k + \left(k^2 - \frac{a''}{a} \right) f_k = 0.$$

(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

$$f_k'' + k^2 f_k \approx 0 . \quad (6.2.21)$$

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 problem¹: 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

$$S[q] = \frac{1}{2} \int dt \left[\dot{q}^2 - \omega^2 q^2 \right] , \quad (6.3.22)$$

and the equation of motion is $\ddot{q} + \omega^2 q = 0$. The conjugate momentum is

$$p = \frac{\partial L}{\partial \dot{q}} = \dot{q} . \quad (6.3.23)$$

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)

$$[\hat{q}, \hat{p}] = i , \quad (\text{I})$$

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

$$\hat{q}(t) = q(t) \hat{a} + q^*(t) \hat{a}^\dagger , \quad (\text{II})$$

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 \hat{a} . Substituting (II) into (I), we get

$$W[q, q^*] \times [\hat{a}, \hat{a}^\dagger] = 1 , \quad (6.3.24)$$

where we have defined the *Wronskian* as

$$W[q_1, q_2^*] \equiv -i (q_1 \partial_t q_2^* - (\partial_t q_1) q_2^*) . \quad (6.3.25)$$

¹The 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 ($q \rightarrow \lambda q$) such that

$$W[q, q^*] \equiv 1 , \quad (\text{III})$$

and hence

$$[\hat{a}, \hat{a}^\dagger] = 1 . \quad (\text{IV})$$

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}

$$\hat{a}|0\rangle = 0 . \quad (\text{V})$$

Excited states are created by repeated application of creation operators

$$|n\rangle \equiv \frac{1}{\sqrt{n!}}(\hat{a}^\dagger)^n|0\rangle . \quad (6.3.26)$$

These states are eigenstates of the number operator $\hat{N} \equiv \hat{a}^\dagger \hat{a}$ with eigenvalue n , i.e.

$$\hat{N}|n\rangle = n|n\rangle . \quad (6.3.27)$$

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 $q(t)$,

$$\begin{aligned} \hat{H} &= \frac{1}{2}\hat{p}^2 + \frac{1}{2}\omega^2\hat{q}^2 \\ &= \frac{1}{2}\left[(\dot{q}^2 + \omega^2q^2)\hat{a}\hat{a} + (\dot{q}^2 + \omega^2q^2)^*\hat{a}^\dagger\hat{a}^\dagger + (|\dot{q}|^2 + \omega^2|q|^2)(\hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a})\right]. \end{aligned} \quad (6.3.28)$$

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

$$\hat{H}|0\rangle = \frac{1}{2}(\dot{q}^2 + \omega^2q^2)^*\hat{a}^\dagger\hat{a}^\dagger|0\rangle + \frac{1}{2}(|\dot{q}|^2 + \omega^2|q|^2)|0\rangle . \quad (6.3.29)$$

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

$$\dot{q} = \pm i\omega q . \quad (6.3.30)$$

For such a function q , the norm is

$$W[q, q^*] = \mp 2\omega|q|^2 , \quad (6.3.31)$$

and positivity of the normalisation condition $W[q, q^*] > 0$ selects the minus sign in (6.3.30)

$$\dot{q} = -i\omega q \quad \Rightarrow \quad q(t) \propto e^{-i\omega t} . \quad (6.3.32)$$

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

$$q(t) = \frac{1}{\sqrt{2\omega}} e^{-i\omega t} . \quad (6.3.33)$$

With this choice of mode function, the Hamiltonian takes the familiar form

$$\hat{H} = \hbar\omega \left(\hat{N} + \frac{1}{2} \right) . \quad (6.3.34)$$

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

$$\begin{aligned} \langle \hat{q} \rangle &\equiv \langle 0 | \hat{q} | 0 \rangle \\ &= \langle 0 | \underbrace{q(t)\hat{a} + q^*(t)\hat{a}^\dagger}_{|0\rangle} | 0 \rangle \\ &= 0 , \end{aligned} \quad (6.3.35)$$

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

$$\begin{aligned} \langle |\hat{q}|^2 \rangle &\equiv \langle 0 | \hat{q}^\dagger \hat{q} | 0 \rangle \\ &= \langle 0 | \underbrace{(q^*\hat{a}^\dagger + q\hat{a})(q\hat{a} + q^*\hat{a}^\dagger)}_{|0\rangle} | 0 \rangle \\ &= |q(t)|^2 \langle 0 | \hat{a} \hat{a}^\dagger | 0 \rangle \\ &= |q(t)|^2 \langle 0 | [\hat{a}, \hat{a}^\dagger] | 0 \rangle \\ &= |q(t)|^2 . \end{aligned} \quad (6.3.36)$$

Hence, we find that the variance of the amplitude of the quantum oscillator is given by the square of the mode function

$$\langle |\hat{q}|^2 \rangle = |q(t)|^2 = \frac{\hbar}{2\omega} . \quad (\text{VI})$$

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

$$\pi \equiv \frac{\partial \mathcal{L}}{\partial f'} = f' . \quad (6.4.37)$$

We perform the canonical quantisation just like in the case of the harmonic oscillator.

6.4.1 Canonical Quantisation

We promote the fields $f(\tau, \mathbf{x})$ and $\pi(\tau, \mathbf{x})$ to quantum operators $\hat{f}(\tau, \mathbf{x})$ and $\hat{\pi}(\tau, \mathbf{x})$. The operators satisfy the equal time CCR

$$[\hat{f}(\tau, \mathbf{x}), \hat{\pi}(\tau, \mathbf{x}')]=i\delta(\mathbf{x}-\mathbf{x}') . \quad (\text{I}')$$

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

$$\begin{aligned} [\hat{f}_{\mathbf{k}}(\tau), \hat{\pi}_{\mathbf{k}'}(\tau)] &= \int \frac{d^3x}{(2\pi)^{3/2}} \int \frac{d^3x'}{(2\pi)^{3/2}} \underbrace{[\hat{f}(\tau, \mathbf{x}), \hat{\pi}(\tau, \mathbf{x}')]_{i\delta(\mathbf{x}-\mathbf{x}')}} e^{-i\mathbf{k}\cdot\mathbf{x}} e^{-i\mathbf{k}'\cdot\mathbf{x}'} \\ &= i \int \frac{d^3x}{(2\pi)^3} e^{-i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{x}} \\ &= i\delta(\mathbf{k} + \mathbf{k}') , \end{aligned} \quad (\text{I}'')$$

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

$$\hat{f}_{\mathbf{k}}(\tau) = f_k(\tau) \hat{a}_{\mathbf{k}} + f_k^*(\tau) \hat{a}_{\mathbf{k}}^\dagger , \quad (\text{II}')$$

where $\hat{a}_{\mathbf{k}}$ is a time-independent operator, $\hat{a}_{\mathbf{k}}^\dagger$ 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

$$f_k'' + \omega_k^2(\tau) f_k = 0 , \quad \text{where } \omega_k^2(\tau) \equiv k^2 - \frac{a''}{a} . \quad (6.4.38)$$

As indicated by dropping the vector notation \mathbf{k} on the subscript, the mode functions, $f_k(\tau)$ and $f_k^*(\tau)$, are the same for all Fourier modes with $k \equiv |\mathbf{k}|$.²

Substituting (II') into (I''), we get

$$W[f_k, f_k^*] \times [\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}^\dagger] = \delta(\mathbf{k} + \mathbf{k}') , \quad (6.4.39)$$

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

$$W[f_k, f_k^*] \equiv 1 . \quad (\text{III}')$$

²Since the frequency $\omega_k(\tau)$ depends only on $k \equiv |\mathbf{k}|$, the evolution does not depend on direction. The constant operators $\hat{a}_{\mathbf{k}}$ and $\hat{a}_{\mathbf{k}}^\dagger$ define initial conditions which may depend on direction.

Eq. (6.4.39) then becomes

$$[\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}^\dagger] = \delta(\mathbf{k} + \mathbf{k}') , \quad (\text{IV}')$$

which is the same as (IV), but for each Fourier mode.

