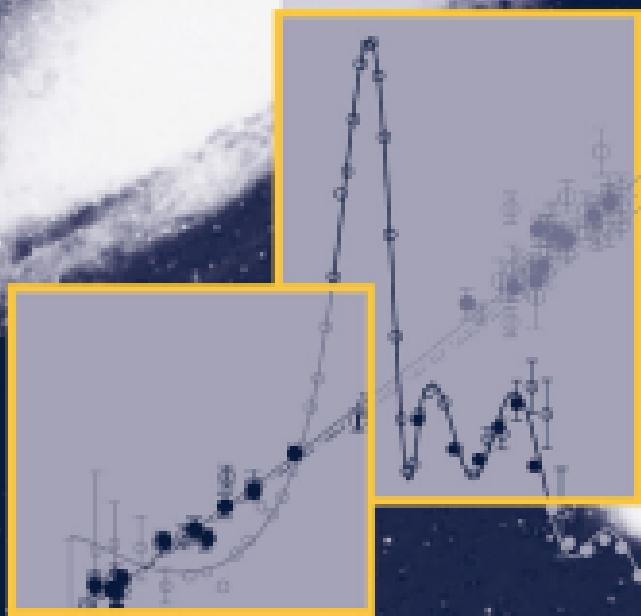


James Rich

Fundamentals of Cosmology

2nd Edition



Springer



POLYTECHNIQUE

Fundamentals of Cosmology



ÉCOLE POLYTECHNIQUE

The Ecole Polytechnique, one of France's top academic institutions, has a long-standing tradition of producing exceptional scientific textbooks for its students. The original lecture notes, the *Cours de l'Ecole Polytechnique*, which were written by Cauchy and Jordan in the nineteenth century, are considered to be landmarks in the development of mathematics.

The present series of textbooks is remarkable in that the texts incorporate the most recent scientific advances in courses designed to provide undergraduate students with the foundations of a scientific discipline. An outstanding level of quality is achieved in each of the seven scientific fields taught at the *Ecole*: pure and applied mathematics, mechanics, physics, chemistry, biology, and economics. The uniform level of excellence is the result of the unique selection of academic staff there which includes, in addition to the best researchers in its own renowned laboratories, a large number of world-famous scientists, appointed as part-time professors or associate professors, who work in the most advanced research centers France has in each field.

Another distinctive characteristic of these courses is their overall consistency; each course makes appropriate use of relevant concepts introduced in the other textbooks. This is because each student at the Ecole Polytechnique has to acquire basic knowledge in the seven scientific fields taught there, so a substantial link between departments is necessary.

The distribution of these courses used to be restricted to the 900 students at the Ecole. Some years ago we were very successful in making these courses available to a larger French-reading audience. We now build on this success by making these textbooks also available in English.

James Rich

Fundamentals of Cosmology

Second Edition

With 124 Figures



Prof. James Rich
SPP-IRFU
CEA-Saclay
91191 Gif-sur-Yvette
France
james.rich@cea.fr

**Instructors and lecturers can download a solutions manual from
<http://www.springer.com/astronomy/general+relativity/book/978-3-642-02799-4>**

ISBN 978-3-642-02799-4 e-ISBN 978-3-642-02800-7
DOI 10.1007/978-3-642-02800-7
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2009937485

© Springer-Verlag Berlin Heidelberg 2010

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: eStudio Calamar S.L.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This is the second edition of a book on the phenomenological foundations of modern cosmology. The first edition was surprisingly well timed. In the preceding 3 years, advances in astronomy had laid the foundations of the dark-energy/dark-matter cosmological model. In 1998, observations of type Ia supernova by the Supernova Cosmology Project and High Z team suggested that the expansion of the universe is accelerating. In 2000, Boomerang and Maxima observed clearly the first acoustic peak in the anisotropy spectrum of the cosmic microwave background. The position of the peak strongly suggested that the density of the universe was near critical. These two observations reconciled popular inflationary models with the heretofore embarrassing insistence by astronomers that the clumping of galaxies required that the mass density of the universe be only about 25% of the critical density. The supernova data told us that the rest was dark energy in the form of an effective vacuum energy or cosmological constant. The so-called Lambda CDM model became the standard model.

Eight years later, the observational data has improved tremendously as a comparison of the figures in the two editions will demonstrate. Surprisingly, all this incredible data has failed to find any clear failure of the ΛCDM model.

Among the observational advances, the most obvious comes from the WMAP satellite who, along the ground-based experiments like ACBAR, yielded the exquisite series of peaks in the CMB temperature anisotropy spectrum shown in Fig. 7.11. WMAP and the QUAD experiment have also seen clearly CMB polarization with the peaks nicely in place. The Supernova Legacy Survey (SNLS) and Essence programs have provided precision supernovae data figure that confirmed the conclusions of the original pioneering experiments. The Sloan Digital Sky Survey (SDSS) redshift and photometric surveys have revolutionized many fields in astronomy. For cosmology, one thinks of their measurements of galaxy–galaxy lensing (Fig. 2.14) that have created a new tool for understanding how dark matter congregates to form structures. But perhaps the biggest observational advance was the SDSS galaxy–galaxy correlation function (Fig. 5.7) showing a “BAO” peak at the position of the acoustic horizon at recombination. Besides yielding a theoretically clean measurement of the mass density, it provided a visually remarkable image of sound waves propagating in the early universe.

For a textbook, these observational advances have clarified many issues that were troublesome 9 years ago. In the previous edition, arguments for a subcritical mass density relied on astrophysical reasoning dealing with ways of hiding baryons or biases in galaxy distributions. The interpretation of supernova data required discussion of the possibility of “gray dust” or luminosity evolution. All these questions remain, and will be important issues for “precision cosmology.” However, they no longer call into question the framework of Λ CDM.

With the precise agreement between observations and the predictions of the standard Λ CDM model, it is hard to imagine how any radically different approach could describe well the data. Models that replace cold-dark matter with modified gravitational laws still seem to work (quite well) at galactic scales but one would be surprised to see them generate the CMB anisotropy spectrum shown in Figure 7.11. Models that explain the apparent acceleration with gravitational “back-reaction” effects of ordinary matter raise interesting questions of principle but it seems unlikely that such an effect could imitate so well a cosmological constant.

This suggests that some form of dark energy will be a permanent part of cosmology. Searches for differences between dark-energy parameters those expected for a standard cosmological constant will be an important part of cosmological research in the next decades. Searches will be pursued for weakly interacting dark matter in its various proposed forms. A discovery of supersymmetry CERN LHC would provide strong encouragement.

So should we be happy with Λ CDM? Perhaps, but the observed proportion of dark energy and dark matter makes a strange mixture if one believes that it should be a direct consequence of the “laws of physics.” People who believe that what we see is only a small part of a much vaster reality are more comfortable with Λ CDM since the observed mixture is manifestly one that permits the existence of observers and it is easy to imagine others that don’t. The last few years have seen a certain grudging acceptance of anthropic selection as perhaps necessary for understanding cosmology. This has received some encouragement from superstring theory which appears to provide us with a “Landscape” of possibilities sprinkled around the “multiverse.” This is perhaps the biggest revolution in cosmology, though one that clearly goes beyond the normal limits of falsifiable science. It will be interesting to see if such ideas are still around 50 years from now.

This new version has made great use of the increasingly common practice of making observational data available on the Internet, thus allowing me to make figures that are pedagogically appropriate. The CAMB package (<http://camb.info/>), expertly piloted by Christophe Magneville, was used to generate the theoretical transfer functions and anisotropy spectra. Thanks finally to Jean-Baptiste Melin, Bruce Miller, Jean-Louis Basdevant, and Jacques Haissinski for comments on various versions of this book.

James Rich
October 22, 2009

Contents

1	Introduction	1
1.1	The Composition of the Universe	3
1.1.1	The Visible Universe: Galaxies	3
1.1.2	Baryons	10
1.1.3	Cold Dark Matter	11
1.1.4	Photons	13
1.1.5	Neutrinos	14
1.1.6	Dark Energy	16
1.2	The Evolution of the Universe	18
1.2.1	The Scale Factor $a(t)$	18
1.2.2	Gravitation and the Friedmann Equation	20
1.2.3	Open and Closed Universes	22
1.2.4	The Evolution of the Temperature	24
1.2.5	An Improved Friedmann Equation	29
1.2.6	The Evolution of the Ω s and Structure Formation	31
1.2.7	The Standard Scenario	33
1.3	Open Questions	35
	Exercises	39
2	Observational Cosmology	41
2.1	Stars and Quasi-stars	41
2.2	Galaxies	53
2.3	Galaxy Clusters	60
2.4	Large-Scale Structure	62
2.5	Dark Matter	64
2.5.1	WIMPs	64
2.5.2	Axions	68
2.5.3	MACHOs	68
2.5.4	Cold Gas	71
2.6	The Cosmological Parameters	73
2.6.1	H_0	73
2.6.2	ρ s and Ω s	78
	Exercises	80

3 Coordinates and Metrics	91
3.1 Relativity and Gravitation	94
3.2 Comoving Coordinates	100
3.3 The Metric I: Mostly Isotropy	102
3.4 The Metric II: Mostly Homogeneity	105
3.5 Photon Propagation	110
3.6 Observable Distances	113
3.7 The Geodesic Equation	116
3.8 Gravitational Lensing	119
Exercises	130
4 The Field Equations	137
4.1 Our Freely Falling Coordinates	138
4.2 A Universe with $\rho = 0$	142
4.3 The Energy-Momentum Tensor	143
4.4 The Friedmann Equation	148
4.5 The Cosmological Parameters	150
4.6 Scalar Fields	152
4.7 The Riemann Tensor	154
4.8 The Einstein Tensor	157
4.9 The General Einstein Equation	158
Exercises	162
5 Friedmannology	165
5.1 The Age of the Universe	167
5.2 Type Ia Supernovae as Standard Candles	170
5.3 The Sound Horizon as a Standard Ruler	175
5.4 The Horizon Problem	181
5.5 The Ω Problem	186
5.6 Inflation	188
5.7 Intergalactic Scattering and Absorption	191
Exercises	193
6 The Thermal History of the Universe	197
6.1 Equilibrium Distributions	200
6.2 The Boltzmann Equation	204
6.3 Electrons and Positrons	209
6.4 Neutrinos	215
6.5 Primordial Nucleosynthesis	216
6.6 WIMPs	225
6.7 Baryogenesis	227
6.8 Irreversibility	229
6.9 The Future	231
Exercises	232

7 Structure Formation	239
7.1 A Spherical Collapse Model	242
7.1.1 A CDM Universe with Scale Invariant Fluctuations	246
7.1.2 Effects of Baryons and Photons	247
7.2 The Spectrum of Density Fluctuations	248
7.2.1 The Power Spectrum for a CDM-Baryon Universe	252
7.2.2 Measurements of the Power Spectrum	256
7.2.3 The Correlation Function	258
7.2.4 Redshift Distortions	258
7.3 The Primordial Spectrum from Inflation	260
7.4 CMB Temperature Anisotropies	264
7.4.1 The Sources of Anisotropy	267
7.4.2 The Cosmological Parameters	272
7.5 CMB Polarization	274
7.6 Alternatives to Λ CDM	278
7.7 Newtonian Evolution	280
7.8 Photon Propagation	283
Exercises	288
A Lorentz Vectors and Tensors	295
B Natural Units	299
C Standard Particles and Beyond	303
D Magnitudes	309
E Solutions and Hints for Selected Exercises	311
F Useful Formulas and Numbers	317
References	321
Index	325

Chapter 1

Introduction

Cosmology is the study of the global characteristics of the universe. In its details, the cosmos is rather complicated with a multitude of fascinating objects ranging from carbonaceous dust grains to quasars. In spite of this, cosmologists like to think of the universe as a simple place characterized by the following:

- A tiny mean density, $\rho \sim 10^{-26} \text{ kg m}^{-3}$.
- The curious “ Λ CDM” mixture of substances given in Table 1.1. Most of the energy is in the form of “dark energy” (an effective vacuum energy or cosmological constant, Λ) and “cold dark matter” (CDM) made of yet-to-be-discovered weakly interacting particles. Ordinary “baryonic” matter (protons and nuclei plus electrons) makes up only 4.5% of the energy. On the other hand, most particles are “Cosmic Microwave Background” (CMB) photons or neutrinos.
- A chemical state out of thermal equilibrium characterized by a deficit of neutrinos (compared to photons) and a deficit of highly bound heavy nuclei (compared to hydrogen and helium).
- A hierarchy of gravitationally bound structures ranging from planets and stars to galaxies and clusters of galaxies.

Most importantly, we note the observation that is the basis for modern cosmology:

- The universe is “expanding” in the sense that the distances between galaxies increase with time. The present universe has thus evolved from an earlier state characterized by high density and temperature.

And finally, a tentative conclusion based on observations of the last decade:

- The expansion is now accelerating.

This conclusion is tentative only because it is so surprising: one would expect that the attractive effects of normal gravitation would *decelerate* the expansion. The role of the hypothesized dark energy is to make the acceleration possible.

The primary aim of observational cosmology is to characterize the thermal, chemical, and structural state of the present universe. The existence of the components of the Λ CDM universe, listed in Table 1.1 were originally inferred from a

Table 1.1 The occupants of the universe in the Λ CDM model assuming a mixture of photons (CMB), neutrinos, baryons, cold dark matter (CDM), and dark energy (DE) in the form of an effective vacuum energy or cosmological constant. For each species, i , the table gives estimated number density of particles, n_i , and the estimated mass or energy density, $\Omega_i = \rho_i/\rho_c$, normalized to the “critical density,” $\rho_c = 0.92 h_{70}^2 \times 10^{-26} \text{ kg m}^{-3}$. Some of the estimated densities depend on the numerical value of the Hubble constant H_0 resulting in the factors of $h_{70} = H_0/(70 \text{ km s}^{-1} \text{ Mpc}^{-1})$. The estimates for Ω_b , Ω_{CDM} , Ω_A , and Ω_T are those of the WMAP [1] collaboration using their CMB anisotropy data combined with supernova data [2, 3] and baryon acoustic oscillation (BAO) data [4]. The CMB densities come from the FIRAS instrument on the COBE satellite [5]. The lower limit on Ω_V comes for the lower limit on the heaviest neutrino mass from the Kamioka atmospheric oscillation data [6]. The upper limit on Ω_V is the limit from WMAP [1]. Also shown are the estimates of the present expansion rate, H_0 , and derived quantities given by WMAP [1]

Species	$n_i (\text{m}^{-3})$	$\Omega_i = \rho_i/\rho_c$
CMB photons	$n\gamma = (4.11 \pm 0.02) \times 10^8$	$\Omega_\gamma = 5.06 h_{70}^{-2} \times 10^{-5}$
ν_e, ν_μ, ν_τ	$n\nu = 3 \times (3/11)n\gamma$ (sum of ν and $\bar{\nu}$)	$0.0004h_{70}^{-2} < \Omega_V < 0.015h_{70}^{-2}$
Baryons	$n_b = \eta n_\gamma$ $\eta = (6.18 \pm 0.2) \times 10^{-10}$	$\Omega_b = 0.0456 \pm 0.0015$ ($\sim 75\%$ H, $\sim 24\%$ He)
Cold dark matter (CDM)	?	$\Omega_{\text{CDM}} = 0.228 \pm 0.013$
Dark energy (DE)	0	$\Omega_A = 0.726 \pm 0.015$
Total		$-0.0179 < 1 - \Omega_T < 0.0081$
Expansion rate (Hubble constant)		$H_0 = (70.5 \pm 1.3) \text{ km sec}^{-1} \text{ Mpc}^{-1}$ $(h_{70} = 1.01 \pm 0.02)$
Expansion time (Hubble time)		$t_H \equiv H_0^{-1} = (13.87 \pm 0.25) \text{ Gyr}$
“Age” of universe		$t_0 = t_H f(\Omega_M, \Omega_A) = (13.72 \pm 0.12) \text{ Gyr}$
Hubble distance		$d_H \equiv c/H_0 = (4250 \pm 8) \text{ Mpc}$
Critical density		$\rho_c \equiv \frac{3H_0^2}{8\pi G} = (0.93 \pm 0.04) \times 10^{-26} \text{ kg m}^{-3}$ = $(5.2 \pm 0.2) (\text{GeV}/c^2) \text{ m}^{-3}$ = $(1.36 \pm 0.05) \times 10^{11} M_\odot \text{ Mpc}^{-3}$

variety of measurements ranging from the observed fluxes from distant supernovae to the spatial correlations between galaxy positions. However, the last 5 years has seen the emergence of the study of the CMB photons as the dominant tool for measuring precisely the densities of each component. The CMB photons have a nearly perfect thermal distribution with a mean temperature of $T_\gamma = 2.725 \pm 0.002 \text{ K}$, measured by the COBE satellite in 1990 [5]. This thermal distribution is due to the fact that in the early universe, the photons electron and baryons formed an ionized gas in thermal equilibrium. Tiny variations in photon temperature, of order $10^{-5} T_\gamma$, are observed depending on the direction of observation. These variations are believed to be due to acoustic oscillations of the photon-electron-baryon fluid in gravitational potential wells created by the CDM. Characterization of the CMB anisotropies by the WMAP satellite [1] has thus allowed one to give the precise estimations of the quantities of all components given in Table 1.1.