As before, the operators $\hat{a}_{\mathbf{k}}^\dagger$ and $\hat{a}_{\mathbf{k}}$ 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

$$\hat{a}_{\mathbf{k}}|0\rangle = 0 , \quad (\text{V}')$$

and by producing excited states by repeated application of creation operators

$$|m_{\mathbf{k}_1}, n_{\mathbf{k}_2}, \dots\rangle = \frac{1}{\sqrt{m!n! \dots}} \left[(a_{\mathbf{k}_1}^\dagger)^m (a_{\mathbf{k}_2}^\dagger)^n \dots \right] |0\rangle . \quad (6.4.40)$$

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

$$\omega_k^2 = k^2 - \frac{a''}{a} \approx k^2 - \frac{2}{\tau^2} \xrightarrow{\tau \rightarrow -\infty} k^2 , \quad (6.4.41)$$

and the Mukhanov-Sasaki equation reduces to

$$f_k'' + k^2 f_k \approx 0 . \quad (6.4.42)$$

But this is just the equation for a free field in Minkowski 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

$$\lim_{\tau \rightarrow -\infty} f_k(\tau) = \frac{1}{\sqrt{2k}} e^{-ik\tau} . \quad (6.4.43)$$

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

$$f_k'' + \left(k^2 - \frac{2}{\tau^2} \right) f_k = 0 . \quad (6.4.44)$$

This has an exact solution

$$f_k(\tau) = \alpha \frac{e^{-ik\tau}}{\sqrt{2k}} \left(1 - \frac{i}{k\tau} \right) + \beta \frac{e^{ik\tau}}{\sqrt{2k}} \left(1 + \frac{i}{k\tau} \right) . \quad (6.4.45)$$

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

$$f_k(\tau) = \frac{e^{-ik\tau}}{\sqrt{2k}} \left(1 - \frac{i}{k\tau} \right). \quad (6.4.46)$$

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

$$\hat{f}(\tau, \mathbf{x}) = \int \frac{d^3 k}{(2\pi)^{3/2}} \left[f_k(\tau) \hat{a}_k + f_k^*(\tau) \hat{a}_k^\dagger \right] e^{i\mathbf{k}\cdot\mathbf{x}}. \quad (6.4.47)$$

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

$$\begin{aligned} \langle |\hat{f}|^2 \rangle &\equiv \langle 0 | \hat{f}^\dagger(\tau, \mathbf{0}) \hat{f}(\tau, \mathbf{0}) | 0 \rangle \\ &= \int \frac{d^3 k}{(2\pi)^{3/2}} \int \frac{d^3 k'}{(2\pi)^{3/2}} \langle 0 | \overline{(f_k^*(\tau) \hat{a}_k^\dagger + f_k(\tau) \hat{a}_k)(f_{k'}(\tau) \hat{a}_{k'}^\dagger + f_{k'}^*(\tau) \hat{a}_{k'}^\dagger)} | 0 \rangle \\ &= \int \frac{d^3 k}{(2\pi)^{3/2}} \int \frac{d^3 k'}{(2\pi)^{3/2}} f_k(\tau) f_{k'}^*(\tau) \langle 0 | [\hat{a}_k, \hat{a}_{k'}^\dagger] | 0 \rangle \\ &= \int \frac{d^3 k}{(2\pi)^3} |f_k(\tau)|^2 \\ &= \int d \ln k \frac{k^3}{2\pi^2} |f_k(\tau)|^2. \end{aligned} \quad (6.4.48)$$

We define the (dimensionless) *power spectrum* as

$$\Delta_f^2(k, \tau) \equiv \frac{k^3}{2\pi^2} |f_k(\tau)|^2. \quad (\text{VI}')$$

As in (VI), the square of the classical solution determines the variance of quantum fluctuations. Using (6.4.46), we find

$$\Delta_{\delta\phi}^2(k, \tau) = a^{-2} \Delta_f^2(k, \tau) = \left(\frac{H}{2\pi} \right)^2 \left(1 + \left(\frac{k}{aH} \right)^2 \right) \xrightarrow{\text{superhorizon}} \left(\frac{H}{2\pi} \right)^2. \quad (6.4.49)$$

We will use the approximation that the power spectrum at horizon crossing is³

$$\Delta_{\delta\phi}^2(k) \approx \left(\frac{H}{2\pi} \right)^2 \Big|_{k=aH}. \quad (6.4.50)$$

³Computing 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 at 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 (\mathbf{II}') and its conjugate momentum operator

$$\hat{f}(\tau, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^{3/2}} \left[f_k(\tau) \hat{a}_k + f_k^*(\tau) a_k^\dagger \right] e^{i\mathbf{k}\cdot\mathbf{x}}, \quad (6.4.51)$$

$$\hat{\pi}(\tau, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^{3/2}} \left[f'_k(\tau) \hat{a}_k + (f_k^*)'(\tau) a_k^\dagger \right] e^{i\mathbf{k}\cdot\mathbf{x}}. \quad (6.4.52)$$

In the superhorizon limit, $k\tau \rightarrow 0$, we have

$$f_k(\tau) \approx -\frac{1}{\sqrt{2}k^{3/2}} \frac{i}{\tau} \quad \text{and} \quad f'_k(\tau) \approx \frac{1}{\sqrt{2}k^{3/2}} \frac{i}{\tau^2}, \quad (6.4.53)$$

and hence

$$\hat{f}(\tau, \mathbf{x}) = -\frac{i}{\sqrt{2}\tau} \int \frac{d^3k}{(2\pi)^{3/2}} \frac{1}{k^{3/2}} \left[\hat{a}_k - a_k^\dagger \right] e^{i\mathbf{k}\cdot\mathbf{x}}, \quad (6.4.54)$$

$$\hat{\pi}(\tau, \mathbf{x}) = \frac{i}{\sqrt{2}\tau^2} \int \frac{d^3k}{(2\pi)^{3/2}} \frac{1}{k^{3/2}} \left[\hat{a}_k - a_k^\dagger \right] e^{i\mathbf{k}\cdot\mathbf{x}} = -\frac{1}{\tau} \hat{f}(\tau, \mathbf{x}). \quad (6.4.55)$$

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),

$$\Delta_{\mathcal{R}}^2 = \frac{1}{2\varepsilon} \frac{\Delta_{\delta\phi}^2}{M_{\text{pl}}^2}, \quad \text{where} \quad \varepsilon = \frac{\frac{1}{2}\dot{\phi}^2}{M_{\text{pl}}^2 H^2}. \quad (6.5.56)$$

Substituting (6.4.50), we get

$$\boxed{\Delta_{\mathcal{R}}^2(k) = \frac{1}{8\pi^2} \frac{1}{\varepsilon} \frac{H^2}{M_{\text{pl}}^2} \Big|_{k=aH}}. \quad (6.5.57)$$

Exercise.—Show that for slow-roll inflation, eq. (6.5.59) can be written as

$$\boxed{\Delta_{\mathcal{R}}^2 = \frac{1}{12\pi^2} \frac{V^3}{M_{\text{pl}}^6 (V')^2}}. \quad (6.5.58)$$

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 $\Delta_{\mathcal{R}}^2(k)$ 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 $\Delta_{\mathcal{R}}^2 \sim k^0$. Near a reference scale k_* , the k -dependence of the spectrum takes a power-law form

$$\Delta_{\mathcal{R}}^2(k) \equiv A_s \left(\frac{k}{k_*} \right)^{n_s - 1}. \quad (6.5.59)$$

The measured amplitude of the scalar spectrum at $k_* = 0.05 \text{ Mpc}^{-1}$ is

$$A_s = (2.196 \pm 0.060) \times 10^{-9}. \quad (6.5.60)$$

To quantify the deviation from scale-invariance we have introduced the *scalar spectral index*

$$n_s - 1 \equiv \frac{d \ln \Delta_{\mathcal{R}}^2}{d \ln k}, \quad (6.5.61)$$

where the right-hand side is evaluated at $k = k_*$ and $n_s = 1$ corresponds to perfect scale-invariance. We can split (6.5.61) into two factors

$$\frac{d \ln \Delta_{\mathcal{R}}^2}{d \ln k} = \frac{d \ln \Delta_{\mathcal{R}}^2}{dN} \times \frac{dN}{d \ln k}. \quad (6.5.62)$$

The derivative with respect to e -folds is

$$\frac{d \ln \Delta_{\mathcal{R}}^2}{dN} = 2 \frac{d \ln H}{dN} - \frac{d \ln \varepsilon}{dN}. \quad (6.5.63)$$

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

$$\ln k = N + \ln H. \quad (6.5.64)$$

Hence, we have

$$\frac{dN}{d \ln k} = \left[\frac{d \ln k}{dN} \right]^{-1} = \left[1 + \frac{d \ln H}{dN} \right]^{-1} \approx 1 + \varepsilon. \quad (6.5.65)$$

To first order in the Hubble slow-roll parameters, we therefore find

$$n_s - 1 = -2\varepsilon - \eta. \quad (6.5.66)$$

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

$$n_s = 0.9603 \pm 0.0073. \quad (6.5.67)$$

Exercise.—For slow-roll inflation, show that

$$n_s - 1 = -3M_{\text{pl}}^2 \left(\frac{V'}{V} \right)^2 + 2M_{\text{pl}}^2 \frac{V''}{V}. \quad (6.5.68)$$

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,

$$ds^2 = a^2(\tau) \left[d\tau^2 - (\delta_{ij} + 2\hat{E}_{ij}) dx^i dx^j \right]. \quad (6.5.69)$$

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

$$S = \frac{M_{\text{pl}}^2}{2} \int d^4x \sqrt{-g} R \quad \Rightarrow \quad S^{(2)} = \frac{M_{\text{pl}}^2}{8} \int d\tau d^3x a^2 \left[(\hat{E}'_{ij})^2 - (\nabla \hat{E}_{ij})^2 \right]. \quad (6.5.70)$$

It is convenient to define

$$\frac{M_{\text{pl}}}{2} a \hat{E}_{ij} \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} f_+ & f_\times & 0 \\ f_\times & -f_+ & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (6.5.71)$$

so that

$$S^{(2)} = \frac{1}{2} \sum_{I=+, \times} \int d\tau d^3x \left[(f'_I)^2 - (\nabla f_I)^2 + \frac{a''}{a} f_I^2 \right]. \quad (6.5.72)$$

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 ,

$$\Delta_t^2 \equiv 2 \times \Delta_{\hat{E}}^2 = 2 \times \left(\frac{2}{a M_{\text{pl}}} \right)^2 \times \Delta_f^2. \quad (6.5.73)$$

Using (6.4.50), we get

$$\boxed{\Delta_t^2(k) = \frac{2}{\pi^2} \frac{H^2}{M_{\text{pl}}^2} \Big|_{k=aH}}. \quad (6.5.74)$$

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

$$\Delta_t^2(k) \equiv A_t \left(\frac{k}{k_*} \right)^{n_t}, \quad (6.5.75)$$

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*

$$r \equiv \frac{A_t}{A_s}. \quad (6.5.76)$$

Tensors have not been observed yet, so we only have an upper limit on their amplitude, $r \lesssim 0.17$.

Exercise.—Show that

$$r = 16\varepsilon \quad (6.5.77)$$

$$n_t = -2\varepsilon . \quad (6.5.78)$$

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.

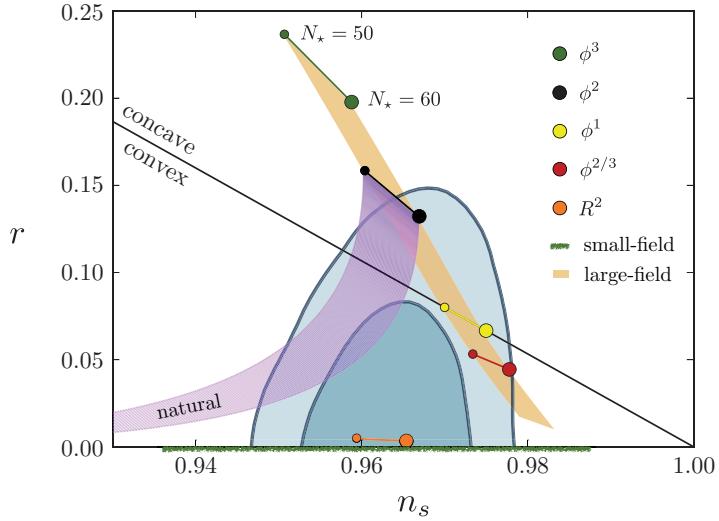


Figure 6.3: Latest constraints on the scalar spectral index n_s and the tensor amplitude r .

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. 5.2. In fig. 6.4 we compare this prediction to the measured matter power.⁴

⁴ 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.

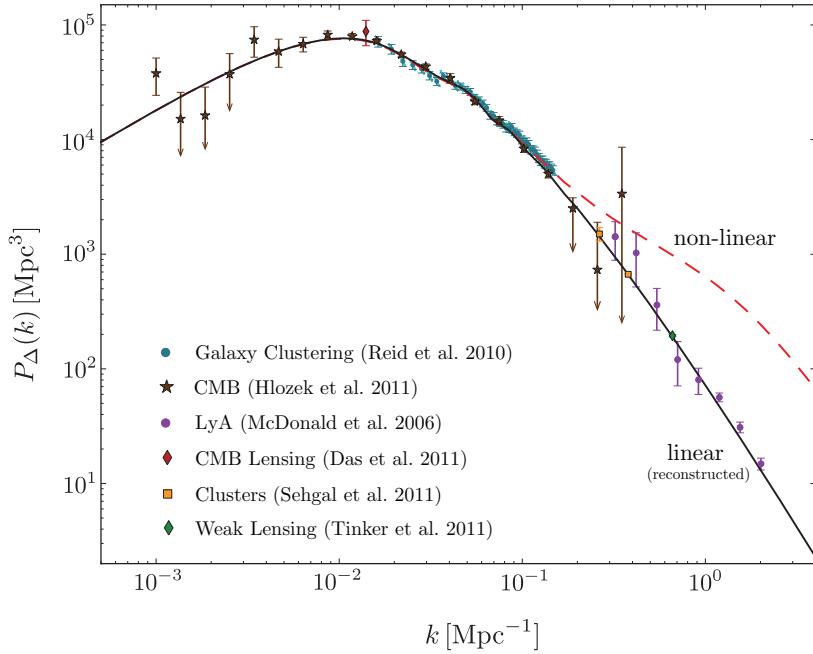


Figure 6.4: Compilation of the latest measurements of the matter power spectrum.

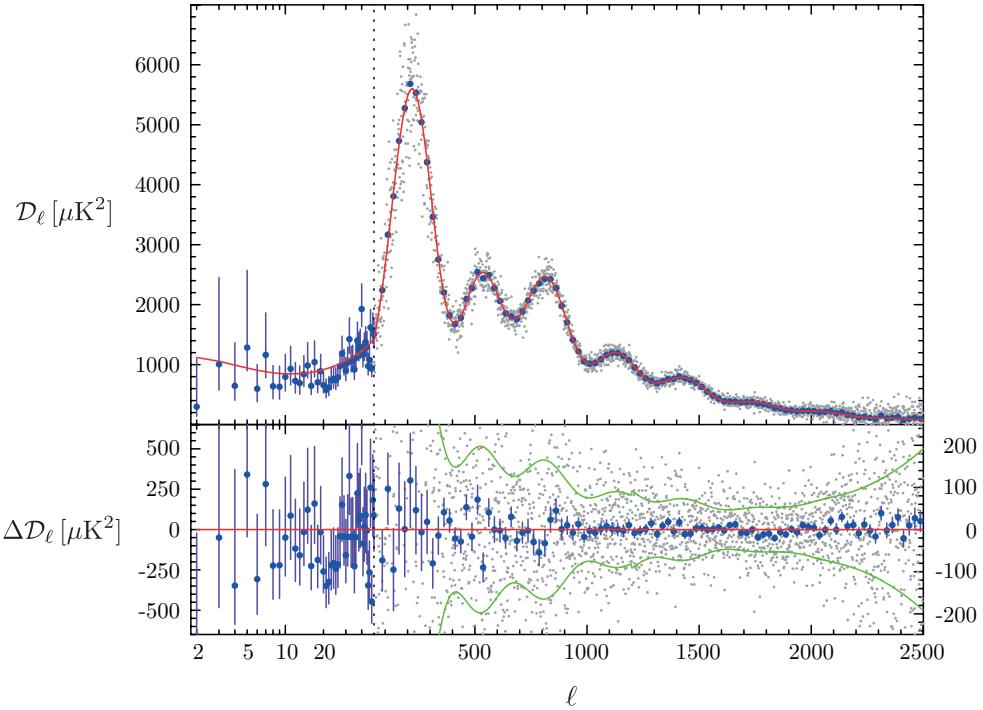


Figure 6.5: The latest measurements of the CMB angular power spectrum by the Planck satellite.

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.