The aim of theoretical cosmology is to explain the present state of the universe in terms of the conditions in the “early universe.” Non-controversial laws of physics

allow us to say with confidence that the expansion has been proceeding at least since an epoch when the density was 40 orders of magnitude greater than at present, with a temperature of $kT > 10 \text{ MeV}$. Going back in time, we would see stars and galaxies melt into a uniform plasma of elementary particles. We will see in Chap. 6 that the chemical composition in Table 1.1 and the thermal non-equilibrium will, in part, be explained by the reactions between particles that took place at temperatures near 1 MeV. In particular, the neutrino–photon ratio n_ν/n_γ will be calculated to high precision and the helium–hydrogen ratio $n_{\text{He}}/n_{\text{H}}$ will be calculated as a function of the baryon–photon ratio $\eta = n_b/n_\gamma$. The other ratios, including η itself, remain to be explained.

The observed hierarchy of gravitationally bound structures is believed to have resulted from the gravitational growth of small density inhomogeneities present in the early universe. The origin of the initial inhomogeneities is the subject of speculation but observations are consistent with them being due to quantum fluctuations of a scalar field in an early “inflationary” epoch that *ended* when the universe was a factor of $\sim 10^{120}$ denser than at present. The production of these seeds and their subsequent growth into the structures that are now present will be described in Chap. 7.

In this introductory chapter, we will start in Sect. 1.1 with a qualitative description of the components of the ΛCDM universe listed in Table 1.1. In Sect. 1.2, we will then introduce the standard cosmological model where the evolution of the universe is determined by gravitation via the Friedmann equation and by statistical mechanics via the Boltzmann equation.

The observational basis of the standard model will be presented in more detail in Chap. 2. The gravitational foundations will be presented in Chaps. 3 and 4, allowing us, in Chap. 5, to understand the measurements of the cosmological densities. The statistical mechanics necessary to understand the chemical evolution of the universe will be discussed in Chap. 6. Finally, the basics of structure formation theory will be introduced in Chap. 7.

1.1 The Composition of the Universe

1.1.1 *The Visible Universe: Galaxies*

The “building blocks” of the visible universe are galaxies (Figs. 1.1 and 1.2) which are themselves composed of stars, interstellar gas, and unidentified dark matter. Galaxies have a wide variety of shapes (spirals, ellipticals, and irregulars), masses, and luminosities. Nevertheless, most of the light in the universe is produced in galaxies containing $10^{10}\text{--}10^{11}$ stars that generate a typical galactic luminosity of

$$\langle L_{\text{gal}} \rangle \sim 2 \times 10^{10} L_\odot, \quad (1.1)$$

where L_\odot is the solar luminosity, $L_\odot = 2.4 \times 10^{45} \text{ eV s}^{-1}$.



Fig. 1.1 The spiral galaxy NGC1232 viewed by the ESO VLT. Courtesy of the European Southern Observatory [7]

Galactic masses are ill-defined because studies of the rotation velocities of galactic clouds (Sect. 2.2) indicate that most of the mass is contained in dark, roughly spherical, halos (Fig. 1.3). The mass of the visible regions M_{vis} is mostly contained in stars and is observed to be roughly proportional to the galactic luminosity, L . Typically, the “mass-to-light” ratio is of order $3M_{\odot}$ per L_{\odot}

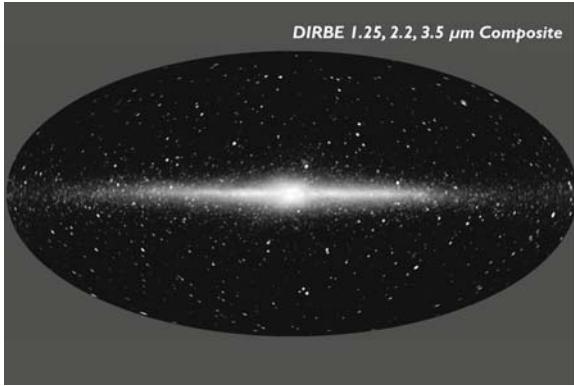


Fig. 1.2 The Milky Way as viewed from our position 8 kpc from the Galactic center. The figure is an all-sky composite image using data taken by the Diffuse Infrared Background Experiment (DIRBE) instrument on the Cosmic Background Explorer (COBE) in the 1.25, 2.2, and 3.5 micrometer wavelength bands. Most of the emission at these wavelengths is from cool, low-mass K stars in the disk and bulge of the Milky Way. Interstellar dust does not strongly obscure emission at these wavelengths; the maps trace emission all the way through the Galaxy. Courtesy of the NASA Goddard Space Flight Center and the COBE Science Working Group [8]

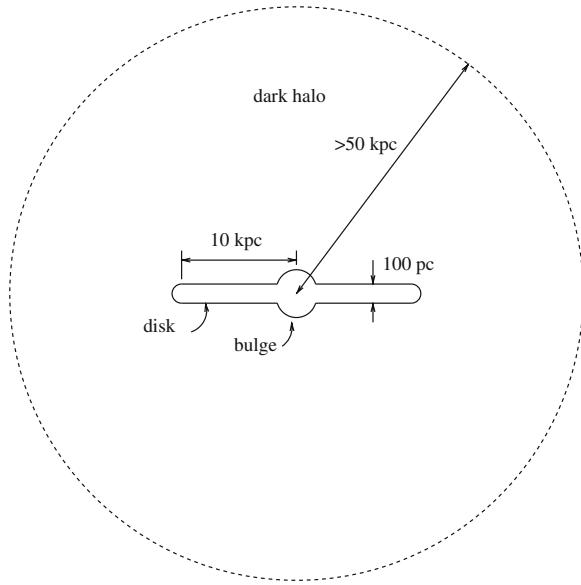


Fig. 1.3 A schematic of a typical spiral galaxy viewed edge-on. The visible zone is in the form of a disk of radius ~ 10 kpc and thickness ~ 100 pc. Most of the mass is believed to be in a spherical “halo.” The density of the halo is roughly $\rho \propto R^{-2}$

$$\langle M_{\text{vis}}/L \rangle \sim 3(M_{\odot}/L_{\odot}), \quad (1.2)$$

where $M_{\odot} = 1.988 \times 10^{30}$ kg is the solar mass. This implies a “visible mass” per galaxy of typically

$$\langle M_{\text{vis}} \rangle \sim 6 \times 10^{10} M_{\odot}. \quad (1.3)$$

The radius of the visible regions of spiral galaxies is typically

$$\langle R_{\text{vis}} \rangle \sim 10 \text{ kpc}, \quad (1.4)$$

where we have introduced the “parsec,” the distance unit most commonly used in cosmology: 1 pc = 3.26 light-year = 3.086×10^{16} m.

Galaxies are not uniformly distributed in space, a fact that is not surprising in view of their mutual gravitational attraction. Galaxies are often grouped in bound clusters, the largest of which contain thousands of galaxies. In spite of this “small-scale” inhomogeneity, at large scales > 100 Mpc the universe appears to be rather uniform with a number density of bright galaxies of

$$n_{\text{gal}} \sim 0.015 \text{ Mpc}^{-3} \quad (1.5)$$

corresponding to a typical intergalactic distance of ~ 4 Mpc. It should be emphasized that this density corresponds to that of “bright” galaxies. The number density

appears to diverge logarithmically at small galactic luminosities which makes the number of galaxies ill-defined. These small galaxies contribute little to the total light output of the universe which has a value of

$$J \sim \langle L_{\text{gal}} \rangle n_{\text{gal}} \sim 10^8 L_{\odot} \text{Mpc}^{-3}. \quad (1.6)$$

The total mass density associated with the visible parts of galaxies is

$$\rho_{\text{vis}} = J \langle M/L \rangle \sim 3 \times 10^8 M_{\odot} \text{Mpc}^{-3} \quad (1.7)$$

Modern cosmology started with Hubble's discovery that galaxies are receding from us with a (recession) velocity dR/dt proportional to their distance R (Fig. 1.4):

$$\frac{dR}{dt} = H_0 R + v_{\text{peculiar}}. \quad (1.8)$$

The factor of proportionality, H_0 , is called the "Hubble constant." Its inverse, H_0^{-1} , gives the characteristic time for significant changes in R , and H_0 can therefore be called the current "expansion rate" of the local universe. The Hubble constant should be distinguished from the "Hubble parameter," $H(t)$, which gives the expansion rate as a function of time.¹

As indicated in (1.8), in addition to the "Hubble velocity," $H_0 R$, galaxies have quasi-random "peculiar" velocities that are typically of order $10^{-3}c \sim 300 \text{ km s}^{-1}$.

Hubble's law (1.8) applies to galaxies that are "near" so that $v \ll c$. In this case, a photon's flight time from the galaxy to us is sufficiently small that we can neglect the variation of R during the flight. To find the "relativistic" generalization (3.59) of Hubble's law, applicable to more distant galaxies, we will have to be much more rigorous in defining what we mean by "distance." This will be done in Chap. 3.

Because of the difficulty in measuring galactic distances, estimates of H_0 have changed by over an order of magnitude since Hubble's original work. The recent measurements from the Hubble Key Project [9] shown in Fig. 1.4 give

$$H_0 = 72 \pm 8 \text{ km s}^{-1} \text{Mpc}^{-1}. \quad (1.9)$$

This value of H_0 corresponds to the local expansion rate measured with nearby galaxies. Techniques leading to this value will be summarized in Sect. 2.6.1. The value deduced by the WMAP [1]

$$H_0 = 70.5 \pm 1.3 \text{ km s}^{-1} \text{Mpc}^{-1}, \quad (1.10)$$

corresponds to a global expansion rate. Its value is consistent with the local value.

¹ Following the usual convention, the subscript 0 denotes the present value of a quantity.

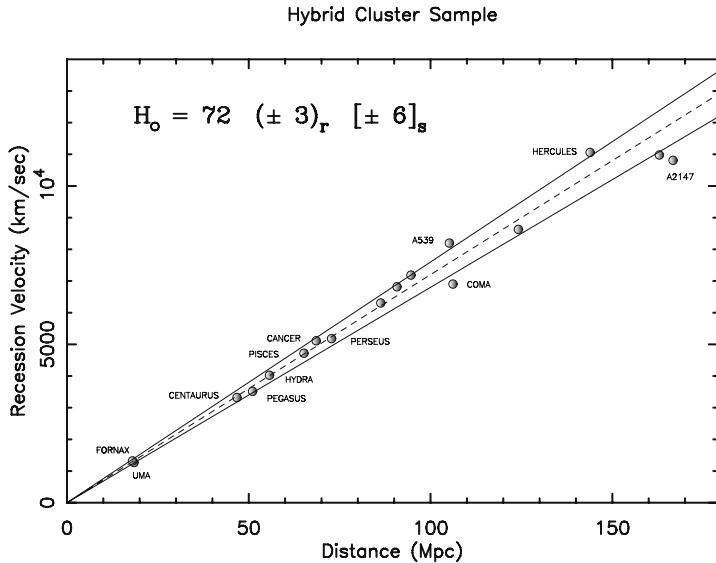


Fig. 1.4 The “Hubble diagram,” of galactic recession velocities versus distance for a set of galaxy clusters as determined by the Hubble Key Project [9]. The velocities are determined by the galaxy redshifts and the distances are determined by a variety of methods discussed in Sect. 2.6.1. The slope of the line is the Hubble constant H_0

Since H_0 is omnipresent in cosmological formulas, it is useful for numerical results to define $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$:

$$H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1} \quad h = 0.705 \pm 0.013. \quad (1.11)$$

While h is widely used in the literature, in this book we prefer to use $h_{70} = H_0/(70 \text{ km s}^{-1} \text{ Mpc}^{-1})$:

$$H_0 = 70 h_{70} \text{ km s}^{-1} \text{ Mpc}^{-1} \quad h_{70} = 1.01 \pm 0.02, \quad (1.12)$$

which will give directly numerical factors in line with present measurements of H_0 .

To measure H_0 we need galactic distances and recession velocities. The latter are easily measured from the “redshift,” z , of the galactic spectral lines:

$$z \equiv \frac{\lambda_0}{\lambda_1} - 1, \quad (1.13)$$

where λ_1 is the photon wavelength that would be measured by an observer in the rest frame of the emitting galaxy and λ_0 is the wavelength measured by us. We can

interpret the redshift as being due to the Doppler effect² of the recession velocity. For nearby galaxies, the recession velocity is much less than the speed of light so we can use the non-relativistic Doppler formula, $(\lambda_0 - \lambda_1)/\lambda_1 = v/c$. Equation (1.13) becomes

$$z \sim c^{-1} \frac{dR}{dt} \quad (z \ll 1). \quad (1.14)$$

Substituting this into (1.8) and ignoring the peculiar velocity, we see that Hubble's law can be written as

$$z = \frac{H_0}{c} R = \frac{R}{d_H} \quad z \ll 1, \quad (1.15)$$

where d_H is the “Hubble distance”:

$$d_H = c H_0^{-1} = 3000 h^{-1} \text{ Mpc} = 4300 h_{70}^{-1} \text{ Mpc}. \quad (1.16)$$

It follows that $z \ll 1$ if $R \ll d_H$. Equation (1.15) allows us to estimate a galaxy's distance from its redshift, $R \sim zd_H$ ($z \ll 1$). This estimate is accurate only if the peculiar velocity is negligible.

In this book we will often neglect to write the “ c ”:

$$d_H = H_0^{-1}. \quad (1.17)$$

This is our first example of the use of “natural units” where c , \hbar , and the Boltzmann constant k are “set equal to 1.” As discussed in Appendix B, we lose no information by neglecting to write explicitly the factors of c and \hbar since they can always be recovered by dimensional analysis.

We can list three important implications of Hubble's law:

1. The universe is dynamic. In the past, it was denser and, as we will see, hotter. Imprudent commentators extrapolate to a moment of infinite density called the “big bang” or “primordial singularity.” It is clear that, since we do not know the laws of physics at infinite density, such an extrapolation cannot be performed with confidence.³ We note, however, that if we use today's recession velocities

² Readers uncomfortable with the Doppler interpretation should not worry—it is only an interpretation. What is really important are relations between observables, e.g. redshifts and photon fluxes. In particular, we observe that galaxies that are faint (because they are distant) are redshifted. It is thus the redshift-apparent luminosity relation that has real significance. This relation will be established in Sect. 3.5.

³ We will see that in order to extrapolate we need to know, amongst other things, the pressure as a function of energy density. We have no empirical guidance on this point at temperatures above ~ 1 GeV.

to extrapolate, a time H_0^{-1} has elapsed since the putative singularity. This time is called the “Hubble time”:

$$t_H = H_0^{-1} = 1.39 \times 10^{10} h_{70}^{-1} \text{ yr}. \quad (1.18)$$

In Chap. 5 we will see that the Hubble time gives the order of magnitude of the elapsed time since the beginning of the epoch when the “known laws of physics apply,” i.e. since the time when the temperature was, say, $\sim 1\text{GeV}$. We also note that in one Hubble time light can travel a distance of $c t_H = d_H$. One might therefore suspect that there is a “horizon” at a distance of the order of the Hubble distance. Beyond this horizon, objects are not yet visible because their light has not yet reached us. This suspicion will be confirmed in Chap. 5.

- As illustrated in Fig. 1.5, the linearity of Hubble’s law means that an observer in a neighboring galaxy will see a universal expansion with the same H_0 . This is true only for a linear relation between recession velocity and distance and any other law would imply that we occupy a privileged position at the “center” of the universe. We are encouraged to conclude that the universe is homogeneous in our neighborhood in the sense that all observers attached to galaxies see the same Hubble law and the same density of matter averaged over sufficiently large volumes. While there is good evidence for this homogeneity within our horizon, the homogeneity of the entire universe is only a matter of conjecture, sometimes raised to the status of the “Cosmological Principle.” This principle is now somewhat out of fashion since in many speculative theories, e.g. chaotic inflation and string-theory landscape, the universe is very inhomogeneous at super-horizon scales.

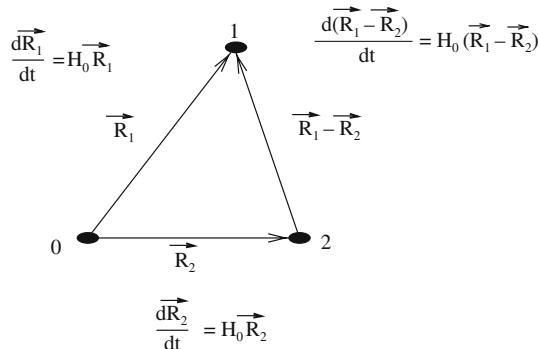


Fig. 1.5 If an observer in galaxy 0 sees a universal expansion with rate H_0 , an observer in galaxies 1 or 2 will see a universal expansion with the same rate H_0 . This is because, if we neglect peculiar velocities, Hubble’s law is $d\mathbf{R}_i/dt = H_0 \mathbf{R}_i$ for the position \mathbf{R}_i of any galaxy i . This implies that $d(\mathbf{R}_i - \mathbf{R}_j)/dt = H_0 (\mathbf{R}_i - \mathbf{R}_j)$ for any pair of galaxies i and j meaning that any galaxy can be taken to be the “center of the universe”

3. With H_0 and Newton's constant $G = G_N$, it is possible to form a quantity with the dimensions of mass (or energy) per unit volume called the "critical density":

$$\rho_c = \frac{3H_0^2}{8\pi G} = 0.91 h_{70}^2 \times 10^{-26} \text{ kg m}^{-3} \quad (1.19)$$

$$= 1.34 h_{70}^2 \times 10^{11} M_\odot \text{ Mpc}^{-3} = 0.51 h_{70}^2 \times 10^{10} \text{ eV m}^{-3}. \quad (1.20)$$

The second line indicates that the critical density corresponds to about one galaxy/Mpc³ or about 5 protons/m³, much greater than the observed densities of galaxies or protons. It turns out that if there is no effective cosmological constant, the expansion of a homogeneous universe with $\Omega_T > 1$ will eventually stop and be followed by an epoch of contraction ending (if we dare extrapolate to infinite density) with a "big crunch." In Chap. 4, we will also see that a homogeneous universe with a supercritical density has a finite volume. If there is a cosmological constant, this correlation between geometry and destiny is broken, but ρ_c still has a role as the "natural unit" of density.

It is generally convenient to normalize cosmological densities to the critical density to form an " Ω ." For example, the total mean density ρ_T divided by ρ_c is Ω_T :

$$\Omega_T \equiv \frac{\rho_T}{\rho_c}. \quad (1.21)$$

The density of visible matter (1.7) divided by ρ_c gives

$$\Omega_{\text{vis}} = \frac{\rho_{\text{vis}}}{\rho_c} \sim 0.002 \quad (\text{vis} = \text{visible stars}). \quad (1.22)$$

We see that if there was only visible matter, the universe would be sub-critical with $\Omega_T \sim 0.002$.

We emphasize that quantities like " ρ_T " and " Ω_T " refer to present values. Values at other times will be denoted by " $\rho_T(t)$ " and " $\Omega_T(t)$." The critical density also depends on time and to avoid confusion we will generally write it out explicitly as $3H_0^2/8\pi G$ for the present epoch or as $3H(t)^2/8\pi G$ at other epochs.

1.1.2 Baryons

The total density of baryonic matter (protons, nuclei, and electrons⁴) is estimated to be an order of magnitude greater than that of visible baryons (1.22):

$$\Omega_b = (0.0456 \pm 0.015). \quad (1.23)$$

⁴ Only protons and neutrons are "baryons" but we include the electrons in "baryonic" matter.

This estimate comes from the WMAP [1] analysis of CMB anisotropies that will be discussed in Chap. 7. However, the original estimates came from the theory of the nucleosynthesis of the light elements [10] which correctly predicts the *relative* abundances of the light elements only if Ω_b is near this value. To briefly summarize the more detailed discussion of Chap. 6, the nuclear composition of the universe changes with time as the stars transform their hydrogen into helium and then into heavier elements. Nevertheless, there appears to be a “primordial” mix of nuclei consisting of about 75% hydrogen (by mass) and 25% ^4He along with traces of ^2H , ^3He , and ^7Li . This mixture is approximately the observed mixture in certain locations relatively unpolluted by stellar nucleosynthesis. The primordial abundances were determined by nuclear reactions that took place when the universal temperature was $\sim 60 \text{ keV}$. Primordial nucleosynthesis calculations predict the abundances as a function of the total baryon density at $T \sim 60 \text{ keV}$, since it is this density that determines the nuclear reaction rates. We will see in Chap. 6 that the predicted nuclear abundances agree reasonably well with the observed abundances if the current baryon density has the value given by (1.23).

The number density of baryons is considerably less than that of photons:

$$n_b = \eta n_\gamma \quad \eta = (6.18 \pm 0.2) \times 10^{-10}. \quad (1.24)$$

Note that since there are four baryons per helium nucleus but only two electrons, the number density of electrons is less than that of baryons:

$$n_e \sim 0.87 n_b \quad (1.25)$$

Since $\Omega_b > \Omega_{\text{vis}}$ one can wonder where the missing “dark” baryons are. Most of them are thought to be in the intergalactic medium in the form of an ionized gas [11]. Some fraction of them may be in dark compact objects such as dead stars (neutron stars or white dwarfs) or stars too light to burn hydrogen (brown dwarfs). It has also been suggested [12, 13] that a significant fraction of the baryons are contained in cold molecular clouds.

Finally, we mention that there are apparently very few antibaryons in the visible universe [14]. Any antimatter consisting of antibaryons and positrons would quickly annihilate in collisions with ordinary matter. Even if the antimatter were somehow separated from the matter, annihilations in intergalactic space at the boundaries between matter and antimatter domains would lead to a flux of high-energy annihilation photons higher than the observed flux from other sources. It thus seems probable that the density of antimatter is negligible within our horizon (Exercise 6.7).

1.1.3 Cold Dark Matter

Galaxies and galaxy clusters were formed by gravitational collapse of non-relativistic matter. The theory of this process will be outlined in Chap. 7. One of the main results

is that it is difficult to understand how baryons of an amount given by (1.23) could have created the observed structures. The basic problem stems from the fact that the photons of cosmological origin (Sect. 1.1.4) are observed to have an energy spectrum that is nearly independent of the direction of observation. We will see that this implies that the baryons were extremely homogeneous at early times. Going from this homogeneous state to the present inhomogeneous state is difficult if there are only baryons because they suffer from their tight coupling (through electron-photon Compton scattering) with the homogeneous photons.

Popular models of structure formation generally assume the existence of dark matter in the form of particles that have been only weakly interacting and non-relativistic since the epoch of $T \sim MeV$. This matter is generically called “cold dark matter” (CDM). CDM particles are called “cold” because they are “slow” (non-relativistic). Often, they are assumed to be in the form of non-baryonic weakly interacting massive particles called generically “WIMPs.” The fact that they are only weakly interacting allows them to gravitate freely without being inhibited by interactions with other particles (baryons and photons).

The present density of CDM is estimated to be a factor of five greater than that of baryons but still less than critical:

$$\Omega_{\text{CDM}} = 0.228 \pm 0.013 . \quad (1.26)$$

This dark matter is believed to make up most of the mass of galactic halos and galaxy clusters.

Unfortunately, there are no known WIMPs in the zoo of elementary particles (Appendix C) and their existence is a bold prediction of cosmology. Some extensions of the standard model of particle physics predict the existence of WIMPs that are sufficiently heavy that they would not yet have been produced at accelerators. Originally, it was proposed that a new heavy neutrino with $m_\nu > 1 \text{ GeV}$ could be the CDM but this possibility was excluded [15] by a combination of accelerator results and direct searches (Sect. 2.5).

A more popular, but speculative, class of models that predict the existence of WIMPs are “supersymmetric” models. In these models, each of the known fermions (bosons) is paired with a heavy supersymmetric partner that is a boson (fermion). The lightest of the supersymmetric partners (LSP) is expected to be stable and to have only weak interactions, making it an ideal WIMP candidate. In Chap. 6 we will see that the parameters of the supersymmetric model can be chosen so that the WIMP has the required present-day density (1.27). The mass would be expected to be between 10 GeV and 10 TeV. Efforts are underway to detect supersymmetric particles at accelerators and in the Milky Way (Sect. 2.5).

Other candidates for non-baryonic dark matter are light neutrinos and primordial black holes [16]. We will see that neutrinos with the density (1.27) would have masses near 10 eV. As discussed in Chap. 7, such light particles have difficulties in forming the observed structures because they were relativistic when $T \sim MeV$ and constitute “hot dark matter.” Primordial black holes work well in structure

formation but cosmologists lack a convincing scenario for their production in the early universe [16].

The total density of non-relativistic matter is the sum of the densities of CDM, baryons, and non-relativistic neutrinos :

$$\Omega_M = \Omega_{CDM} + \Omega_b + \Omega_V = 0.27 \pm 0.01 . \quad (1.27)$$

(One generally supposes $\Omega_V \ll \Omega_{CDM}$.) In many applications, e.g. in the calculations of cosmological distances and time scales in Chap. 5, Ω_M is the important parameter.

1.1.4 Photons

The most abundant particles in the universe are the photons of the “cosmic microwave background” (CMB). These photons have a nearly perfect thermal spectrum as shown in Fig. 1.6. The photon temperature is $T_\gamma = 2.725 \text{ K} = 2.35 \times 10^{-4} \text{ eV}$ corresponding to a number density of $n_\gamma = 411 \text{ cm}^{-3}$. This is considerably greater than the number of photons that have been generated by stars (Exercise 2.2). Despite

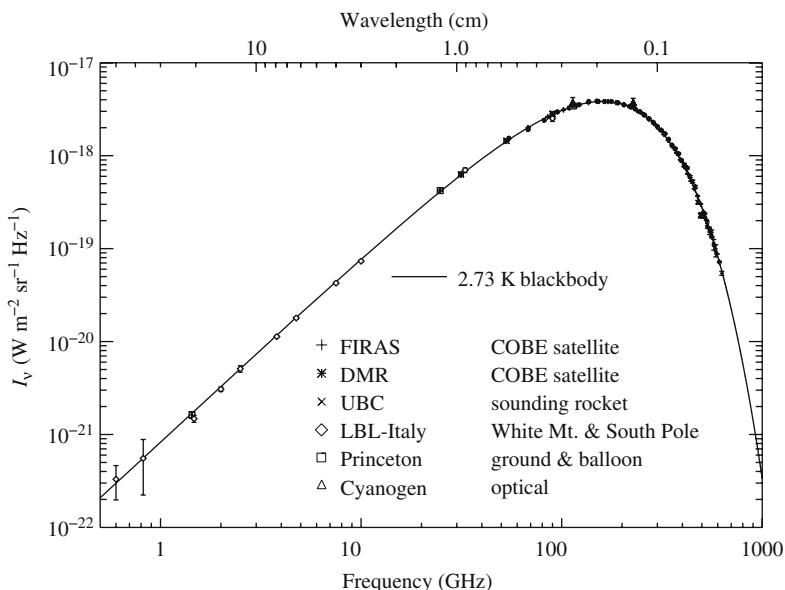


Fig. 1.6 The observed spectrum of the cosmic (microwave) background radiation (CMB) [17]. The points at wavelengths $< 1 \text{ cm}$ come from ground-based experiments. At shorter wavelengths the Earth's atmosphere is opaque and measurements must be made from balloons, rockets, or satellites. The high precision points around the peak of the spectrum were made by the FIRAS instrument of the COBE satellite which observed from 1989 to 1995 [5]. Compilation courtesy of the Particle Data Group

their great abundance, the low temperature of the CMB results in a small photon energy density:

$$\Omega\gamma = 5.04 h_{70}^{-2} \times 10^{-5} . \quad (1.28)$$

The present-day universe is nearly transparent to photons (Exercise 2.3). In Chap. 6, we will see that the CMB photons were thermalized in the early universe when the temperature was > 0.26 eV and baryonic matter was completely ionized. At $T \sim 0.26$ eV, baryonic matter “recombined” to form atoms and the resulting decrease in the photon-matter cross-section made the universe transparent.

As already emphasized, the CMB temperature is not completely isotropic but is observed to vary by factors of order 10^{-5} according to the direction of observation. These small variations are believed to be due to the density inhomogeneities present at the moment of recombination. As such, the temperature anisotropies provide information about the “initial conditions” for structure formation. As we will see in Chap. 7, the spectrum of anisotropies when interpreted within the framework of the Λ CDM model allows the determination of the densities in Table 1.1.

1.1.5 Neutrinos

In addition to thermal photons, it is believed that the universe is filled with neutrinos, ν_e , ν_μ , and ν_τ and the corresponding antineutrinos. The three neutrino “flavors” should be identified with the neutrino created in interactions with the associated charged leptons e , μ , and τ . For instance, the ν_e is created in β decay in association with an electron or positron:

$$(A, Z) \rightarrow (A, Z \mp 1) e^\pm \nu_e (\bar{\nu}_e) , \quad (1.29)$$

where A is the number of nucleons and Z is the number of protons in the nucleus. The ν_μ is created in pion decay in association with a muon:

$$\pi^\pm \rightarrow \mu^\pm \nu_\mu (\bar{\nu}_\mu) . \quad (1.30)$$

The ν_τ is created in τ decay, e.g.:

$$\tau^\pm \rightarrow \pi^\pm \bar{\nu}_\tau (\nu_\tau) . \quad (1.31)$$

It is suspected that the neutrinos of definite “flavor”, ν_e , ν_μ , and ν_τ , are, in fact, linear combinations (in the quantum-mechanical sense) of the neutrinos of definite mass, ν_1 , ν_2 , and ν_3 :

$$\nu_f = \sum_{i=1}^3 \alpha_{fi} \nu_i \quad (1.32)$$

for $f = e, \mu, \tau$.

Neutrinos interact even less than the CMB photons but they had a sufficiently high interaction rate at $T > 1 \text{ MeV}$ to have been thermalized. In Chap. 6 we will see that for temperatures $T < m_e$, relativistic neutrinos have a temperature slightly less than the photon temperature:

$$T_\nu = (4/11)^{1/3} T_\gamma . \quad (1.33)$$

It turns out that this corresponds to a neutrino (+ antineutrino) energy density per relativistic species of

$$\rho_\nu = (7/8) \times (4/11)^{4/3} \rho_\gamma = 0.227 \rho_\gamma \quad \text{per relativistic species} \quad (1.34)$$

and to a neutrino (+ antineutrino) number density of

$$n_\nu = (3/11)n_\gamma \quad \text{per species.} \quad (1.35)$$

Even if the neutrinos are now non-relativistic, this last relation still holds so we expect $n_\nu = 112 \text{ cm}^{-3}$ per species today.

For an effectively massless neutrino species, i.e. $m_\nu \ll T_\nu$, the summed neutrino and antineutrino contribution of that species to the energy density is even less than that of CMB photons:

$$\Omega_\nu = 1.14 h_{70}^{-2} \times 10^{-5} \quad \text{if } m_\nu \ll 10^{-4} \text{ eV} . \quad (1.36)$$

For a species of mass greater than the calculated temperature, the neutrinos are currently non-relativistic and the summed neutrino and antineutrino mass density is

$$\Omega_\nu = \frac{m_\nu n_\nu}{\rho_c} = 0.022 h_{70}^{-2} \frac{m_\nu}{1 \text{ eV}} \quad \text{if } m_\nu \gg 10^{-4} \text{ eV} . \quad (1.37)$$

If one of the neutrino species has a mass in the eV range it would contribute significantly to Ω_M and we need $m_\nu < 10 \text{ eV}$ to have $\Omega_\nu < \Omega_M$. A stricter limit comes from the lack of any observed effect of a light neutrino on the spectrum of density and CMB fluctuations. This gives (Table 1.1) $\Omega_\nu < 0.015$, i.e. $m_\nu < 0.7 \text{ eV}$. This limit is stricter than the limits on the neutrino masses that come from kinematic studies of the decays (1.29), (1.30) and (1.31) [17] : $m_1 < 10 \text{ eV}$, $m_2 < 170 \text{ keV}$ and, $m_3 < 18.2 \text{ MeV}$, where $i = 1, 2, 3$ correspond to the neutrinos of definite mass most closely aligned with ν_e , ν_μ , and ν_τ .

Evidence for non-zero neutrino masses comes from searches for “neutrino oscillations”, i.e. the transformation of given flavor neutrino into a different flavor. For example the production of a ν_μ

$$\pi^+ \rightarrow \mu^+ \nu_\mu, \quad (1.38)$$

might be followed at some distance by the neutrino acting as a ν_τ , with the reaction

$$\nu p \rightarrow \tau^- p \pi^+. \quad (1.39)$$

These experiments are, unfortunately, only sensitive to *differences* in the squares of neutrino masses.⁵ Recent observation of oscillations of neutrinos produced in the atmospheric interactions of cosmic rays [6] have given results that are most easily interpreted as

$$m_3^2 - m_2^2 = (2.5 \pm 0.5) \times 10^{-3} \text{ eV}^2 \quad \Rightarrow m_3 > 0.04 \text{ eV}. \quad (1.40)$$

Anomalies in the spectrum of solar neutrinos and reactor neutrinos can be explained by [17]

$$m_2^2 - m_1^2 \sim (8.0 \pm 0.3) \times 10^{-5} \text{ eV}^2 \quad \Rightarrow m_2 > 0.008. \quad (1.41)$$

The lower limit on m_3 and the upper limit from the lack of effects on the density inhomogeneities combine to give

$$0.0009 h_{70}^{-2} < \Omega_V < 0.015. \quad (1.42)$$

Finally, we note that, because of their extremely weak interactions, there is little hope of directly detecting the cosmic neutrino background [18].

1.1.6 Dark Energy

Certainly the most surprising recent discovery is that the universe appears to be dominated by “dark energy,” i.e. an apparent “vacuum energy” or “cosmological constant” Λ :

$$\Omega_\Lambda \sim \frac{\Lambda}{3H_0^2} = 0.726 \pm 0.015. \quad (1.43)$$

Vacuum energy is, by definition, energy that is not associated with particles and is therefore not diluted by the expansion of the universe. Unless the present vacuum

⁵ The differences determine the “oscillation length,” i.e. the characteristic distance traveled before a neutrino starts to interact like a neutrino of a different flavor.

is only metastable, this implies the vacuum energy density is independent of time. The value implied by $\Omega_\Lambda = 0.726$ is

$$\rho_\Lambda(t) = 3.75 h_{70}^2 \times 10^9 \text{ eV m}^{-3}. \quad (1.44)$$

The original observational evidence for the existence of such an energy will be discussed in Chap. 5. It involves the apparent luminosity of high-redshift objects which can provide information on whether the universal expansion is accelerating or decelerating (as would be expected from normal gravitation). The observations indicate that the expansion is accelerating and, as we will see in Chap. 4, this can be explained by a positive vacuum energy density.