A

Elements of General Relativity

The following is a brief introduction to the basic elements of general relativity (GR). A familiarity with special relativity and classical dynamics will be assumed, but otherwise we will start from first principles. We will be schematic, so this is no substitute for a proper course on the subject. Our goal is to present the minimal theoretical background required for a course in cosmology.¹

A.1 Gravity as Geometry

The origin of general relativity lies in the following simple question: Why do objects with different masses fall at the same rate? We think we know the answer: the mass of an object cancels in Newton's law

$$\cancel{m} \mathbf{a} = \cancel{m} \mathbf{g}, \quad (\text{A.1.1})$$

where \mathbf{g} is the local gravitational acceleration. However, Einstein pointed out that the meaning of ‘mass’ on the left-hand side and the right-hand side of (A.1.1) is quite different. We should really distinguish between the two masses by giving them different names:

$$m_i \mathbf{a} = m_g \mathbf{g}. \quad (\text{A.1.2})$$

The *gravitational mass*, m_g , is a source for the gravitational field (just like the charge q is a source for an electric field), while the *inertial mass*, m_i , characterizes the dynamical response to any forces. It is a nontrivial result that experiments find²

$$\frac{m_i}{m_g} = 1 \pm 10^{-13}. \quad (\text{A.1.3})$$

But is there a deeper reason that the gravitational force is proportional to the inertial mass? There are two other forces which are also proportional to the inertial mass. These are

$$\text{Centrifugal force : } \mathbf{F} = -m_i \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}). \quad (\text{A.1.4})$$

$$\text{Coriolis force : } \mathbf{F} = -2m_i \boldsymbol{\omega} \times \dot{\mathbf{r}}. \quad (\text{A.1.5})$$

In both of these cases, we understand why the force is proportional to the inertial mass, namely because these are “fictitious forces”, arising in a non-inertial frame. (In this case, one that is rotating with frequency $\boldsymbol{\omega}$). Could gravity also be a fictitious force, arising only because we are in a non-inertial frame?

¹ A GR primer with a slightly more mathematical focus is Sean Carroll's *No-Nonsense Introduction to GR*.

²Note that (A.1.2) defines both m_g and \mathbf{g} . For any given material (say test masses made of platinum), we can therefore define $m_g = m_i$ by the rescaling $\mathbf{g} \rightarrow \lambda \mathbf{g}$ and $m_g \rightarrow \lambda^{-1} m_g$. What is nontrivial is that (A.1.3) holds for other bodies made of other materials.

A.1.1 The Happiest Thought

"I was sitting in a chair in the patent office in Bern when all of a sudden a thought occurred to me: "If a person falls freely he will not feel his own weight." I was startled. This simple thought made a deep impression on me. It impelled me toward a theory of gravitation."

Albert Einstein

Consider the equation of motion of a particle in a gravitational field

$$\ddot{\mathbf{x}} = \mathbf{g}(\mathbf{x}(t), t). \quad (\text{A.1.6})$$

Solutions of this equation are uniquely determined by the initial position and velocity of the particle. Any two particles with the same initial position and velocity will follow the same trajectory. Now consider a new frame of reference moving with constant acceleration \mathbf{a} relative to the first one. The equation of motion in this frame is

$$\ddot{\mathbf{x}}' = \mathbf{g}' - \mathbf{a} \equiv \mathbf{g}'. \quad (\text{A.1.7})$$

The dynamics of the particle in the new frame is the same as in the old frame, but with a different gravitational field \mathbf{g}' . Choosing $\mathbf{a} = \mathbf{g}$ (i.e. going to the frame of a freely falling observer), we find $\mathbf{g}' = 0$ (no gravitational field). This was Einstein's happiest thought: a freely falling observer doesn't feel a gravitational field. Conversely, even if $\mathbf{g} = 0$, a non-zero gravitational field $\mathbf{g}' = -\mathbf{a}$ can be experienced in an accelerating frame. Einstein elevated these observations to the *equivalence principle*:

uniform acceleration is indistinguishable from a uniform gravitational field.

Even if the gravitational field is not uniform, it can be approximated as uniform for experiments performed in a region of spacetime which is sufficiently small so that the non-uniformity is negligible.

A.1.2 Bending of Light

An immediate experimental consequence of the equivalence principle is that light bends in a gravitational field. To see this, compare the following two situations (cf. Fig. A.1): *i*) a rocket in outer space moving with constant acceleration $\mathbf{a} = g\hat{\mathbf{z}}$, and *ii*) the same rocket in a uniform gravitational field $\mathbf{g} = -g\hat{\mathbf{z}}$ (e.g. on the surface of the Earth). Imagine shinning a beam of light from one side of the rocket to the other. During this time, the accelerating rocket moves forward, so the light reaches the second wall slightly below the height that it left the first wall (see the figure, if the words aren't clear enough). By the equivalence principle, the same has to be true for the stationary rocket on the surface of the Earth. Light has to bend in a gravitational field. And it does! General relativity's prediction for the bending of light was first confirmed by Eddington, who observed the deflection of starlight passing close to the Sun during a solar eclipse. Since then we have observed many examples of the gravitational bending of light, such as gravitational lensing of the light from entire galaxies by other galaxies along the line-of-sight.

A.1.3 Gravitational Redshift

Consider what happens when we shine the light along the rocket, rather than across (see Fig. A.1). Due to the Doppler effect, the light will reach to top of the rocket at a lower frequency

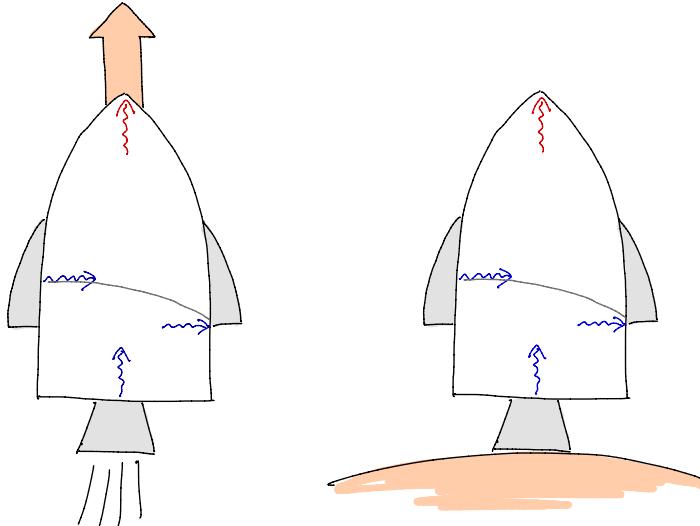


Figure A.1: The equivalence principle relates effects occurring inside a rocket moving with constant acceleration $\mathbf{g} = g\hat{\mathbf{z}}$ (left) to those in a uniform gravitational field $\mathbf{g} = -g\hat{\mathbf{z}}$ (right).

(longer wavelength). By the equivalence principle, the same effect must be observed for light “climbing out” of a gravitational potential well. And it is! The effect was first measured in the Harvard physics department in 1959.

A.1.4 Gravitational Time Dilation

Another observable consequence of the equivalence principle is that time slows down in a gravitational field. To derive this effect, consider the following (thought) experiment. Alice (A) and Bob (B) are at rest in a uniform gravitational field. Their positions are $z_A = h$ and $z_B = 0$ (e.g. at the top and bottom of the rocket in Fig. A.1). At regular intervals $\Delta\tau_A$, Alice sends light signals to Bob. In the insert below, we will show that the proper time between the signals as measured by Bob is

$$\Delta\tau_B \approx \left(1 - \frac{gh}{c^2}\right) \Delta\tau_A. \quad (\text{A.1.8})$$

By the equivalence principle, the same effect must be observed for light in a gravitational field. Noticing that gh is equal to the difference in the gravitational potentials experienced by Alice and Bob, $\Delta\Phi \equiv \Phi_B - \Phi_A$, we can write (A.1.8) as

$$\Delta\tau_B \approx \left(1 + \frac{\Phi_B - \Phi_A}{c^2}\right) \Delta\tau_A. \quad (\text{A.1.9})$$

This shows that time slows in a region of stronger gravity (smaller Φ).