Fundamental physics cannot currently be used to calculate the value of the vacuum energy even though it is a concept used throughout modern gauge theories of particle physics. It is expected to be a temperature-dependent quantity which changes in a calculable manner during phase transitions, e.g. the electroweak transition at $T \sim 300 \text{ GeV}$ when the intermediate vector bosons, W^\pm and Z^0 , became massive. Unfortunately, all calculable quantities involving vacuum energy concern differences in energy densities and there are no good ideas on how to calculate the absolute value.

Despite the lack of ideas, the existence of a vacuum energy density of the magnitude given by (1.44) is especially surprising. In natural units, an energy density has the dimension of the fourth power of mass, so a vacuum energy density can be associated with a mass scale M

$$\rho_\Lambda \sim \frac{M^4}{(\hbar c)^3}. \quad (1.45)$$

Particle physicists are tempted to choose the Planck mass $m_{\text{pl}} = (\hbar c^5/G)^{1/2} \sim 10^{19} \text{ GeV}$ as the most fundamental scale giving $\rho_\Lambda \sim 3 \times 10^{132} \text{ eV m}^{-3}$. This is 123 orders of magnitude too large making it perhaps the worst guess in the history of physics. In fact, the density (1.44) implies a scale of $M \sim 10^{-3} \text{ eV}$ which is not obviously associated with any other fundamental scale in particle physics, though it is near the estimated masses of the neutrinos.

A second problem with an energy density (1.43) is that it is comparable to the matter density $\Omega_M \sim 0.3$. Since the matter density changes with the expansion of the universe while the vacuum energy does not, it appears that we live in a special epoch when the two energies are comparable. This problem will be further discussed in Sect. 1.3.

Finally, we note that the observational evidence for a strictly constant vacuum energy density is also consistent with a component that varies little over a Hubble time. In this general context, instead of speaking of vacuum energy one speaks of “dark energy” of density ρ_{de} . The present time derivative of ρ_{de} is characterized by the parameter w :

$$\frac{\dot{\rho}_{\text{de}}}{\rho_{\text{de}}} = \frac{3(1+w)}{t_{\text{H}}}, \quad (1.46)$$

with $w = -1$ corresponding to a constant vacuum energy. We will see in Chap. 4 that w is the ratio between the dark energy pressure and the dark energy density. Present measurements [1] give

$$w = -1 \pm 0.1 \quad (1.47)$$

meaning that ρ_{de} varies by less than 30% over a Hubble time. Models with $w \neq -1$ generally involve hypothetical slowly varying scalar fields corresponding to a substance generically called “Quintessence” [19].

1.2 The Evolution of the Universe

1.2.1 The Scale Factor $a(t)$

It is useful to parameterize the expansion of the universe by a time-dependent function that is proportional to the distances between galaxies. This function is called the “scale factor,” $a(t)$:

$$a(t) \propto \langle \text{intergalactic distances} \rangle. \quad (1.48)$$

Hubble’s law informs us that the logarithmic derivative of $a(t)$ is currently equal to H_0 :

$$\left[\frac{\dot{a}}{a} \right]_{t_0} = H_0 \quad t_0 \equiv \text{today}. \quad (1.49)$$

The number density of galaxies n_{gal} is proportional to $\langle \text{distances} \rangle^{-3}$ so, if the number of galaxies does not change with time, the definition (1.48) can be replaced with

$$a(t) \equiv a_0 \left(\frac{n_{\text{gal}}(t_0)}{n_{\text{gal}}(t)} \right)^{1/3} \quad a_0 \equiv a(t_0). \quad (1.50)$$

While the number of galaxies is not strictly conserved (there were none in the early universe), the current limit on the proton lifetime ($\tau_p > 10^{32}$ yr) means that baryon number is conserved to a very high precision. It is thus even better to define $a(t)$ in terms of the mean baryonic number density:

$$a(t) \equiv a_0 \left(\frac{n_b(t_0)}{n_b(t)} \right)^{1/3}. \quad (1.51)$$

During the present epoch, there are few antibaryons in the observable universe and n_b can be taken to be simply the number density of baryons. In the early universe at temperatures $T > \text{GeV}$, antibaryons were present in a plasma of quarks and anti-quarks. At this epoch, n_b should be taken to be the net baryon number density, i.e. the number density of baryons minus the number density of antibaryons.

The present value of the scale parameter a_0 is, for most applications, an arbitrary constant that can be given the dimensions of distance. For this reason, we will often use the dimensionless “reduced scale factor”

$$\hat{a}(t) \equiv \frac{a(t)}{a_0} . \quad (1.52)$$

However, for certain applications involving high-redshift objects, we will need the dimensioned scale factor and in Chap. 4 we will propose a definition in terms of the Hubble distance $d_H = H_0^{-1}$:

$$a_0 = \frac{H_0^{-1}}{\sqrt{|1 - \Omega_T|}} = \frac{d_H}{\sqrt{|1 - \Omega_T|}} . \quad (1.53)$$

We will see in Chap. 3 that the singularity at $\Omega_T = 1$ causes no problems.

The scale factor will be essential in Chap. 3 when we define “comoving” coordinates. As illustrated in Fig. 1.7, if peculiar velocities are neglected, the distance $R_i(t)$ between us and the galaxy i is proportional to $a(t)$:

$$R_i = a(t) \chi_i . \quad (1.54)$$

The constant of proportionality, χ_i , is time independent and called the radial “comoving” coordinate of the galaxy.

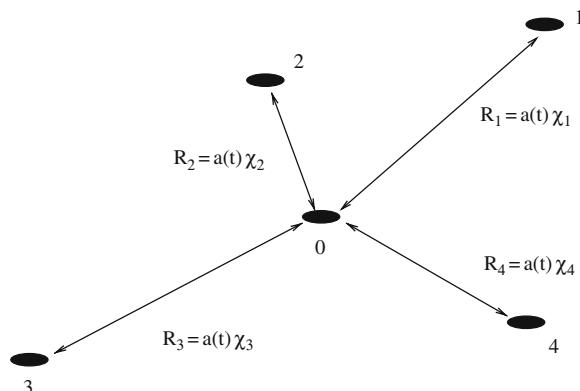


Fig. 1.7 If peculiar velocities are neglected, the distance R_i between us (galaxy 0) and galaxy i is given by $R_i = a(t) \chi_i$ where χ_i is time independent. The quantity χ_i is the “comoving radial coordinate” of the galaxy i

The energy densities in the universe as a function of time can be simply expressed in terms of the scale factor. The density associated with non-relativistic particles, $\rho_M(t)$, is proportional to the number density of particles from which it follows that

$$\rho_M(t) = \rho_M(t_0) \left(\frac{a_0}{a(t)} \right)^3 = \rho_M(t_0) \hat{a}^{-3}. \quad (1.55)$$

The time dependence of the densities of other types of energy will be found in the next sections.

1.2.2 Gravitation and the Friedmann Equation

In the absence of gravitation, the recession velocities of galaxies would be constant, implying $\ddot{a} = 0$. In the presence of the attractive effects of gravitation, we might expect that the expansion would be decelerated, $\ddot{a} < 0$. We will use general relativity in Chap. 4 to find the correct equation for \ddot{a} . As a preview, in this section we will use a Newtonian argument (of dubious validity) to find an equation that will turn out to be correct in a universe that is dominated by non-relativistic matter.

Referring to Fig. 1.8, we place a galaxy of mass m at a distance $R = \chi a(t)$ from the “center” of a universe of uniform density ρ . Since the mass distribution is spherically symmetric, Gauss’s theorem “suggests” that the galaxy is subject to a

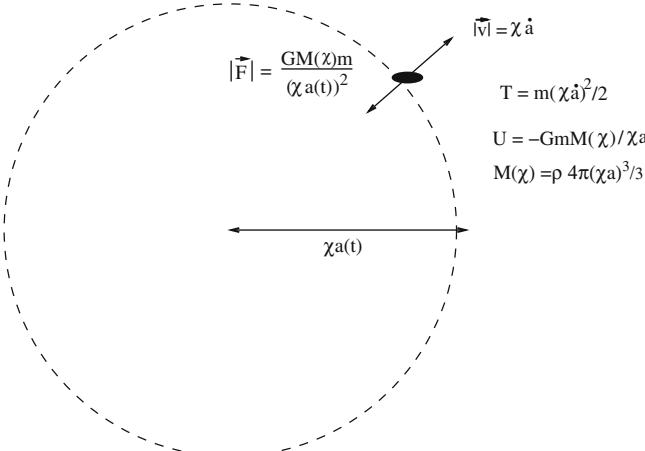


Fig. 1.8 A Newtonian treatment of the universal expansion. A galaxy of mass m placed in a universe of uniform density ρ at a distance $\chi a(t)$ from the “center of the universe.” The spherical symmetry suggests that the Newtonian force on the galaxy will be directed toward the origin with a magnitude $F = GM(\chi)m/(\chi a(t))^2$, where $M(\chi)$ is the total mass at a distance $< \chi a(t)$ from the origin. For a uniform mass density ρ , $M(\chi) = 4\pi(\chi a)^3 \rho/3$

gravitational force directed toward the origin that is proportional to the total mass at a distance $< \chi a(t)$ from the origin:

$$|\mathbf{F}| = \frac{GM(\chi)m}{\chi^2 a^2}, \quad (1.56)$$

where

$$M(\chi) = 4\pi(\chi a)^3 \rho / 3. \quad (1.57)$$

(We ignore the question of whether Gauss's theorem applies in an infinite medium.) Using $|\mathbf{F}| = m \ddot{\mathbf{R}} = m\chi \ddot{a}$ we find an equation for the deceleration of the universe:

$$\frac{\ddot{a}}{a} = \frac{-4\pi G\rho}{3} \quad \text{if } \rho \sim \rho_M. \quad (1.58)$$

This is the correct equation if ρ is dominated by non-relativistic matter. (In the relativistic generalization we will find in Chap. 4, the factor of ρ in (1.58) will be replaced by $\rho + 3p$ where p is the pressure.)

Multiplying and dividing by H_0^2 the right side of (1.58) we find the present value of the deceleration:

$$\left[\frac{\ddot{a}}{a} \right]_{t_0} = -H_0^2 \frac{\Omega_M}{2} \quad \text{if } \rho \sim \rho_M. \quad (1.59)$$

(In the relativistic generalization, the factor of $\Omega_M/2$ will be replaced by $(\Omega_R + \Omega_M/2 - \Omega_\Lambda)$). The characteristic time for a significant change in a recession velocity is $(\ddot{a}/a)^{-1}$ which today would have the value of $2t_H/\Omega_M$ if the universe were dominated by non-relativistic matter.

To find the solution $\dot{a}(t)$ of (1.58) we use the fact that, for a universe dominated by non-relativistic matter, ρa^3 is independent of time. In this case it is simple to verify the following solution:

$$\dot{a}^2 = \left[\frac{8\pi G\rho a^3}{3} \right] a^{-1} + \text{const},$$

where the bracketed quantity is time independent. To evaluate the constant we use the present values $\dot{a}(t_0) = H_0 a_0$ and $8\pi G\rho(t_0)/3 = H_0^2 \Omega_T$ to find

$$\dot{a}^2 = \frac{8\pi G\rho a^2}{3} + H_0^2 a_0^2 (1 - \Omega_T). \quad (1.60)$$

Dividing (1.60) by a^2 we find the “Friedmann equation”:

$$\left(\frac{\dot{a}}{a} \right)^2 - \frac{8\pi G\rho}{3} = H_0^2 (1 - \Omega_T) \hat{a}^{-2}. \quad (1.61)$$

In spite of the fact that (1.58) applies only to a universe dominated by non-relativistic matter, we will find in Chap. 4 that the Friedmann equation (1.61) is completely general for a homogeneous universe.

Equation (1.60) can be interpreted as the conservation of energy for the galaxy in Fig. 1.8:

$$\frac{T + U}{m\chi^2/2} = \dot{a}^2 - \frac{8\pi G\rho a^2}{3} = a_0^2 H_0^2 (1 - \Omega_T), \quad (1.62)$$

where the galaxy's kinetic energy T is

$$T = \frac{1}{2}m\dot{R}^2 = \frac{1}{2}m\chi^2\dot{a}^2, \quad (1.63)$$

and where the galaxy's gravitational potential energy U is

$$U(\chi) = -\frac{GmM(\chi)}{\chi a(t)} = -\frac{8\pi G\rho a^2}{3} m\chi^2/2. \quad (1.64)$$

We see that $\Omega_T > 1$ ($\Omega_T < 1$) corresponds to a negative (positive) total energy per galaxy.

In the next section, we will find the solution $a(t)$ of the Friedmann equation and, to do this, we will need the explicit form of $\rho(t)$. For non-relativistic matter, we have already noted

$$\rho_M(a) = \rho_M(a_0)\hat{a}^{-3} = \Omega_M \frac{3H_0^2}{8\pi G} \hat{a}^{-3}. \quad (1.65)$$

In this case the Friedmann equation becomes

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0^2 (\Omega_M \hat{a}^{-3} + (1 - \Omega_T) \hat{a}^{-2}) \quad \text{if } \Omega_T \sim \Omega_M. \quad (1.66)$$

With this form, we see explicitly that the Friedmann equation gives the expansion rate ($H(t) = \dot{a}/a$) as a function of the present expansion rate H_0 . The proportionality factor depends on $\Omega_M = \Omega_T$ and on the reduced expansion factor \hat{a} .

1.2.3 Open and Closed Universes

The solution $a(t)$ of the Friedmann equation for a homogeneous matter-dominated universe (1.66) depends on the value of $\Omega_T \sim \Omega_M$. Figure 1.9 shows three examples with $\Omega_T = \Omega_M$ and also the Λ CDM model with ($\Omega_M = 0.27$, $\Omega_\Lambda = 0.73$). Two general cases for $\Omega_T = \Omega_M$ are distinguished by their behaviors at large time:

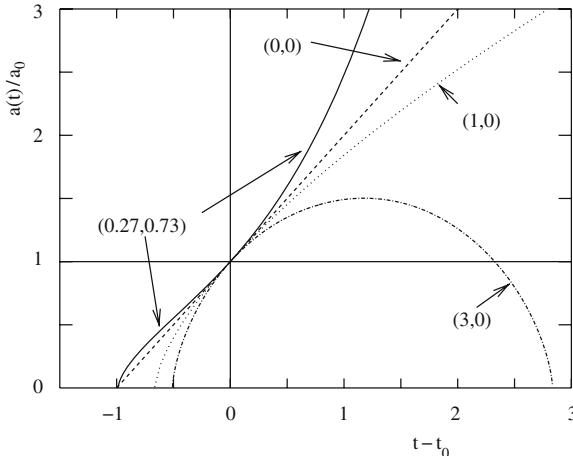


Fig. 1.9 The scale factor $a(t)$ for four combinations of $(\Omega_M, \Omega_\Lambda)$. The four curves have the same values of a_0 and H_0 . The elapsed time between the initial singularity and today (t_0) is t_H ($\Omega_M = \Omega_\Lambda = 0$), $(2/3)t_H$ ($\Omega_M = 1, \Omega_\Lambda = 0$), $0.5t_H$ ($\Omega_M = 3, \Omega_\Lambda = 0$), and $0.99t_H$ ($\Omega_M = 0.27, \Omega_\Lambda = 0.73$). This last combination corresponds to the standard Λ CDM parameters given in [1]

1. $\Omega_T = \Omega_M \leq 1$. In this case, (1.66) indicates that \dot{a} is positive for all values of \dot{a} . It follows that the expansion will be eternal. This is perhaps not surprising in view of the fact that the Newtonian energy of a galaxy (1.62) is positive for $\Omega_T < 1$. A homogeneous universe with $\Omega_T < 1$ is called an “open” universe. Simple expressions for $a(t)$ can be found in two cases. For an “empty universe,” $\Omega_T = \Omega_M = 0$, the first term on the right of (1.66) is zero and we find

$$a(t) = a_0 \frac{t}{H_0^{-1}} \quad (\Omega_T = \Omega_M = 0). \quad (1.67)$$

The age of the universe (the elapsed time between $a = 0$ and $a = a_0$) is the Hubble time. For a “critical” universe, $\Omega_T = 1$, the second term on the right of (1.66) is zero and we find

$$a(t) = a_0 \left(\frac{t}{(2/3)H_0^{-1}} \right)^{2/3} \quad (\Omega_T \sim \Omega_M = 1). \quad (1.68)$$

The age of the universe is $(2/3)$ of the Hubble time. The age is smaller than in the case of $\Omega_T = 0$ because gravitation has *decelerated* the expansion. It follows that expansion was faster in the past than it is today and less time is required for the universe to reach its present “size.”

2. $\Omega_T = \Omega_M > 1$. The case $\Omega_T > 1$ is called a “closed” universe. For $\Omega_T = \Omega_M > 1$ the expansion stops ($\dot{a} = 0$) when $a = a_{\max}$:

$$a_{\max} = a_0 \frac{\Omega_M}{\Omega_T - 1} \quad (\Omega_T = \Omega_M > 1). \quad (1.69)$$

Since (1.58) implies $\ddot{a} < 0$, it follows that the present expansion will be followed by a period of contraction ending (perhaps) with a big crunch. We will show in Chap. 3 that a closed homogeneous universe has a finite spatial volume. The connection between geometry (finite volume) and destiny (big crunch) is maintained only if the vacuum energy vanishes. In the case of a non-zero vacuum energy, a closed universe ($\Omega_T > 1 \Rightarrow$ finite volume) can expand eternally.

1.2.4 The Evolution of the Temperature

With the expansion of the universe, the energy density of non-relativistic matter falls as $\rho_M \propto a^{-3}$. This behavior is simply due to the dilution of the particles, $n \propto a^{-3}$. In this section we will show that the energy density of relativistic matter (photons and massless neutrinos) falls as $\rho_R \propto a^{-4}$. The difference between the two behaviors comes from the fact that a comoving observer sees a CMB energy distribution that is redshifted by a factor $\propto a^{-1}$ so that the mean photon energy⁶ falls as a^{-1} . This effect combines with the dilution, $n_R \propto a^{-3}$, to give $\rho_R \propto a^{-4}$.

The redshifting of the CMB photons is no different from the redshifting of stellar spectral lines that leads to the Hubble law. Consider two galaxies separated by a distance dR . If the universe is homogeneous, observers in the two galaxies see the same distribution of CMB photons. In particular, the two observers will see the same ratio between the mean CMB energy and the energy of a given atomic transition produced by local stars. Some of the stellar photons and some of the CMB photons in one galaxy are directed toward the other galaxy. When the stellar photons from one galaxy reach the other, they are observed to be redshifted by a factor $(1 - H_0 dR/c)$. Since there is no fundamental difference between stellar photons and CMB photons, the CMB photons must be redshifted by the same factor $dE/E = -H_0 dR/c$. Since all photons are redshifted by the same factor, the Planckian form of the spectrum is maintained and the temperature decreases.

The time for the exchange of photons between the two galaxies is $dt = dR/c$ so each observers sees a CMB temperature that decreases with time according to:

$$\frac{dT}{T} = -H_0 dt \quad \Rightarrow \quad \frac{dT}{da} = -\frac{T}{a}, \quad (1.70)$$

where in the second form we use $H_0 = \dot{a}/a$ at t_0 . The solution of this equation is

⁶ We will see that it is actually the momentum that is redshifted and that a background of non-relativistic particles suffers the same momentum redshift. However, for non-relativistic particles, the kinetic energy is unimportant so we can ignore the effect in the calculation of the energy density.

$$T(t) = T(t_0) \frac{a_0}{a(t)}. \quad (1.71)$$

It follows that in the past the universe was hotter than it is today⁷ and that in the future the universe will be colder. The fact that the momentum distribution is uniformly redshifted means that a Planck distribution remains a Planck distribution.⁸ This remarkable fact explains why we see a thermal distribution today even in the absence of thermalizing collisions. As we will see in Chap. 6, the photons acquired their Planck distribution in the early universe when the density, and therefore the collision rate, was high.

Since the photon energy density is proportional to T^4 , we have

$$\rho\gamma(a) = \rho\gamma(a_0) \hat{a}^{-4}. \quad (1.72)$$

The total density of relativistic matter must include relativistic neutrinos (and other relativistic species at early times). Including only neutrinos for the not-too distant past and using (1.34) we find

$$\rho_R(a) = \rho\gamma(a_0) \hat{a}^{-4} [1 + 0.227 N_{VR}(a)], \quad (1.73)$$

where $N_{VR}(a)$ is the number of species of relativistic neutrinos at the epoch $a(t)$. The present density of relativistic matter normalized to the present critical density is then

$$\Omega_R = \Omega\gamma [1 + 0.227 N_{VR}(a_0)]. \quad (1.74)$$

Since the density of relativistic matter (1.65) is proportional to a^{-3} while that of non-relativistic matter (1.73) is proportional to a^{-4} , relativistic matter must come to dominate for $a \rightarrow 0$. The moment when relativistic and non-relativistic densities were equal is called “ t_{eq} .” Assuming the three neutrino species were relativistic at this time, we find

$$\hat{a}_{eq} = \hat{a}(t_{eq}) = \frac{1.68\Omega\gamma}{\Omega_M} \sim 3.13 \times 10^{-4} \frac{0.27}{\Omega_M h_{70}^2}. \quad (1.75)$$

This gives a temperature at t_{eq} of

$$T_{eq} = 2.8 \Omega_M h_{70}^2 \text{ eV} = 0.75 \text{ eV} \frac{\Omega_M h_{70}^2}{0.27}. \quad (1.76)$$

⁷ This has been verified by measuring, in high-redshift galaxies, the spin temperatures of systems that are excited by CMB photons [20].

⁸ The fact that the normalization of the distribution remains Planckian, i.e. $n\gamma \propto T^3$ and $\rho\gamma \propto T^4$ follows from the conservation of the number of photons, $n\gamma \propto a^{-3}$. Since $T \propto a^{-1}$ it follows that $n\gamma \propto T^3$ preserving the Planckian normalization. This delicate point will be made clearer in the following discussion of Liouville’s equation.

Equations (1.75) and (1.76) apply as long as all neutrinos have masses $\ll T_{\text{eq}}$. We will generally assume this to be true.

As a preview of Chap. 3, we note that the derivation of the temperature redshift (1.71) also applies to non-CMB photons. Consider a photon of energy E_1 emitted at t_1 by a distant galaxy and detected by us at t_0 . By the same argument, the photon will be observed by us to have an energy

$$E_0 = E(t_1) \frac{a_1}{a_0}. \quad (1.77)$$

Since photon wavelengths are inversely proportional to photon energies, this gives us a formula for the redshift:

$$z + 1 \equiv \frac{\lambda_0}{\lambda_1} = \frac{a_0}{a_1}. \quad (1.78)$$

This elegant expression says that the wavelength of a photon scales by the same factor as the universe itself.⁹ We will re-derive this formula in Chap. 3 in the usual way by using comoving coordinates.

We end this section on a note meant only for readers who like equations. The law (1.71) can be derived more formally by using the “Boltzmann equation” that governs the phase-space distribution of particles

$$F = \frac{dN}{d^3r d^3p}. \quad (1.79)$$

Here we will consider the Boltzmann equation for photons in the absence of collisions between particles, i.e. Liouville’s equation. This approximation is justified today since the universe is sufficiently dilute to make collisions between photons and matter very rare (Exercise 2.3). In Chap. 6 we will derive the Boltzmann equation with collisions, allowing us to treat the dense early universe.

Today ($t = t_0$) we observe locally ($\mathbf{R} = 0$) a thermal distribution of photons with $T(t_0) = T_0 \sim 2.7$ K. This corresponds to a Planckian phase-space distribution $F(\mathbf{p}, \mathbf{R}, t)$

$$F(\mathbf{p}, \mathbf{R} = 0, t_0) = \frac{1}{(2\pi)^3} \frac{1}{\exp(E_p/kT_0) - 1} \left(\frac{1}{\hbar c}\right)^3. \quad (1.80)$$

We want to calculate the phase-space distribution at ($\mathbf{R} = 0$) in the future. As we have seen, the photons that will be here tomorrow are over there today which means

⁹ This sometimes leads people to say that the wavelength of a photon grows because space is expanding. While the author does not claim to understand what is meant by this statement, he does not deny that it is conceivable that a meaning can be found. It should, however, be emphasized that space is apparently not expanding uniformly because the sizes of bound objects like atoms and galaxies do not grow with time.

$$F(\mathbf{p}, \mathbf{R} = 0, t_0 + dt) \propto F(\mathbf{p}, \mathbf{R} = -\mathbf{v}_p dt, t_0). \quad (1.81)$$

The momentum \mathbf{p} on both sides of this equation refer to the momentum *in our Galilean reference frame*. In particular, the \mathbf{p} on the right-hand side is not the \mathbf{p} that would be measured by an observer at rest in a receding galaxy at the position $\mathbf{R} = -\mathbf{v}_p dt$. The use of such Galilean frames will be justified in Chap. 4 where we will see that space-time can be considered to be flat on scales much less than d_H .

From Fig. 1.10, we see that the constant of proportionality in (1.81) is unity (Liouville's theorem). It follows that

$$\frac{\partial F}{\partial t} = -\mathbf{v}_p \cdot \nabla_{\mathbf{R}} F. \quad (1.82)$$

This is Liouville's equation in the absence of forces. Gravitational forces can be neglected because the gradient of the gravitational potential (1.64) vanishes at the origin. Once again, a relativistic justification of this must await Chap. 4.

Photon momenta at $d\mathbf{R}$ in our Galilean frame are related to the momenta seen by an observer in the receding galaxy at $d\mathbf{R}$ by a Lorentz transformation with $\beta = v/c = H_0 dR/c \ll 1$, i.e. $\mathbf{p} \rightarrow \mathbf{p} - E_p \boldsymbol{\beta}$. It follows that

$$F(\mathbf{p}, d\mathbf{R}, t_0) \propto G(\mathbf{p} - H_0 E_p d\mathbf{R}, 0, t_0), \quad (1.83)$$

where $G(\mathbf{p})$ is the phase-space distribution seen by an observer in the receding galaxy at $d\mathbf{R}$ (corresponding to the origin in his frame). Once again, the constant of proportionality is unity (Fig. 1.11) because the Lorentz transformation preserves the phase-space density.

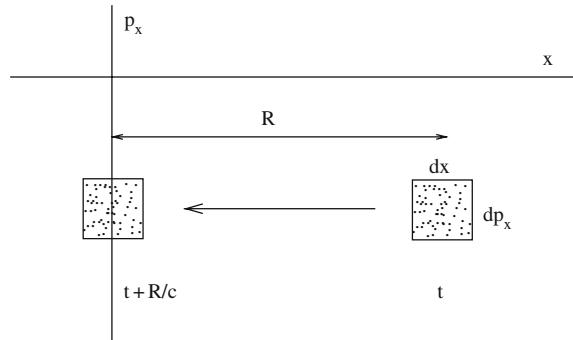


Fig. 1.10 Photons in phase space. The box on the right contains photons with $p_x < 0$ and $p_y = p_z = 0$ that at time t are at the position ($x \sim R$, $y = z = 0$). By following individual trajectories, we can see that at time $t + R/c$, the photons have the same momenta but are in a box at $x \sim 0$. The size and shape of the box does not change because all photons have the same velocity. The fact that the size does not change implies that the phase space density does not change, i.e. that the constant of proportionality in (1.81) is unity

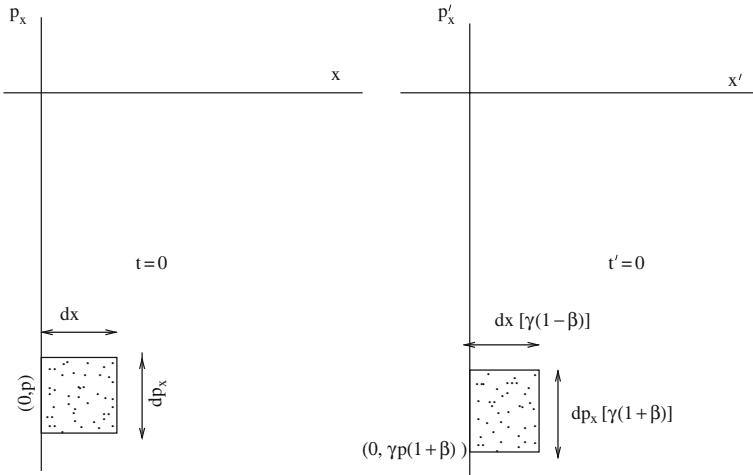


Fig. 1.11 Photons with $p_y = p_z = 0$ and $y = z = 0$ viewed in two frames related by a Lorentz transformation along the x direction. The fact that the area of the box is the same in both frames implies that the constant of proportionality in (1.84) is unity

If the universe is homogeneous, the function G is the same function as that observed in our Galilean frame at the origin. Using this fact, (1.83) becomes

$$F(\mathbf{p}, d\mathbf{R}, t_0) = F(\mathbf{p} - H_0 E_p d\mathbf{R}, 0, t_0), \quad (1.84)$$

which is equivalent to

$$\nabla_R F = -H_0 E_p \nabla_p F. \quad (1.85)$$

Substituting this into (1.82) we find

$$\frac{\partial F}{\partial t} = H_0 \mathbf{p} \cdot \nabla_p F. \quad (1.86)$$

Changing variables $t \rightarrow a(t)$ and using $H_0 = \dot{a}/a(t_0)$, we find

$$a \frac{\partial F}{\partial a} = \mathbf{p} \cdot \nabla_p F. \quad (1.87)$$

The general solution is

$$F(\mathbf{p}, a) = F(p a/a_0, a_0), \quad (1.88)$$

i.e. the photon momenta are redshifted by a factor a_0/a . A particular solution is

$$F(\mathbf{R} = 0, \mathbf{p}, a) = \frac{1}{(2\pi)^3} \frac{1}{\exp(E_p/T(a)) - 1}, \quad (1.89)$$

with

$$T(a) = T_0 \frac{a_0}{a}. \quad (1.90)$$

The evolution of $n\gamma$ can be found by integrating over $d^3 p$ the equation (1.87) or its solution (1.88)

$$\frac{dn\gamma(r=0, t)}{dt} = -3 \frac{\dot{a}}{a} n\gamma \quad \Rightarrow \quad n\gamma = n\gamma(a_0) \left(\frac{a_0}{a} \right)^3. \quad (1.91)$$

1.2.5 An Improved Friedmann Equation

We must modify the Friedmann Equation (1.66) to take into account energy that is not in the form of non-relativistic matter. Two types of energy come to mind. The first is relativistic matter with an energy density given by (1.73). Another possible source would be a vacuum energy. This energy is not diluted during the expansion so we expect

$$\rho_\Lambda(a) = \rho_\Lambda(a_0) = \frac{3H_0^2}{8\pi G} \Omega_\Lambda. \quad (1.92)$$

In Chap. 4 we will see that we need only to add these other sources of energy to the non-relativistic Friedmann Equation (1.66). We obtain

$$\left(\frac{\dot{a}}{a} \right)^2 = H_0^2 (\Omega_M \hat{a}^{-3} + \Omega_R \hat{a}^{-4} + \Omega_\Lambda + (1 - \Omega_T) \hat{a}^{-2}), \quad (1.93)$$

where Ω_T is the sum of the Ω s:

$$\Omega_T = \Omega_M + \Omega_R + \Omega_\Lambda. \quad (1.94)$$

The effect of the four terms in (1.93) is clearer if we multiply by a^2 :

$$a^2 = H_0^2 a_0^2 (\Omega_M \hat{a}^{-1} + \Omega_R \hat{a}^{-2} + \Omega_\Lambda \hat{a}^2 + (1 - \Omega_T)). \quad (1.95)$$

We see that the effect of relativistic and non-relativistic matter is to decelerate the universe: $\dot{a} \rightarrow \infty$ for $\hat{a} \rightarrow 0$. This is as we would expect for normal gravitation. On the other hand, a positive vacuum energy has the effect of accelerating the expansion: $\dot{a} \rightarrow \infty$ for $\hat{a} \rightarrow \infty$ if $\Omega_\Lambda > 0$. This conclusion is confirmed by differentiating (1.95):

$$\ddot{a} = -H_0^2 a_0 \left(\frac{\Omega_M}{2} \hat{a}^{-2} + \Omega_R \hat{a}^{-3} - \Omega_\Lambda \hat{a} \right). \quad (1.96)$$

It is useful to define the “deceleration parameter” q_0 :

$$q_0 \equiv - \left[\frac{a\ddot{a}}{\dot{a}^2} \right]_{t_0} = \frac{\Omega_M}{2} - \Omega_\Lambda, \quad (1.97)$$

where we have neglected $\Omega_R \ll \Omega_M$. With the first and second derivatives (1.95) and (1.96) we can give a Taylor expansion for $a(t)$ that is useful for $(t - t_0) \ll t_H = H_0^{-1}$.

$$a(t) = a_0 \left[1 + \frac{t - t_0}{t_H} - \frac{q_0}{2} \left(\frac{t - t_0}{t_H} \right)^2 + \dots \right] \quad (1.98)$$

This expression will be used in Chap. 5 for the analysis of the fluxes from distant galaxies.

With the \hat{a} dependence of the four terms in (1.93), we see that the universe passes through a succession of epochs where one of the four terms dominates, as summarized in Table 1.2.