Derivation.—We choose our freely falling frame so that Alice and Bob are at rest at $t = 0$, but accelerate with $\mathbf{a} = g\hat{\mathbf{z}}$. At a later time, the velocities of Alice and Bob are $v(t) = gt$. We will assume that the time taken to perform the experiment is short enough, so that $v/c \ll 1$ and we can neglect special relativistic effects in the dynamics. The trajectories of Alice and Bob then are

$$z_A(t) = h + \frac{1}{2}gt^2, \quad z_B(t) = \frac{1}{2}gt^2. \quad (\text{A.1.10})$$

Let us denote by t_1 the time at which Alice sends out the first signal. The trajectory of the light ray is $z_\gamma(t) \equiv z_A(t_1) - c(t - t_1) = h + \frac{1}{2}gt_1^2 - c(t - t_1)$. The signal will reach Bob at time T_1 when this equals $z_B(T_1) = \frac{1}{2}gT_1^2$, i.e.

$$h + \frac{1}{2}gt_1^2 - c(T_1 - t_1) = \frac{1}{2}gT_1^2. \quad (\text{A.1.11})$$

Alice emits the second signal at time $t_1 + \Delta\tau_A$. The trajectory of this light ray is $z_A(t_1 + \Delta\tau_A) - c(t - t_1 - \Delta\tau_A)$. The signal will reach Bob at $T_1 + \Delta\tau_B$, where

$$h + \frac{1}{2}g(t_1 + \Delta\tau_A)^2 - c(T_1 + \Delta\tau_B - t_1 - \Delta\tau_A) = \frac{1}{2}g(T_1 + \Delta\tau_B)^2. \quad (\text{A.1.12})$$

Subtracting (A.1.11) from (A.1.12), we get

$$c(\Delta\tau_A - \Delta\tau_B) + \frac{1}{2}g\Delta\tau_A(2t_1 + \Delta\tau_A) = \frac{1}{2}g\Delta\tau_B(2T_1 + \Delta\tau_B). \quad (\text{A.1.13})$$

Terms quadratic in $\Delta\tau$ will be negligible as long as $g\Delta\tau \ll c$. In that case, we find

$$c(\Delta\tau_A - \Delta\tau_B) + g\Delta\tau_A t_1 \approx g\Delta\tau_B T_1. \quad (\text{A.1.14})$$

Solving for $\Delta\tau_B$, we obtain

$$\Delta\tau_B = \left(1 + \frac{gT_1}{c}\right)^{-1} \left(1 + \frac{gt_1}{c}\right) \Delta\tau_A \approx \left(1 - \frac{g(T_1 - t_1)}{c}\right) \Delta\tau_A, \quad (\text{A.1.15})$$

where we have dropped terms of order $g^2T_1^2/c^2$ (or higher) in the binomial expansion. At leading order, we have $T_1 - t_1 \approx h/c$ and hence

$$\Delta\tau_B \approx \left(1 - \frac{gh}{c^2}\right) \Delta\tau_A, \quad (\text{A.1.16})$$

which is the desired result (A.1.8).

A.1.5 Curved Spacetime

Consider a spacetime in which the proper time between two nearby events is not given by $c^2d\tau^2 = c^2dt^2 - dx^2$, but by

$$c^2d\tau^2 = \left(1 + \frac{2\Phi(x)}{c^2}\right) c^2dt^2 - \left(1 - \frac{2\Phi(x)}{c^2}\right) dx^2, \quad (\text{A.1.17})$$

with $\Phi \ll c^2$. In these coordinates, Alice sends signals at times t_A and $t_A + \Delta t$, and Bob receives them at t_B and $t_B + \Delta t$. The proper time interval between the signals sent by Alice is

$$\Delta\tau_A^2 = \left(1 + \frac{2\Phi_A}{c^2}\right) \Delta t^2, \quad (\text{A.1.18})$$

where we have used that $\Delta x = 0$ because her signals are sent from the same spatial position. Using $\Phi/c^2 \ll 1$, we can write

$$\Delta\tau_A = \left(1 + \frac{2\Phi_A}{c^2}\right)^{1/2} \Delta t \approx \left(1 + \frac{\Phi_A}{c^2}\right) \Delta t. \quad (\text{A.1.19})$$

Similarly, the proper time between the signals received by Bob is

$$\Delta\tau_B \approx \left(1 + \frac{\Phi_B}{c^2}\right) \Delta t. \quad (\text{A.1.20})$$

Combining (A.1.19) and (A.1.20), we find

$$\Delta\tau_B = \left(1 + \frac{\Phi_B}{c^2}\right) \left(1 + \frac{\Phi_A}{c^2}\right)^{-1} \Delta\tau_A \approx \left(1 + \frac{\Phi_B - \Phi_A}{c^2}\right) \Delta\tau_A, \quad (\text{A.1.21})$$

which is the same as (A.1.9)! The difference in the rates of the two clocks has therefore been explained by the geometry of spacetime. In other words, Newtonian gravity is encoded in the curved spacetime (A.1.17).

In general relativity, gravity will be described by a more general curved spacetime

$$ds^2 \equiv c^2 d\tau^2 = g_{\mu\nu} dX^\mu dX^\nu, \quad (\text{A.1.22})$$

where the metric can depend on both position and time, $g_{\mu\nu}(X) = g_{\mu\nu}(t, \mathbf{x})$, and we have used the Einstein summation convention, in which all repeated indices are summed over. The homogeneous and isotropic spacetime of our universe is described by the *Robertson-Walker metric*

$$ds^2 = c^2 dt^2 - a^2(t) d\mathbf{x}^2, \quad (\text{A.1.23})$$

where the time-dependent scale factor $a(t)$ captures the expansion of the universe.

Indices.—You should be familiar with the manipulation of indices from special relativity. In particular, you should recall that the metric is used to raise and lower indices on four-vectors,

$$A_\mu = g_{\mu\nu} A^\nu \quad \text{and} \quad A^\mu = g^{\mu\nu} A_\nu, \quad (\text{A.1.24})$$

where $g^{\mu\nu}$ is the inverse of $g_{\mu\nu}$. Sometimes, A_μ is called a *covariant vector*, in order to distinguish it from the *contravariant vector* A^μ . A contravariant vector and a covariant vector can be *contracted* to produce a *scalar*,

$$S \equiv A \cdot B = A^\mu B_\mu = g_{\mu\nu} A^\mu B^\nu. \quad (\text{A.1.25})$$

Just as the metric can turn an upper index on a vector into a lower index, the metric can be used to raise and lower indices on *tensors* with arbitrary number of indices. For example, raising the indices on the metric tensor itself leads to

$$g^{\mu\nu} = g^{\mu\alpha} g^{\nu\beta} g_{\alpha\beta}. \quad (\text{A.1.26})$$

Convince yourself that the last relation implies

$$g^{\nu\beta} g_{\beta\alpha} = \delta_\alpha^\nu = \begin{cases} 1 & \text{for } \nu = \alpha \\ 0 & \text{for } \nu \neq \alpha \end{cases}. \quad (\text{A.1.27})$$

In following, we will first discuss how particles move in a curved spacetime (§A.2) and then show how the presence of matter creates the curvature of the spacetime (§A.3).

A.2 Geodesic Equation

In the absence of any non-gravitational forces, particles move along special paths in the curved spacetime. In this section, we will derive an equation that determines these trajectories.

A.2.1 Timelike Geodesics

Let \mathcal{C} be a timelike curve connecting two points p and q . Use a parameter λ to label points along the curve, with $\mathcal{C}(0) = p$ and $\mathcal{C}(1) = q$. The total proper time between p and q , can then be written as

$$\Delta\tau = \int_0^1 d\lambda \sqrt{g_{\mu\nu} \frac{dX^\mu}{d\lambda} \frac{dX^\nu}{d\lambda}}. \quad (\text{A.2.28})$$

A small deformation of a timelike curve remains timelike hence there exist infinitely many timelike curves connecting p and q . The path taken by a massive particle in general relativity is that which extremises the proper time (A.2.28). [Recall that the relativistic action of a massive particle is $S = -m \int d\tau$.] This special curve is called a *geodesic*. Solving the Euler-Lagrange problem, we find an equation satisfied by particles moving along geodesics:

$$\frac{d^2X^\mu}{d\tau^2} + \Gamma_{\alpha\beta}^\mu \frac{dX^\alpha}{d\tau} \frac{dX^\beta}{d\tau} = 0, \quad (\text{A.2.29})$$

where $\Gamma_{\alpha\beta}^\mu$ are called *Christoffel symbols*, and are defined by

$$\boxed{\Gamma_{\alpha\beta}^\mu \equiv \frac{1}{2} g^{\mu\gamma} (\partial_\alpha g_{\gamma\beta} + \partial_\beta g_{\gamma\alpha} - \partial_\gamma g_{\alpha\beta})}. \quad (\text{A.2.30})$$

Derivation.—Let us write (A.2.28) as

$$\Delta\tau[X^\mu(\lambda)] = \int_0^1 d\lambda \left(g_{\mu\nu}(X) \dot{X}^\mu \dot{X}^\nu \right)^{1/2} \equiv \int_0^1 d\lambda L[X^\mu, \dot{X}^\mu], \quad (\text{A.2.31})$$

where $\dot{X}^\mu \equiv dX^\mu/d\lambda$. The geodesic equation for the path of extremal proper time follows from the Euler-Lagrange equation

$$\frac{d}{d\lambda} \left(\frac{\partial L}{\partial \dot{X}^\mu} \right) - \frac{\partial L}{\partial X^\mu} = 0. \quad (\text{A.2.32})$$

The derivatives in (A.2.32) are

$$\frac{\partial L}{\partial \dot{X}^\mu} = -\frac{1}{L} g_{\mu\nu} \dot{X}^\nu, \quad \frac{\partial L}{\partial X^\mu} = -\frac{1}{2L} \partial_\mu g_{\nu\rho} \dot{X}^\nu \dot{X}^\rho. \quad (\text{A.2.33})$$

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

$$\left(\frac{d\tau}{d\lambda} \right)^2 = g_{\mu\nu} \dot{X}^\mu \dot{X}^\nu = L^2, \quad (\text{A.2.34})$$

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

$$\frac{d}{d\tau} \left(g_{\mu\nu} \frac{dX^\nu}{d\tau} \right) - \frac{1}{2} \partial_\mu g_{\nu\rho} \frac{dX^\nu}{d\tau} \frac{dX^\rho}{d\tau} = 0. \quad (\text{A.2.35})$$

Expanding the first term, we get

$$g_{\mu\nu} \frac{d^2 X^\nu}{d\tau^2} + \partial_\rho g_{\mu\nu} \frac{dX^\rho}{d\tau} \frac{dX^\nu}{d\tau} - \frac{1}{2} \partial_\mu g_{\nu\rho} \frac{dX^\nu}{d\tau} \frac{dX^\rho}{d\tau} = 0, \quad (\text{A.2.36})$$

where ∂_ρ is shorthand for $\partial/\partial X^\rho$. 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 (A.2.36) with the inverse metric and relabelling indices, we find

$$\frac{d^2 X^\mu}{d\tau^2} + \Gamma_{\alpha\beta}^\mu \frac{dX^\alpha}{d\tau} \frac{dX^\beta}{d\tau} = 0. \quad (\text{A.2.37})$$

This is the desired result (A.2.29).

Newtonian limit.—Let us use the geodesic equation (A.2.29) to study the dynamics of a massive test particle moving slowly in the spacetime (A.1.17). “Moving slowly” (with respect to the speed of light, $c \equiv 1$) means

$$\frac{dx^i}{dt} \ll 1 \Rightarrow \frac{dx^i}{d\tau} \ll \frac{dt}{d\tau}. \quad (\text{A.2.38})$$

The geodesic equation then becomes

$$\frac{d^2 X^\mu}{d\tau^2} \approx -\Gamma_{00}^\mu \left(\frac{dt}{d\tau} \right)^2. \quad (\text{A.2.39})$$

The relevant component of the Christoffel symbol is

$$\Gamma_{00}^\mu = \frac{1}{2} g^{\mu\gamma} (\partial_0 g_{0\gamma} + \partial_0 g_{0\gamma} - \partial_\gamma g_{00}). \quad (\text{A.2.40})$$

All time derivatives vanish since the metric (A.1.17) is static. Hence, we find

$$\Gamma_{00}^\mu = -g^{\mu j} \partial_j \Phi. \quad (\text{A.2.41})$$

Since the metric is diagonal, $g^{0j} = 0$, we get $\Gamma_{00}^0 = 0$. The $\mu = 0$ component of (A.2.39) therefore becomes

$$\frac{d^2 t}{d\tau^2} = 0 \Rightarrow \frac{dt}{d\tau} = \text{const.} \quad (\text{A.2.42})$$

The $\mu = i$ component of (A.2.39) then reads

$$\frac{d^2 x^i}{d\tau^2} \approx -\Gamma_{00}^i \left(\frac{dt}{d\tau} \right)^2 = -\partial^i \Phi \left(\frac{dt}{d\tau} \right)^2 + \mathcal{O}(\Phi^2). \quad (\text{A.2.43})$$

Dividing by $(dt/d\tau)^2$ and using $dt/d\tau = \text{const.}$, we find

$$\frac{d^2 x^i}{dt^2} \approx -\partial^i \Phi, \quad (\text{A.2.44})$$

which is Newton’s law if we identify Φ with the gravitational potential. Newtonian gravity has therefore been expressed geometrically as geodesic motion in the curved spacetime (A.1.17).

The geodesic equation (A.2.29) can be written in a more elegant form. First, recall that $dX^\mu/d\tau$ is the four-velocity U^μ of the particle, so that (A.2.29) reads

$$\frac{dU^\mu}{d\tau} + \Gamma_{\alpha\beta}^\mu U^\alpha U^\beta = 0. \quad (\text{A.2.45})$$

Using

$$\frac{dU^\mu}{d\tau} = \frac{dX^\alpha}{d\tau} \frac{dU^\mu}{dX^\alpha} = U^\alpha \partial_\alpha U^\mu, \quad (\text{A.2.46})$$

we get

$$U^\alpha \left(\partial_\alpha U^\mu + \Gamma_{\alpha\beta}^\mu U^\beta \right) \equiv U^\alpha \nabla_\alpha U^\mu, \quad (\text{A.2.47})$$

where we have introduced the *covariant derivative* of the four-vector, ∇_α . The geodesic equation therefore takes the following form

$$U^\alpha \nabla_\alpha U^\mu = 0. \quad (\text{A.2.48})$$

In a proper GR course, you would derive this form of the geodesic equation from the concept of “parallel transport”. Finally, note that in terms of the four-momentum $P^\mu = mU^\mu$ the geodesic equation (A.2.48) becomes

$$\boxed{P^\alpha \nabla_\alpha P^\mu = 0} \quad \text{or} \quad P^\alpha \left(\partial_\alpha P^\mu + \Gamma_{\alpha\beta}^\mu P^\beta \right) = 0. \quad (\text{A.2.49})$$

This form of the geodesic equation is particularly convenient since it also applies to massless particles.

A.2.2 Null Geodesics

For massless particles, we can't parameterize the geodesic in terms of proper time since $d\tau = 0$ for null separated points. Instead, it is conventional to define

$$\frac{dX^\mu}{d\tau} \equiv P^\mu, \quad (\text{A.2.50})$$

where $P^\mu = (E, p^i)$ is the four-momentum of the particle. With this definition, the geodesic equation (A.2.49) holds even for massless particles.

Redshift.—Let us apply the geodesic equation to massless particles, e.g. photons, in an expanding universe described by the FRW metric (A.1.23). To study the evolution of the energy of the particles, we consider the $\mu = 0$ component of (A.2.49). Since E is independent of \mathbf{x} (due to the homogeneity of space), we have

$$E \frac{dE}{dt} = -\Gamma_{ij}^0 p^i p^j = -\dot{a}a \delta_{ij} p^i p^j, \quad (\text{A.2.51})$$

where we have used that only the spatial components of $\Gamma_{\alpha\beta}^0$ are non-zero, namely $\Gamma_{ij}^0 = \dot{a}a \delta_{ij}$. For a massless particle, we have

$$0 = g_{\mu\nu} P^\mu P^\nu = E^2 - a^2 \delta_{ij} p^i p^j. \quad (\text{A.2.52})$$

Equation (A.2.51) can therefore be written as

$$\frac{1}{E} \frac{dE}{dt} = -\frac{\dot{a}}{a}. \quad (\text{A.2.53})$$

This equation has the following solution

$$E \propto \frac{1}{a}. \quad (\text{A.2.54})$$

Hence, the energy of a massless particle decreases as the universe expands. This effect is responsible for the redshifting of light, $\lambda \propto a$.

A.3 Einstein Equation

So far, we have described how test particles move in arbitrary curved spacetimes. Next, we want to discuss how the curvature of spacetime is determined by the local matter distribution. We are in search of the following relationship:

$$\begin{pmatrix} \text{a measure of local} \\ \text{spacetime curvature} \end{pmatrix} = \begin{pmatrix} \text{a measure of local} \\ \text{stress-energy density} \end{pmatrix}. \quad (\text{A.3.55})$$

We will start on the left-hand side.

A.3.1 Curvature

It is important to realise that the motion of a single test particle tells us nothing about spacetime curvature. In a frame falling freely with the particle, the test particle remains at rest. Its motion is therefore indistinguishable from that of a test particle in flat spacetime. One test particle is not enough to detect curvature; the motion of at least two particles is needed. In the presence of spacetime curvature the separation between the two particles will change in time.