Table 1.2 The epochs of an ever-expanding universe with a cosmological constant. For a_{eq} , we suppose three neutrino species with $m_V < 1$ eV

Epoch	$a(t)$
Radiation: $\hat{a} < \hat{a}_{eq} = 1.68\Omega_\gamma/\Omega_M$ $= 3.13 \times 10^{-4} \times 0.27 / (\Omega_M h_{70}^2)$	$a(t) \propto t^{1/2}$
Matter: $a_{eq} < a < (\Omega_M/\Omega_\Lambda)^{1/3}$	$a(t) \propto t^{2/3}$
Vacuum: $a > (\Omega_M/\Omega_\Lambda)^{1/3}$ (if $\Omega_\Lambda > 0$)	$a(t) \propto \exp(H_0\Omega_\Lambda^{1/2}t)$

In all cases, the universe starts in a “radiation” epoch that ends at t_{eq} to be followed by the matter epoch. If there were no vacuum energy, what follows would then only depend on Ω_T . For $\Omega_T > 1$ (and $\Omega_\Lambda \sim 0$), the expansion reaches a maximum at

$$a_{max} = a_0 \frac{\Omega_M}{\Omega_T - 1} \quad \Omega_T > 1, \quad \Omega_\Lambda \sim 0. \quad (1.99)$$

In this case the expansion is followed by a contraction and the matter epoch ends with a new radiation epoch. On the other hand, for $\Omega_T < 1$ (and $\Omega_\Lambda \sim 0$), the matter epoch is followed by the “curvature” epoch when the Friedmann equation is dominated by the $(1 - \Omega_T)$

$$\hat{a} > \frac{\Omega_T}{1 - \Omega_T} \quad a(t) \propto t \quad \text{curvature epoch.}$$

The expansion is “free”, $a(t) \propto t$ because the density of matter is not sufficient to decelerate the universe.

For $\Omega_A > 0$, the matter epoch can be followed by the vacuum epoch that is characterized by an exponential expansion

$$\hat{a} > \left(\frac{\Omega_M}{\Omega_A} \right)^{1/3} \quad a(t) \propto \exp \left(H_0 \Omega_A^{1/2} t \right) \quad \text{vacuum epoch.}$$

The expansion has different time dependences in the curvature and vacuum epochs. Nevertheless, from a practical point of view they are similar because, as we will see in the next section, the formation of structure is suppressed after matter domination ends. For $\Omega_M \sim 0.27$ and $\Omega_A \sim 0.73$, structure formation has more or less ended since $\hat{a} = (0.27/0.73)^{1/3} \sim 0.75$.

1.2.6 The Evolution of the Ω s and Structure Formation

The Ω s are defined by the current values of the densities and expansion rate:

$$\Omega_I \equiv \frac{\rho_I(t_0)}{3H_0^2/8\pi G} \quad I = T, M, R, \Lambda, \dots . \quad (1.100)$$

This definition can be generalized to give the Ω s as a function of time:

$$\Omega_I(t) \equiv \frac{\rho_I(t)}{3(\dot{a}/a)^2/8\pi G} \quad I = T, M, R, \Lambda, \dots . \quad (1.101)$$

The Friedmann equation (1.93) gives the expansion rate in the denominator. The time dependences of the energy components are given by (1.65), (1.73), and (1.92). During the matter epoch, we find

$$\Omega_T(a) \sim \Omega_M(a) \sim \frac{1}{1 + \hat{a}(1 - \Omega_T)/\Omega_M}. \quad (1.102)$$

We see that $\Omega_T(a) \rightarrow 1$ for $a \rightarrow 0$. We could have guessed this directly from the Friedmann equation (1.93) since the curvature term proportional to $(1 - \Omega_T)\hat{a}^{-2}$ becomes negligible compared to $\Omega_M\hat{a}^{-3}$ for $a \rightarrow 0$.

The precision with which Ω_T was near unity in the past is impressive. Since Ω_T is now within 10^{-2} of unity, at the beginning of the matter epoch, $\hat{a}_{\text{eq}} \sim 10^{-4}$, $\Omega_T(a)$ was within 10^{-6} of unity.

During the radiation epoch, Ω_T continues to approach unity for $a \rightarrow 0$:

$$\Omega_T(a) \sim 1 - \frac{1 - \Omega_T(a_0)}{\Omega_R(a_0)} \hat{a}^2 \quad a < a_{\text{eq}}. \quad (1.103)$$

At the epoch of nucleosynthesis, $\hat{a} \sim 10^{-9}$, $\Omega_T(a)$ was within 10^{-14} of unity.

What is remarkable here is that $\Omega_T(a)$ was sufficiently close to unity for it to remain close to unity for a period of $\sim t_H \sim 10^{10}$ yr. It is easy to show that if, at the epoch t_1 , we have $|\Omega_T(a_1) - 1| = \epsilon \ll 1$, Ω_T will start to differ significantly from unity after the universe expands by a factor of $\epsilon^{-1/n}$ with $n = 2$ ($n = 1$) for the radiation (matter) epochs. If $\Omega_T(a)$ had differed significantly from unity in the distant past, the universe would have either quickly reached a maximum expansion, and then contracted, or quickly become curvature or vacuum-dominated. The fact that Ω_T was sufficiently close to unity is natural in inflationary models, as discussed in Sect. 5.6.

The long duration of the matter epoch with $\Omega_M(a) \sim \Omega_T(a) \sim 1$ is important because a simple Newtonian argument shows that gravitationally bound structures can be formed from small density fluctuations only when $\Omega_M(a) \sim \Omega_T(a) \sim 1$. The argument is illustrated in Fig. 1.12 which shows a universe whose mean density is critical. Because $\Omega_T = 1$, the mean Newtonian energy (1.62) of a particle in this critical universe is zero, permitting an eternal expansion. The universe contains a region with a small over-density $\rho + \delta\rho$, $\delta\rho \ll \rho$. A particle in the over-dense region will have its potential energy lowered so that its total energy can be negative $T(r) + U(r) < 0$. Particles in the over-dense region will thus participate in the

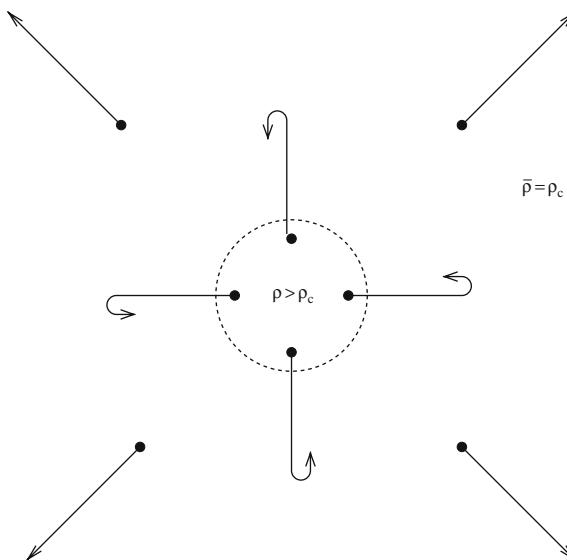


Fig. 1.12 Structure formation in a universe with $\Omega_M = \Omega_T \sim 1$. In such a universe, an over-dense region ($\rho > \rho_c$) can evolve like a mini-closed universe. The region will expand to a maximum radius and then detach from the universal expansion by contracting to form a gravitationally bound object while the rest of the universe continues its expansion. In a universe with Ω_M significantly less than unity, a small over-density will not be sufficient to make the region supercritical and the region will expand eternally. In a universe with Ω_M significantly greater than unity, slightly over-dense regions do not have time to form structures before the universe as a whole stops expanding and starts to contract

expansion for a finite time before reaching a maximal expansion and then falling back on themselves to form a gravitationally bound structure like a galaxy or cluster of galaxies. This process will be studied in Chap. 7.

In a sub-critical universe, a small over-density will not be sufficient to make a particle's energy negative. An over-dense region will continue expanding eternally, though at a slower rate.

On the other hand, in a significantly overcritical universe, a small over-density will not have time to form a structure since the universe as a whole starts to contract after a few Hubble times.

For the future, $a \rightarrow \infty$, it is easy to show that $\Omega_T(a) \rightarrow 0$ if $\Omega_T < 0$ and $\Omega_\Lambda = 0$. On the other hand in the presence of vacuum energy $\Omega_\Lambda > 0$ the universe becomes critical at large times $\Omega_T(a) = \Omega_\Lambda(a) \rightarrow 1$. In either case, structure formation ceases at the end of matter-domination. Only if $\Omega_M = \Omega_T = 1$ will structures of ever-increasing size continue to be formed. For example, if $\Omega_M = \Omega_T = 1$, our local group of galaxies will eventually form a bound structure with the neighboring Virgo cluster.

1.2.7 The Standard Scenario

We have seen that in the standard Λ CDM cosmological model, the universe passes through a succession of epochs when the universe is dominated by radiation, by non-relativistic matter, and then by vacuum energy. The energy densities as a function of temperature are shown in Fig. 1.13.

In Table 1.3, we list some formative events in the history of the universe according to this scenario. Non-controversial physics allows us to follow with confidence the succession of events starting at, say, $T \sim 1$ GeV when the universe was a nearly homogeneous soup of quarks, gluons, and leptons. With time, the universe cooled and a succession of bound states were formed, hadrons, nuclei, atoms, and finally the gravitationally bound stars and galaxies. The moments of the formation of bound states are called “recombinations.” The recombination that resulted in the formation of atoms caused the universe to become effectively transparent to photons. Surprisingly, the atoms lasted a relatively short time since they were “reionized” by photons produced by the first stars and quasars. By this time, the universe was sufficiently dilute that it remained transparent (Exercise 2.3).

We should add that the nature of the radiation changed with temperature. Today, the radiation consists of photons and any neutrinos with masses less than $\sim 10^{-4}$ eV. At temperatures $T > m_e$, electron–positron pairs could be produced and we will show in Chap. 6 that these pairs were in thermal equilibrium with the photons and formed a blackbody spectrum similar to those of the photons and neutrinos. Going back in time, each time the temperature rose above a particle–antiparticle threshold, a new blackbody component was created. During this period, the numbers of electrons and positrons and the numbers of baryons and antibaryons were nearly equal. The small number of electrons and baryons present today resulted from the small excess ($\sim 10^{-10}$) of particles over antiparticles present when $T \gg m_e$.

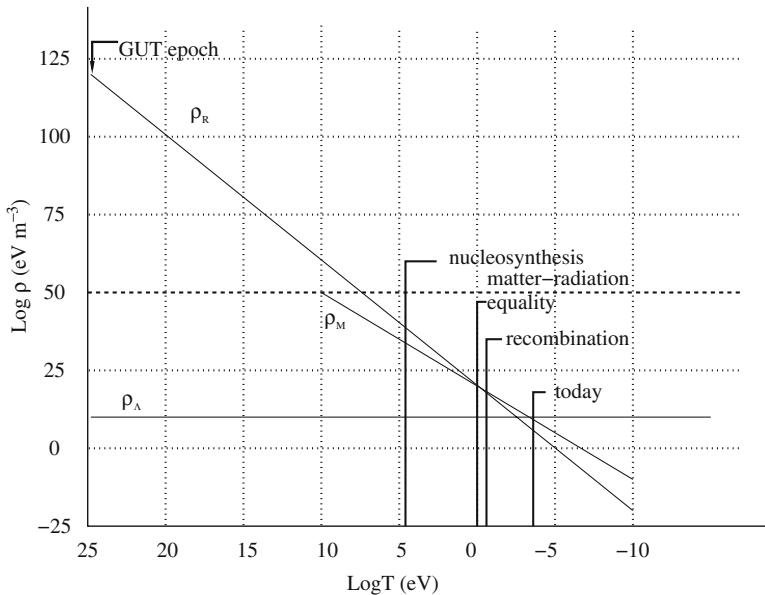


Fig. 1.13 The energy density of matter, radiation and vacuum as a function of temperature. The temperature scale starts at the expected “grand unification” scale of $\sim 10^{16}$ GeV. We suppose that the CDM particles have masses greater than ~ 10 GeV so the line of ρ_M starts at 10 GeV

We do not know the particles present for temperatures > 100 GeV because accelerators have not yet produced particles of such high mass. A variety of speculative theories predict such particles. For example, supersymmetric models often have

Table 1.3 Some formative events in the past. The values of t_0 , t_{rec} , and t_{eq} depend on $(h_{70}, \Omega_M, \Omega_A)$ and we have used (1, 0.27, 0.73)

t	T_γ (eV)	Event
$t_0 = 1.37 \times 10^{10}$ yr	2.349×10^{-4}	Today
$\sim 5 \times 10^8$ yr	$\sim 2.3 \times 10^{-3}$	Formation of the first structures, atomic reionization
$t_{\text{rec}} \sim 3.8 \times 10^5$ yr	0.25	“Recombination” (formation of atoms), universe becomes transparent
$t_{\text{eq}} \sim 5.7 \times 10^4$ yr	0.75	Matter-radiation equality
3 min	6×10^4	Nucleosynthesis (formation of light nuclei, $A = 2, 3, 4, 6, 7$)
1 s	10^6	$e^+ e^- \rightarrow \gamma\gamma$
4×10^{-6} s	$\sim 4 \times 10^8$	QCD phase transition (formation of hadrons from quarks and gluons)
$< 4 \times 10^{-6}$ s	$> 10^9$	Baryogenesis (?) (generation of baryon-antibaryon asymmetry) Inflation (?) (generation of density fluctuations)

particles in the range 100GeV–1000GeV and the lightest of these particles could be the cold dark matter.

Above supersymmetric energies is the “Grand Unification Theory” (GUT) scale estimated to be of order $E_{\text{GUT}} \sim 10^{15}$ to 10^{16} GeV. At this energy, the effective coupling constants of the weak, electromagnetic, and strong interactions appear to converge to a common value. Super-heavy bosons with masses at this scale are expected to mediate baryon number non-conserving interactions causing proton decay. They may have had a role in the generation of the baryon–antibaryon asymmetry.

Cosmologically, GUT interactions are expected to produce stable “topological” objects like magnetic monopoles. The expected monopole relic density is far in excess of the observational upper limits on the present density of monopoles. For this reason, a period of inflation is required to occur at a temperature below the GUT scale in order to diminish the monopole density through an exponential expansion of the universe. The quantum fluctuations of a scalar field driving the inflation may also have seeded today’s density inhomogeneities.

Current thinking suggests that the basic theoretical ideas based on quantum field theory should continue to be valid up to the Planck energy, 10^{19} GeV, beyond which unknown quantum gravity effects should be important. Theoretical work on this epoch involved quantum gravity theories or superstring theories and are clearly at the limits of speculative cosmology.

1.3 Open Questions

The standard Λ CDM model is consistent with a very wide variety of cosmological measurements. The precision of its parameters in Table 1.1 reflects the fact the agreement with the data (e.g., Fig. 7.11) is very good indeed. This does not, of course, mean that cosmology has reached its final goal. The Λ CDM model works on distance scales that are sufficiently big that the universe is to first approximation homogeneous. An important open question is:

- Does Λ CDM correctly describe the formation and structure of galaxies?

It is, in fact, often stated that models with CDM do not correctly predict the observed characteristics of galaxies [21]. Before concluding, it should be stressed that cosmologists do not understand the details of how galaxies and stars detach from this nearly homogeneous mixture. This process involves complicated gravitational fluid dynamics that is only beginning to be simulated accurately with computer models. “Feedback” from the formation of stars is also important. For example, supernova explosions create “winds” that can eliminate gas from galaxies.

Apart from these astrophysical questions on galactic scales, we can identify a number of fundamental questions that are stimulating current cosmological research. The most ambitious and provocative is:

- Are there alternative theoretical frameworks that are consistent with the data but that do not require the existence of dark energy and/or dark matter?

The excellent agreement between data and the Λ CDM model suggests that this is unlikely. Nevertheless, persistent declarations by some that models with CDM are not consistent with the observed properties of galaxies continues to stimulate attempts to eliminate CDM by modifying gravitational laws [22]. Modifications of gravity have also been suggested as a way to avoid dark energy. Some modifications include to the introduction of hypothetical extra (invisible) spatial dimensions [23] that modify the Friedmann equation. Finally, we mention attempts to do away with the need for dark energy by appealing to “back-reaction” in the gravitational dynamics [24]. These models are inspired by the fact that the Friedmann equation is invariably used to interpret cosmological data in spite of the fact that this equation only applies to a homogeneous universe. The hope is that the back-reaction due to the inhomogeneities of the real universe will generate terms in an effective Friedmann equation that cause accelerated expansion.

Whatever the value of these suggestions for modifying gravity, an important question for cosmology is

- Is it possible to test the laws of gravitation at cosmological scales?

General relativity has been tested on the scale of the solar system but tests at cosmological scales are only beginning [25]. These tests involve determining if the mixture of dark energy and dark matter that controls the expansion of the universe also explains the rate at which matter inhomogeneities grow through gravitational attraction. It has been suggested [26] that this can be done by precisely measuring “redshift distortions” of the matter distribution, as discussed in Sect. 7.2.

Staying within the paradigm of general relativity with dark-energy and dark-matter, an important question is:

- Are there small deviations from the Λ CDM model that better describe the data?

One possibility would be a dark energy that is not completely time independent. The time derivative of ρ_Λ is parameterized by w (Eq. 1.46) and a deviation from Λ CDM would be characterized by $w \neq -1$. Another would be Ω_T slightly different from unity. The ability to place strict limits on such deviations from Λ CDM are often used to judge the quality of observational proposals.

A question that is fundamental for the credibility of the Λ CDM model is:

- What is the dark matter?

The question “what is the dark energy” might not have any empirical answer beyond “the uniform substance that makes the expansion accelerate.” On the other hand, the dark matter supposedly consists of real objects that should be found and characterized. Before this is done, doubts will remain that Λ CDM cosmology has anything to do with reality.