Tidal Forces

Consider two particles with positions $\mathbf{x}(t)$ and $\mathbf{x}(t) + \mathbf{b}(t)$. Let us first analyse the evolution of these particles in Newtonian gravity. In the presence of a gravitational potential $\Phi(\mathbf{x})$, the separation between the particles satisfies

$$\frac{d^2 b^i}{dt^2} \approx -\delta^{ij} \left(\frac{\partial^2 \Phi}{\partial x^j \partial x^k} \right) b^k. \quad (\text{A.3.56})$$

Equation (A.3.56) holds as long as the separation \mathbf{b} is small relative to the scale of variation of the potential $\Phi(\mathbf{x})$.

Derivation.—In an inertial frame, the equation of motion for the position $\mathbf{x}(t)$ of a particle moving in a gravitational potential $\Phi(\mathbf{x})$ is

$$\frac{d^2 x^i}{dt^2} = -\delta^{ij} \frac{\partial \Phi(x^k)}{\partial x^j}. \quad (\text{A.3.57})$$

The equation of motion of the second particle is

$$\frac{d^2(x^i + b^i)}{dt^2} = -\delta^{ij} \frac{\partial}{\partial x^j} \Phi(x^k + b^k). \quad (\text{A.3.58})$$

If the particles are close to each other, we can Taylor expand on the right-hand side,

$$\frac{\partial \Phi(x^k + b^k)}{\partial x^j} = \frac{\partial \Phi(x^k)}{\partial x^j} + \frac{\partial}{\partial x^k} \left(\frac{\partial \Phi(x^i)}{\partial x^j} \right) b^k + \dots. \quad (\text{A.3.59})$$

Subtracting (A.3.57) from (A.3.58), we get (A.3.56).

We notice the important role in (A.3.56) played by the *tidal (acceleration) tensor*

$$\frac{\partial^2 \Phi}{\partial x^i \partial x^j}. \quad (\text{A.3.60})$$

It determines the forces that pull nearby particles apart or bring them closer together. The trace of the tidal tensor appears in the *Poisson equation*,

$$\nabla^2 \Phi = \delta^{ij} \left(\frac{\partial^2 \Phi}{\partial x^i \partial x^j} \right) = 4\pi G \rho. \quad (\text{A.3.61})$$

We would like to generalise this equation to GR.

Geodesic Deviation

Let us first find the equivalent of (A.3.56) in GR. The algebra will be a bit more involved. Try not to get scared, the physics is the same as in the Newtonian example. The analog of the tidal tensor will give us a local measure of spacetime curvature.

We consider two particles separated by a four-vector B^μ . We wish to determine how B^μ evolves with respect to the proper time τ of an observer comoving with one of the particles. If the observer is freely falling, i.e. its four-velocity is $U^\mu = (1, 0, 0, 0)$, then the geodesic equation implies

$$\frac{d^2 B^\mu}{d\tau^2} = -R^\mu_{0\alpha 0} B^\alpha, \quad (\text{A.3.62})$$

where $R^\mu_{0\beta 0}$ are components of the *Riemann tensor*

$$R^\alpha_{\beta\gamma\delta} = \partial_\gamma \Gamma^\alpha_{\beta\delta} - \partial_\delta \Gamma^\alpha_{\beta\gamma} + \Gamma^\alpha_{\gamma\epsilon} \Gamma^\epsilon_{\beta\delta} - \Gamma^\alpha_{\delta\epsilon} \Gamma^\epsilon_{\beta\gamma}. \quad (\text{A.3.63})$$

Derivation.—Let $X^\mu(\tau)$ and $X^\mu(\tau) + B^\mu(\tau)$ be two nearby geodesics. To find the evolution of the separation vector B^μ , it is necessary to compute its second derivative with respect to the observer's proper τ . This is a bit non-trivial. Instead of ordinary derivatives with respect to τ , we need to consider covariant derivatives along the geodesics. For example, the derivative of $B^\mu(X)$ evaluated on the observer's worldline $X^\mu(\tau)$ is

$$\begin{aligned} \frac{DB^\mu}{D\tau} &\equiv U^\beta \nabla_\beta B^\mu = \frac{dX^\beta}{d\tau} \nabla_\beta B^\mu \\ &= \frac{dB^\mu}{d\tau} + \Gamma^\mu_{\beta\gamma} U^\beta B^\gamma. \end{aligned} \quad (\text{A.3.64})$$

Similarly, the second derivative along the geodesic is

$$\begin{aligned} \frac{D^2 B^\mu}{D\tau^2} &\equiv U^\alpha \nabla_\alpha (U^\beta \nabla_\beta B^\mu) \\ &= \frac{d}{d\tau} \left(\frac{dB^\mu}{d\tau} + \Gamma^\mu_{\beta\gamma} U^\beta B^\gamma \right) + \Gamma^\mu_{\alpha\delta} U^\alpha \left(\frac{dB^\delta}{d\tau} + \Gamma^\delta_{\beta\gamma} U^\beta B^\gamma \right). \end{aligned} \quad (\text{A.3.65})$$

To manipulate the complex expression in (A.3.65), we use

$$\frac{d^2 B^\mu}{d\tau^2} = -2\Gamma^\mu_{\alpha\beta} U^\alpha \frac{dB^\beta}{d\tau} - \partial_\gamma \Gamma^\mu_{\alpha\beta} B^\gamma U^\alpha U^\beta + \mathcal{O}(B^2), \quad (\text{A.3.66})$$

$$\frac{dU^\mu}{d\tau} = -\Gamma^\mu_{\alpha\beta} U^\alpha U^\beta, \quad (\text{A.3.67})$$

$$\frac{d\Gamma^\mu_{\alpha\beta}}{d\tau} = U^\gamma \partial_\gamma \Gamma^\mu_{\alpha\beta}, \quad (\text{A.3.68})$$

where (A.3.66) follows from subtracting the geodesic equations of the two particles and expanding the right-hand side to first order in B^μ . Notice that all time derivatives of B^μ cancel in (A.3.65). If you keep your head straight, you will then find

$$\frac{D^2 B^\mu}{D\tau^2} = -R^\mu{}_{\beta\gamma\delta} U^\beta U^\delta B^\gamma, \quad (\text{A.3.69})$$

with $R^\alpha{}_{\beta\gamma\delta}$ as defined in (A.3.63). Equation (A.3.69) is called the *geodesic deviation equation*. For a freely falling observer, $U^\mu = (1, 0, 0, 0)$, it reduces to (A.3.62).

We see that the Riemann tensor plays the role of the tidal tensor. In fact, for the metric (A.1.17), we have

$$R^i{}_{0j0} = \frac{\partial \Gamma^i_{00}}{\partial x^j} = \delta^{ik} \frac{\partial^2 \Phi}{\partial x^k \partial x^j}. \quad (\text{A.3.70})$$

Hence, we recover the Newtonian limit (A.3.59). The “trace” of the Riemann tensor defines the *Ricci tensor*:

$$R_{\mu\nu} \equiv g^{\alpha\beta} R_{\alpha\mu\beta\nu} = \partial_\alpha \Gamma_{\mu\nu}^\alpha - \partial_\nu \Gamma_{\mu\alpha}^\alpha + \Gamma_{\alpha\beta}^\alpha \Gamma_{\mu\nu}^\beta - \Gamma_{\nu\beta}^\alpha \Gamma_{\mu\alpha}^\beta. \quad (\text{A.3.71})$$

The relativistic generalisation of $\nabla^2 \Phi = 0$ is the Einstein equation in vacuum,

$$R_{\mu\nu} = 0. \quad (\text{A.3.72})$$

Next, we will show how this equation needs to be extended to take into account how the curvature of spacetime is sourced by the energy and momentum of matter.

A.3.2 Energy-Momentum

In relativity, the energy and momentum densities of a continuous distribution of matter are components of the energy-momentum tensor:

$$T_{\mu\nu} = \left(\begin{array}{c|c} T_{00} & T_{0i} \\ \hline T_{i0} & T_{ij} \end{array} \right) = \left(\begin{array}{c|c} \text{energy density} & \text{energy flux} \\ \hline \text{momentum density} & \text{stress tensor} \end{array} \right). \quad (\text{A.3.73})$$

This is a natural object to appear on the right-hand side of the Einstein equation. An important property of the energy-momentum tensor is that it is locally conserved, in the sense that

$$\nabla^\mu T_{\mu\nu} = 0. \quad (\text{A.3.74})$$

The conservation of the energy and momentum densities correspond to the $\nu = 0$ and $\nu = i$ components, respectively. To work with (A.3.74), we need to unpack the meaning of the covariant derivative, which we will do in the following insert.