The Λ CDM model general supposes “initial” conditions were set by a period of inflationary exponential expansion. An important question is therefore

- Is there any “direct” evidence for inflation?

The most widely discussed possibility are gravitational waves produced during inflation (Sect. 7.3). Such waves would imprint certain polarization patterns (“B modes”) on the CMB (Sect. 7.5) that may be observable. It may also be possible to directly detect the relic gravitational waves [27].

Finally, a question that goes to the heart of cosmology:

- How did the mixture of substances in Table 1.1 arise?

Within the framework of conventional science, an explanation for the mixture would involve knowing the “initial” conditions of the universe at some early time and knowing the laws of physics that say how the universe subsequently evolved. This would allow one to calculate the current condition in terms of the initial condition.

The only number that is now understood in Table 1.1 is the ratio between the number of neutrinos and the number of photons. Non-controversial physics suggests that at early times the temperature and density were sufficiently high that the neutrinos and photons were in thermal equilibrium. In such a case, the initial condition is simple: black body spectra for photons and each neutrino species that differ only because photons are boson and neutrinos are fermions. Starting with these initial conditions, one can calculate how the photon–neutrino ratio evolves with time. It does not remain constant because the neutrino–photon thermal contact is lost at a time determined by the known neutrino scattering cross-section. Thus, knowing the initial condition (thermal equilibrium) and the laws of physics (neutrino cross-section) allows us to calculate the neutrino–photon ratio today, as we will do in Chap. 6.

It is hoped that a similar line of reasoning will explain the baryon–photon ratio. As discussed in more detail in Chap. 6, the existence of baryons today is due to a tiny excess of baryons over antibaryons when the temperature was much greater than the proton mass so that both baryons and antibaryons had blackbody spectra. Such an asymmetry could arise from a baryon–antibaryon symmetric initial condition at high temperature, $kT \gg \text{GeV}$, through interactions between particles that do not respect “CP” invariance. (CP is an operation that inverses spatial coordinates and changes particles to antiparticles.) While the general requirements for the generation of the observed baryon–antibaryon asymmetry is understood, the details are lacking because of no empirical guidance on the particles that were present for temperatures greater than 100 GeV nor on their interactions. The reason is that particle accelerators have not yet explored in detail this energy regime. The recently commissioned Large Hadron Collider at CERN will extend our knowledge into the TeV range and some guidance on the physics relevant for the baryon asymmetry may come.¹⁰

The understanding of the CDM-photon ratio is more difficult than understanding the baryon–photon ratio simply because we do not know what the CDM consists of.

¹⁰ It should be mentioned that even applying “known” physics to the early universe requires us to suppose that the masses and couplings of known particles are time-independent. While such hypothetical time variations are rather at the margins of cosmology, the study of their cosmological implications has a long and fascinating history [28].

One popular candidate is the lightest new particle in “supersymmetric” particle theories (see Appendix C). This particle would be expected to be a stable neutral weakly interacting particle (WIMP). The discovery of supersymmetric particles at the LHC would lend support to this CDM candidate. If such particles do make up the CDM, the WIMP–photon ratio can be calculated in a way that closely parallels the calculation of the neutrino–photon ratio. The critical parameters are the WIMP mass and annihilation cross-section, as we will see in Chap. 6.

Supersymmetric WIMPs are an example of CDM whose density relative to photons is determined, via the annihilation cross-section, by the parameters of the underlying particle theory (Sect. 6.6). An instructive alternative case is that of the axion (Sect. 2.5.2), a light scalar particle associated with a symmetry proposed to avoid large CP violation in the strong interactions. The axion–photon ratio depends not only on the parameters of the theory but on certain dynamical variables that can be time and position dependent. Under certain circumstances [29] this can lead to a present-day universe where the axion–photon ratio is constant over individual Hubble volumes (volumes of radius greater than the Hubble distance) but highly variable over super-Hubble scales. The theory thus does not predict a definite axion–photon ratio but only a spectrum of ratios with a well-defined probability distribution. A prediction for the likelihood of observing a particular ratio must take into account not only the underlying probability distribution but also the probability that a given axion–photon ratio leads to the evolution of observers. For example, large axion–photon ratios may be disfavored if they are correlated with small baryon–axion ratios (since observers are made of baryons, not axions).

Axion models with super-Hubble variations are examples of “multiverse” models where the observed properties of “our” universe are determined in part by “anthropic” selection effects. Such models have been encouraged by our current understanding of superstring theory which suggests that the production of a spectrum of Hubble volumes with different physical parameters is natural. In this context, the spectrum of possible characteristics of individual universes is called the superstring “landscape” [30].

The most popular use of multiverse theories involves the vacuum energy density [31], observed to be $\rho_A = 3.75 h_{70}^2 \times 10^9 \text{ eV m}^{-3}$. This value is not predicted by any accepted fundamental theory. The only clue to its value is that larger values would lead to less structure formation. For example, if it were a factor 10^4 larger, the radiation epoch would have been followed immediately by a vacuum epoch. Since galaxies and stars can only form in the matter epoch, such a large vacuum energy would result, presumably, in no observers. The small value of ρ_A in these models is the result, in part, of anthropic selection. If one assumes an underlying probability distribution for ρ_A that is a constant for small values of ρ_A , it turns out that most galaxies form in Hubble volumes with observed ρ_A not too far from our observed value.

A glance at Fig. 1.13 makes the approximate equality of ρ_M and ρ_A appear to be an extraordinary coincidence since at the GUT epoch about 120 magnitudes separated the radiation and vacuum densities whereas today the matter and vacuum

densities are only separated by a factor ~ 2 . One can ask if it is conceivable that physics operating at the GUT scale could have “fine tuned” ρ_A so that it becomes comparable to the matter density just when intelligent observers were evolving [32]. Anthropic selection is, for the moment, the most plausible explanation.

Multiverse theories are dynamical applications of the “weak anthropic principle” [33] which states that we can only observe physical laws that are consistent with our existence. While the use of anthropic reasoning is still controversial, it follows the historical development of science which has often replaced physical “laws” with arbitrary initial conditions. For instance, before the discovery of Newtonian dynamics, some believed that the radii of planetary orbits should be calculable from first principles. It is now known that they result from the complicated condensation of pre-planetary nebulae. The orbital radius of our own planet is determined by the anthropic requirement that life requires complex chemistry that arises only over a limited temperature range. Anthropic selection in cosmology is no different, apart from the unfortunate fact that, unlike other planets, the other Hubble volumes are not observable.

It is also important to keep in mind that even things that we normally consider as being calculable, e.g. the abundances of light elements, depend on the values of various “fundamental” constants whose origin is not known. It is possible that the fundamental constants depend on position in the multiverse and that a truly fundamental theory only gives probabilities that they take on any particular value [34]. This possibility is sometimes considered to be an “explanation” for why the values of the fundamental constants allow for the production of carbon in stars. For example, because of peculiarities in the level structures of ^8Be , ^{12}C , and ^{16}O , changing the value of the fine-structure constant by a few percent leads to very little production of carbon [35], an element obviously necessary for carbon-based biochemistry.

It must be emphasized that predictions of multiverse models combined with anthropic selection require understanding not only of physics but also of biology, or at least the understanding of how biology can arise. Generally, one assumes that the emergence of observers requires the existence of baryons that concentrate in bound structures (galaxies) where stable planetary systems orbiting stars can evolve. This places constraints not only on the mixture baryons, CDM, and dark energy but also on the spectrum of inhomogeneities within individual Hubble volumes (as discussed in Chap. 7).

Exercises

1.1 Plot the functions $\Omega_T(a)$, $\Omega_M(a)$, and $\Omega_A(a)$ for the cases of the current values ($\Omega_M = \Omega_T = 1$), ($\Omega_M = \Omega_T = 0.3$), and ($\Omega_M = 0.3$, $\Omega_A = 0.7$).

Structure formation can be said to cease when $\Omega_M(a)$ falls below, say, $\Omega_M = 0.5$. For the three cases, determine the value of $a(t)$ when this happens.

1.2 The Friedmann Equation (1.93) can be integrated to calculate $t(a)$.

- (a) Estimate the time that has passed since the universe became vacuum dominated at $\hat{a} = (\Omega_M/\Omega_\Lambda)^{1/3}$. The integral cannot be done analytically so you can either do it numerically or evaluate it approximately by including only the dominant term of the Friedmann equation. (This approximation means you can make at most an error of $\sqrt{2}$ which is sufficient for this exercise and those that follow.)
- (b) Calculate the time the universe spent in the matter dominated epoch $a_{\text{eq}} < a < a_0(\Omega_M/\Omega_\Lambda)^{1/3}$.
- (c) The radiation-dominated epoch is believed to have begun at the end of an inflationary epoch when the scale parameter had the value $a_{\text{inf}} \sim 10^{-28}a_0$. Calculate the time between the end of inflation and the beginning of matter domination.
- (d) The radiation temperature is approximately proportional to a^{-1} so one would expect that in the early universe it was sufficiently hot that bound structures (atoms, nuclei) were destroyed by photo-dissociation. This process prevented the formation of nuclei until $kT \sim 60 \text{ keV}$ corresponding to $a_{\text{nuc}} \sim 3 \times 10^{-9}a_0$. Estimate the time when the first nuclei formed.

1.3 Why is the universe expanding?

Chapter 2

Observational Cosmology

This chapter will present in a more detail the observational foundations of the standard cosmology. We first quickly pass in review the astronomical objects found on various scales. Section 2.1 is concerned with the compact objects that produce the visible photons in the universe either through combustion of nuclear fuel (in stars) or through gravitational collapse (in core collapse supernovae and quasars). The two subsequent sections show how stars are grouped together in galaxies and in clusters of galaxies. In Sect. 2.5 we will review present efforts to elucidate the nature of the dark matter that dominates the mass of the universe. Finally, in Sect. 2.6, we summarize the present measurements of the parameters of the standard model of cosmology i.e., the expansion rate and the various universal densities. A discussion of observations of the CMB photons will be delayed until Chap. 7.

2.1 Stars and Quasi-stars

Among the several thousands of astronomical objects in the sky visible to the naked eye, nearly all are nearby stars ($1\text{ pc} < R < 50\text{ pc}$). The exceptions are the Sun ($R = 4.8457 \times 10^{-6}\text{ pc}$), the Moon, 5 planets, 4 galaxies, and occasional comets. The galaxies are M31 (Andromeda), the Large and Small Magellanic Clouds, and our own Milky Way. While observational cosmology is primarily concerned with supergalactic scales, our understanding of the universe would be impossible without some understanding of stars. In this short section we can only hope to give a superficial account of the astrophysics necessary for cosmology.

Stars begin their lives as diffuse clouds composed primarily of hydrogen and helium. The first generation had the primordial mixture (75% hydrogen and 25% helium by mass) that was produced by nuclear reactions in the early universe. As the clouds contract gravitationally, the negative gravitational binding energy is compensated by increasing the thermal energy of the cloud and by radiating photons (Exercise 2.9). The contraction stops (temporarily) when the core is sufficiently hot to initiate nuclear reactions.¹ At this point, the energy radiated is compensated by

¹ Nuclear reactions occur only at high temperature, $T \sim 1\text{ keV}$, because at low temperatures they are exponentially inhibited by the Coulomb barrier between the positively charged nuclei.

the released nuclear energy and the star can exist in a steady state as long as its nuclear fuel holds out.

The first series of nuclear reactions transforms hydrogen into helium. Stars in this phase of nuclear burning are called “main sequence stars.” The luminosity of a main sequence star is roughly proportional to the third power of its mass (Exercise 2.9). Since the amount of nuclear fuel is roughly proportional to the mass, the total duration of the hydrogen-burning phase is proportional to M^{-2} . Table 2.1 gives some mass-dependant characteristics of main sequence stars.

After exhausting their hydrogen, helium can be burned to heavy elements in the “giant” phase. The helium is burned to carbon, to oxygen, and, in the most massive stars, to ^{56}Fe , the third most highly bound nucleus.²

Some of the heavy elements produced in stars will eventually be dispersed into interstellar space during supernova explosions, the final event in the lives of some stars with masses greater than $\sim 2M_{\odot}$. The result of this heavy element dispersion

Table 2.1 Characteristics of some representative main sequence stars [36]. Stars are classified according to their spectral type (O, B, A, F, G, K, M: “Oh Be A Fine Gnu, Kiss Me”) ranging from hot to cold. Each class is subdivided into subclasses (0-9). The table shows five examples ranging from O5 (heavy, bright, hot) to M5 (light, dim, cold). The first two lines show the correlation between mass and luminosity (heavier-brighter). The lifetime τ on the main sequence corresponds to the time necessary to burn the hydrogen in the stellar core (about 10% of the total hydrogen) at the roughly constant luminosity with bright-heavy stars having a shortest lifetimes. The surface effective temperature T_s and the radius R are related by Stephan’s law with bright-heavy stars being hot and big. The next two lines show how the luminosity and surface temperature are reflected in the absolute visual magnitude M_V and color index $B - V$. The last line gives the number density of stars per unit magnitude near our position in the Milky Way. It shows that most stars have relatively small masses

Type	O5	A0	G2	K5	M5
M/M_{\odot}	60	2.9	1.0	0.67	0.21
$L_V/L_{V\odot}$	8×10^5	50	1.0	0.15	0.01
$\tau/10^{10} \text{ yr}$	10^{-4}	0.05	1	4	200
T_s (Kelvin)	44500	9520	5860	4350	3240
R/R_{\odot}	12.	2.4	1.0	0.72	0.27
M_V	-5.7	0.6	4.7	7.4	12.3
$B - V$	-0.33	-0.02	0.65	1.15	1.64
$n (\text{pc}^{-3} \text{mag}^{-1})$	10^{-8}	10^{-4}	3×10^{-3}	3×10^{-3}	10^{-2}

² The most highly bound nuclei are ^{58}Fe and ^{62}Ni but they cannot be produced in stars by two-body reactions.

is that later generations of stars will have small admixtures of elements heavier than helium (“metals” according to the astronomical jargon). For example, the Sun started its life 4.5×10^9 years ago with $\sim 28\%$ helium (by mass) and $\sim 2\%$ metals. We see that even recently formed stars have a nuclear composition that is not too far from primordial. The great variety of stars is thus due mostly to differences in stellar masses which vary from $\sim 0.1 M_\odot$ to $\sim 50 M_\odot$ (Table 2.1).

The observable spectrum of photons radiated from stellar surfaces is only indirectly related to stellar structure. Photons in stellar interiors random-walk through the star until they happen to reach a radius where the density is sufficiently low that they escape. The shell from which average photons escape is called the “photosphere.” If photons of all wavelengths interacted with the same cross-section, the escape radius would be wavelength independent and stellar spectra would be nearly perfect blackbody spectra reflecting the temperature of the photosphere. The presence near the photosphere of atoms and molecules that are not completely ionized results in a wavelength-dependent photon cross-section. Photons with high cross-sections escape at larger, and therefore colder, radii leading to lower fluxes at the corresponding wavelengths. Two examples of spectra are shown in Fig. 2.1.

In the (rather poor) approximation that the photosphere of a star is a blackbody of a unique temperature, the luminosity of the star is given by Stefan’s law, $L = \sigma T_s^4 \pi D^2$ where D is the diameter of the photosphere. A measurement of the flux $f = L / 4\pi R^2$ then gives an estimation of the angular size of the star:

$$\Delta\theta = \frac{D}{R} = 2 \left(\frac{f}{\sigma T_s^4} \right)^{1/2}, \quad (2.1)$$

where R is the distance to the star.

Of the quantities listed in Table 2.1 only the surface temperature, T_s , and “color index,” $B - V$, are directly measurable from the shape of the observed photon spectrum. The other quantities can be deduced only if the distance to a star is known. We therefore now turn to the fundamental problem of determining stellar distances.

Apart from the distance to the Sun, which can be accurately determined by radar, stellar distances are extremely difficult to measure. Distances to stars of known diameter can be estimated through (2.1). Diameters can be directly determined for stars in binary systems that happen to be oriented so that the two stars periodically eclipse each other (Exercise 2.8). The diameter of a star whose photosphere is expanding, e.g. pulsing stars or supernovae, can be determined from the photosphere’s velocity as determined by the position and shape of its spectral lines (Sect. 2.6.1).