Covariant derivative.—So far, we have only encountered the covariant derivative of a contravariant vector (i.e. a four-vector with upper index):

$$\nabla_\mu A^\nu = \partial_\mu A^\nu + \Gamma_{\mu\alpha}^\nu A^\alpha. \quad (\text{A.3.75})$$

In this insert, we will derive the action of the covariant derivative on general tensors:

- First, we note that there is no difference between the covariant derivative and the partial derivative if it acts on a scalar

$$\nabla_\mu f = \partial_\mu f. \quad (\text{A.3.76})$$

- To determine how the covariant derivative acts on a covariant vector, B_ν , let us consider how it acts on the scalar $f \equiv B_\nu A^\nu$. Using (A.3.76), we can write this as

$$\begin{aligned} \nabla_\mu (B_\nu A^\nu) &= \partial_\mu (B_\nu A^\nu) \\ &= (\partial_\mu B_\nu) A^\nu + B_\nu (\partial_\mu A^\nu). \end{aligned} \quad (\text{A.3.77})$$

Alternatively, we can also write

$$\begin{aligned} \nabla_\mu (B_\nu A^\nu) &= (\nabla_\mu B_\nu) A^\nu + B_\nu (\nabla_\mu A^\nu) \\ &= (\nabla_\mu B_\nu) A^\nu + B_\nu (\partial_\mu A^\nu + \Gamma_{\mu\alpha}^\nu A^\alpha), \end{aligned} \quad (\text{A.3.78})$$

where we have used (A.3.75) in the second equality. Comparing (A.3.77) and (A.3.78), we get

$$(\nabla_\mu B_\nu) A^\nu + B_\nu \Gamma_{\mu\alpha}^\nu A^\alpha = (\partial_\mu B_\nu) A^\nu. \quad (\text{A.3.79})$$

Writing $B_\nu \Gamma_{\mu\alpha}^\nu A^\alpha$ as $B_\alpha \Gamma_{\mu\nu}^\alpha A^\nu$, this gives

$$(\nabla_\mu B_\nu) A^\nu = (\partial_\mu B_\nu - \Gamma_{\mu\nu}^\alpha B_\alpha) A^\nu. \quad (\text{A.3.80})$$

Since the vector A^ν is arbitrary, the factors multiplying it on each side must be equal, so we get

$$\boxed{\nabla_\mu B_\nu = \partial_\mu B_\nu - \Gamma_{\mu\nu}^\alpha B_\alpha}. \quad (\text{A.3.81})$$

Notice the change of the sign of the second term relative to (A.3.75) and the placement of the dummy index.

- The covariant derivative of the mixed tensor $T^\mu{}_\nu$, can be derived similarly by considering $f \equiv T^\mu{}_\nu A^\nu B_\mu$. This gives

$$\boxed{\nabla_\sigma T^\mu{}_\nu = \partial_\sigma T^\mu{}_\nu + \Gamma_{\sigma\alpha}^\mu T^\alpha{}_\nu - \Gamma_{\sigma\nu}^\alpha T^\mu{}_\alpha}. \quad (\text{A.3.82})$$

Writing out the covariant derivative in (A.3.74), we therefore find

$$\begin{aligned} 0 &= \nabla^\mu T_{\mu\nu} = \nabla_\mu T^\mu{}_\nu \\ &= \partial_\mu T^\mu{}_\nu + \Gamma_{\mu\alpha}^\mu T^\alpha{}_\nu - \Gamma_{\mu\nu}^\alpha T^\mu{}_\alpha. \end{aligned} \quad (\text{A.3.83})$$

Perfect fluid.—In cosmology, the coarse-grained energy-momentum tensor takes the form of a perfect fluid:

$$T_{\mu\nu} = (\rho + P) U_\mu U_\nu - P g_{\mu\nu}, \quad (\text{A.3.84})$$

where U^μ is the *four-velocity* of the fluid, and ρ and P are its *energy density* and the *pressure* as measured by a comoving observer (i.e. an observer with four-velocity U^μ). The energy-momentum tensor measured by a free-falling observer takes the following simple form

$$T_{\mu\nu} = \text{diag}(\rho, +a^2 P, +a^2 P, +a^2 P), \quad (\text{A.3.85})$$

$$T^\mu{}_\nu \equiv g^{\mu\alpha} T_{\alpha\nu} = \text{diag}(\rho, -P, -P, -P). \quad (\text{A.3.86})$$

Using the FRW metric (A.1.23), the $\nu = 0$ component of (A.3.83) then leads to

$$\frac{d\rho}{dt} = -3\frac{\dot{a}}{a}(\rho + P). \quad (\text{A.3.87})$$

This equation determines how the energy densities of the different matter components that fill the universe evolve with time.

A.3.3 Einstein Equation

Einstein's first guess for the field equation of GR was

$$R_{\mu\nu} \stackrel{?}{=} \kappa T_{\mu\nu}, \quad (\text{A.3.88})$$

where κ is a constant. However, this doesn't work because, in general, we can have $\nabla^\mu R_{\mu\nu} \neq 0$, which wouldn't be consistent with $\nabla^\mu T_{\mu\nu} = 0$. When we add matter, the Ricci tensor $R_{\mu\nu}$ isn't the only measure of curvature. An alternative curvature tensor, made from $R_{\mu\nu}$ and $g_{\mu\nu}$, is

$$G_{\mu\nu} \equiv R_{\mu\nu} + \lambda g_{\mu\nu} R, \quad (\text{A.3.89})$$

where $R = R^\mu_\mu = g^{\mu\nu} R_{\mu\nu}$ is the Ricci scalar, and λ is a constant that still needs to be determined. To use $G_{\mu\nu}$ in a field equation of the form $G_{\mu\nu} = \kappa T_{\mu\nu}$, we require $\nabla^\mu G_{\mu\nu} = 0$. This Bianchi identity is satisfied iff $\lambda = -\frac{1}{2}$. The combination in (A.3.89) is then called the *Einstein tensor*. The constant of proportionality between $G_{\mu\nu}$ and $T_{\mu\nu}$ is fixed by matching to the Newtonian limit (A.3.61). This gives $\kappa = 8\pi G$ and hence the final form of the Einstein equation is

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} \quad \text{or} \quad R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi G T_{\mu\nu}. \quad (\text{A.3.90})$$

Note that if $T_{\mu\nu} = 0$, then $0 = g^{\mu\nu}G_{\mu\nu} = -R$, and the Einstein equation reduces to the vacuum form (A.3.72).

Newtonian limit.—The exotic appearance of the Einstein equation should not obscure the fact that it is a natural extension of Newtonian gravity. To illustrate this, consider the metric (A.1.17). At linear order in Φ , the Ricci tensor is

$$R_{00} = \nabla^2 \Phi, \quad R_{0i} = 0, \quad R_{ij} = \nabla^2 \Phi \delta_{ij}. \quad (\text{A.3.91})$$

and the Ricci scalar is

$$R = -2\nabla^2 \Phi. \quad (\text{A.3.92})$$

The 00-component of the Einstein equation

$$\begin{aligned} G_{00} &= R_{00} - \frac{1}{2}g_{00}R = 2\nabla^2 \Phi \\ &= 8\pi G T_{00} = 8\pi G \rho, \end{aligned} \quad (\text{A.3.93})$$

therefore equals the *Poisson equation*

$$\nabla^2 \Phi = 4\pi G \rho. \quad (\text{A.3.94})$$

This confirms that the Einstein equation includes Newtonian gravity in the appropriate limit.

Friedmann equation.—The Ricci tensor associated with the metric (A.1.23) is

$$R_{00} = -3 \frac{\ddot{a}}{a}, \quad R_{0i} = 0, \quad R_{ij} = (2\dot{a}^2 + \ddot{a}a) \delta_{ij}. \quad (\text{A.3.95})$$

and the Ricci scalar is

$$R = -6 \left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a} \right)^2 \right]. \quad (\text{A.3.96})$$

The 00-component of the Einstein equation therefore is

$$G_{00} = R_{00} - \frac{1}{2}g_{00}R = 3 \left(\frac{\dot{a}}{a} \right)^2 \quad (\text{A.3.97})$$

$$= 8\pi G T_{00} = 8\pi G \rho. \quad (\text{A.3.98})$$

Defining the *Hubble parameter*, $H \equiv \partial_t \ln a$, this becomes the *Friedmann equation*

$$H^2 = \frac{8\pi G}{3} \rho. \quad (\text{A.3.99})$$

This is the key evolution equation in cosmology.