Other than these and other relatively rare exceptions, distances can be directly determined only for nearby stars via their “parallax,” i.e. their apparent annual movement with respect to more distant stars that results from the Earth’s movement around the Sun. The principle of this technique is illustrated in Fig. 2.2. The excellent angular resolution of the telescope on the Hipparcos satellite [40] that

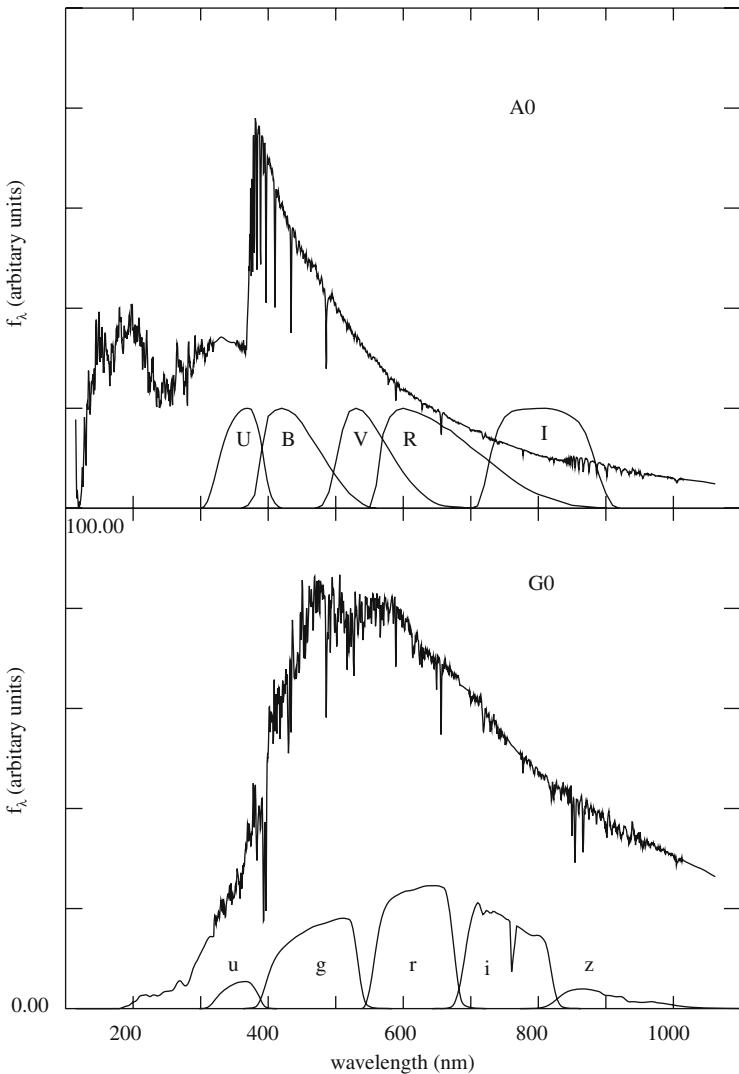


Fig. 2.1 The spectrum (energy per unit wavelength) of an “A0” star (*top*) and a “G0” star (*bottom*) [37]. For A0 stars (like Vega), the photosphere is sufficiently hot that much of the hydrogen is in the $n = 2$ atomic state so the Balmer series of hydrogen lines ($n = 2 \rightarrow n'$) is clearly present. The strong absorption at wavelengths shorter than the “Balmer break” at ~ 370 nm is due to the large cross-section for photo-ionization of $n = 2$ hydrogen and to absorption by closely spaced lines. For G stars (like the Sun), the photosphere is much cooler and there is little flux in the ultraviolet. Superimposed on the A0 spectra are the transmissions of the standard UBVRI filters of the Johnson–Cousins system [38] and on the G0 spectrum the ugriz transmissions (CCD response included) of the SDSS filter system [39]

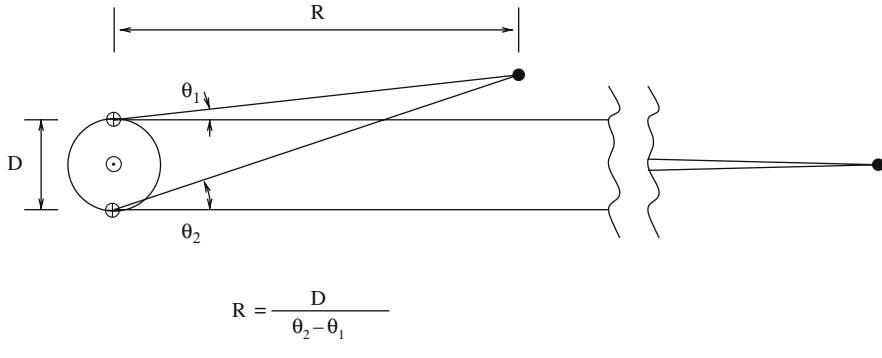


Fig. 2.2 The measurement of the distance R of a nearby star by parallax. The angular separation θ between the nearby star and a distant star undergoes an annual modulation resulting from the Earth’s movement around the Sun. For the realistic case of small modulations, the amplitude is $\theta_2 - \theta_1 = D/R$ where $D = 2AU = 3 \times 10^{11}\text{m}$ is the diameter of the Earth’s orbit. The modulation is 1 arcsec for a star at a distance of 1 pc (which explains the origin of the unit)

observed from 1989 to 1993 allowed the measurement of the distances of $\sim 10^5$ stars within ~ 200 pc to a precision of order 10%.

For stars with distances R determined by parallax, the luminosity can be calculated from the measured photon flux $f = L/4\pi R^2$, after corrections are made for absorption by interstellar dust.³ The calculated luminosity as a function of observed surface temperature is shown in the “color-magnitude” diagram for Hipparcos stars in Fig. 2.3. The luminosity L_V in the wavelength band “V” ($\lambda \sim 550\text{ nm}$) is given on an inverse logarithmic scale of “absolute magnitude,” M_V :

$$M_V = -2.5 \log L_V + \text{constant}. \quad (2.2)$$

(The magnitude system is explained in Appendix D.) The horizontal scale gives the “color index” $B - V$ which is the difference in magnitude in the “B” band ($\lambda \sim 450\text{ nm}$) and the V band. Stars with relatively low surface temperatures have large $B - V$ and stars with relatively large surface temperatures have small $B - V$.

Most of the stars in the color-magnitude diagram are on the diagonal strip that the theory of stellar structure identifies as the main sequence. The heaviest and brightest stars have the highest surface temperatures while the lightest and dimmest stars have the lowest surface temperatures.

Certain post-main sequence stars are on the diagonal extension to the cold side of the main sequence. Especially prominent are the helium-burning “clump” giants near $M_V \sim 0.75$ and $(B - V) \sim 1.0$.

³ The local absorption length of light depends strongly on the line of sight. In the galactic plane it is typically of order 1 kpc for photons with $\lambda \sim 550\text{ nm}$ [41]. Since absorption by dust is a decreasing function of wavelength, the absorption can be estimated by the “reddening” of a star’s spectrum compared to spectra of nearby unreddened stars of the same type.

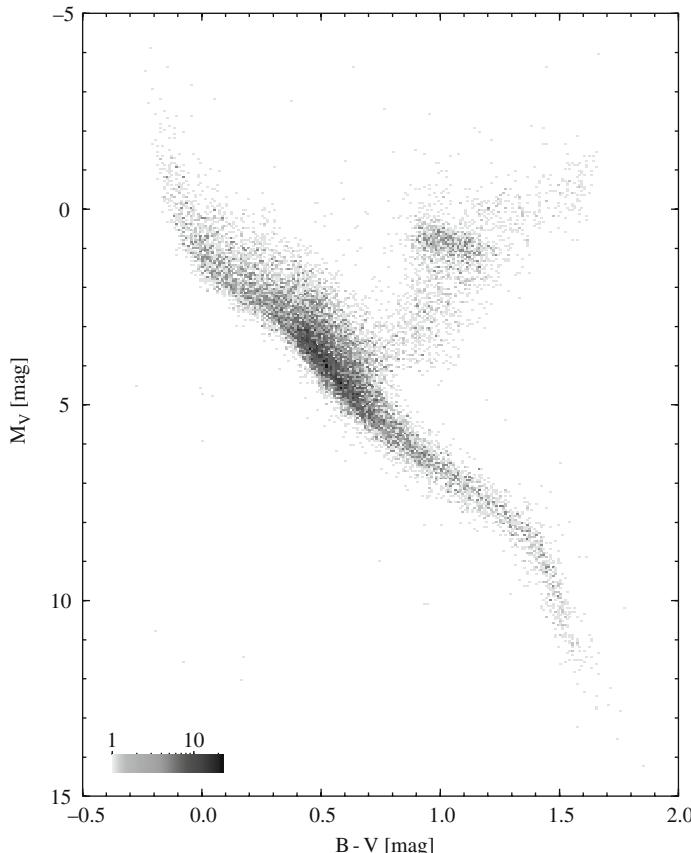


Fig. 2.3 The “color-magnitude” diagram for stars with parallax distances determined with a precision of better than 10% by the Hipparcos satellite [40]. The ordinate is the absolute magnitude in the V band ($\lambda \sim 550$ nm): $M_V = -2.5 \log(L_V) + \text{constant}$ where L_V is the luminosity in the V band. (The magnitude system is explained in appendix D.) The abscissa is the color index ($B - V$), i.e. the difference in magnitudes between the B ($\lambda \sim 450$ nm) and V bands. The color index is an indicator of a star’s surface temperature. Stars that are bright and hot are on the upper left while stars that are dim and cold are on the lower right. The diagonal band corresponds to stars on the main sequence (hydrogen burners). Post-main sequence stars are in the diagonal extension on the cold side of the main sequence. Stars in the accumulation at $M_V \sim 0.75$ and $(B - V) \sim 1.0$ are helium burning “clump”giants. A small number of dead stars (white dwarfs) are present near $M_V \sim 12$, $(B - V) \sim 0$. Courtesy of the European Space Agency

The correlations between *luminosity* and color (Fig. 2.3) generate correlations between *flux* and color for stars grouped at a given distance. Figure 2.4 shows the color-magnitude diagram for stars in the Large Magellanic Cloud (LMC). The flux f_V in the wavelength band V is given on an inverse logarithmic scale of “apparent magnitude,” V :

$$V = -2.5 \log f_V + \text{constant}' \quad (2.3)$$

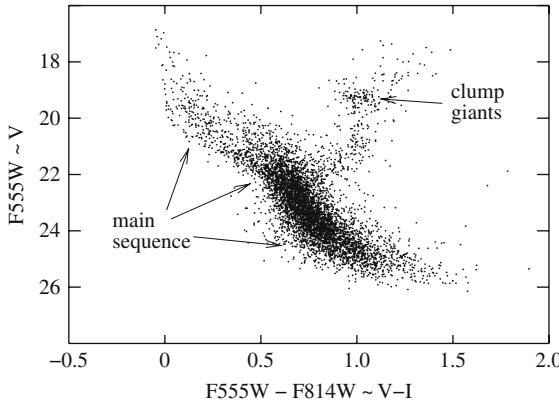


Fig. 2.4 The color-(apparent) magnitude diagram for stars in the Large Magellanic Cloud measured with the Hubble Space Telescope [42]. The ordinate is the apparent magnitude in the F555W band: $F555W = -2.5 \log(f_{555W}) + \text{constant}$ where f_{555W} is the flux in the 555W band. (The mean wavelength of F555W, 525 nm, is near that of the standard V band of Fig. 2.3). The abscissa is the color index ($F555W - F814W$), i.e. the difference in magnitudes in the F555W band and the F814W band ($\lambda \sim 827$ nm). The color index is an indicator of a star's surface temperature. A comparison of the apparent magnitudes of stars in the LMC with the absolute magnitude of the corresponding stars in the Milky Way (Fig. 2.3) allows one to estimate the distance to the LMC

The abscissa is a color index that is an indicator of surface temperature. One is struck by the similarity between this diagram and the Hipparcos color magnitude diagram in Fig. 2.3. The main sequence and clump giants are correctly positioned suggesting that the intrinsic luminosities of the stars in the LMC are nearly the same as those near the solar system. Under this assumption, we can easily estimate the distance to the LMC by using $f_V = L_V / 4\pi R^2$. In the magnitude system, this is easy to do because the constants in the magnitude definitions were chosen so that in the absence of absorption, the apparent magnitude is equal to the absolute magnitude of a star at a distance of 10 pc:

$$V = M_V + 5 \log(R/10 \text{ pc}) + A , \quad (2.4)$$

where A takes into account absorption. Using the V band magnitudes of the clump giants ($M_V \sim 0.75$ and $V \sim 19.25$ in the LMC) and ignoring correction due to absorption (small in this case), we find

$$R_{\text{LMC}} \sim 10 \text{ pc} \times 10^{0.2(19.25 - 0.75)} \sim 50 \text{ kpc} . \quad (2.5)$$

This turns out to be correct to within 10%.

The traditional “distance ladder” techniques for measuring galactic distances are variations of this calculation of the LMC distance. They all use different types of objects of known luminosity called collectively “standard candles.” The technique is accurate only if absorption is negligible or estimable and if the intrinsic difference in

luminosities between the distant and nearby candles can be estimated. In the present example, it is known that the chemical composition of stars in the LMC is slightly different from that of stars in the solar neighborhood. This can make the luminosity of, e.g., clump giants in the solar neighborhood slightly different from those in the LMC and thereby modify distance estimates if not taken into account [43]. We note that the use of different standard candles gives LMC distances that differ by of order 5% [9].

The use of main sequence stars or clump giants as distance indicators is possible only for distances < 1 Mpc beyond which these stars are too dim to be resolved. For distances up to ~ 50 Mpc it is possible to use Cepheid variable stars as distance indicators. Cepheids are a class of post-main sequence stars that have periodic luminosities. Their absolute magnitudes are $M_V \sim -4$ or about 100 times brighter than clump giants and thus identifiable at greater distances. Cepheid luminosities are well-defined functions of their periods as can be seen in Fig. 2.5 for LMC Cepheids. The absolute magnitudes of Cepheids can be determined from the small number that have Hipparcos parallax measurements [44]:

$$M_V = -1.43 - 2.81 \log P , \quad (2.6)$$

where P is the period in days. However, in view of the great number of LMC Cepheids, most authors prefer to use LMC Cepheids to calibrate the period–

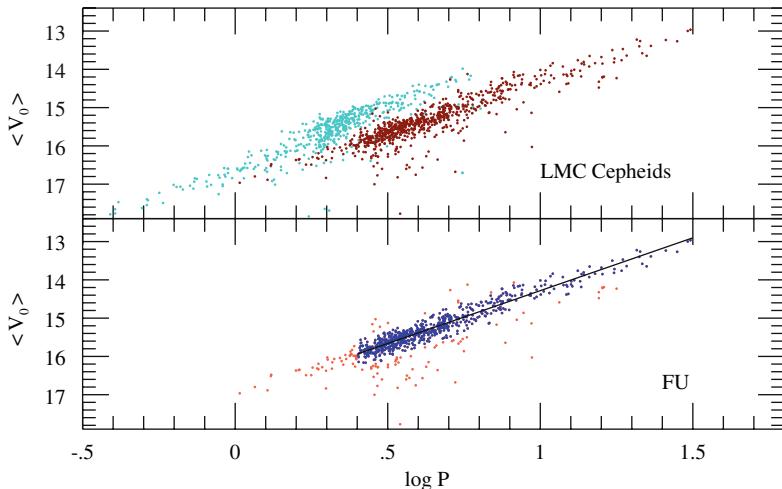


Fig. 2.5 The apparent magnitude, V , ($= V_0$) as a function of the period P in days for a sample of Cepheids in the LMC from the OGLE collaboration [45]. The *upper panel* shows two types of Cepheids, “fundamental” Cepheids that pulse in the fundamental frequency and “first overtone Cepheids” that have periods one half that of fundamental Cepheids. The *lower panel* shows only the fundamental Cepheids where the best fit is $V = -2.765 \log P + 17.044$. The Cepheid apparent magnitudes can be used to determine the distance to the LMC by using the absolute magnitudes of Hipparcos Cepheids. Alternatively, the apparent magnitudes can be used to find the Cepheid absolute magnitudes by using the LMC distance determined by other methods

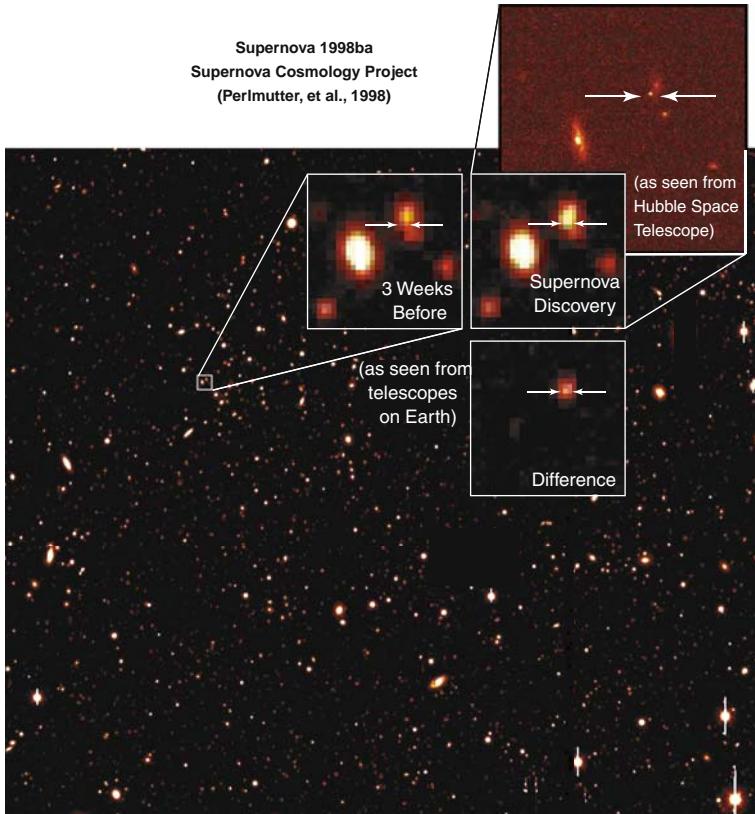


Fig. 2.6 The discovery of a supernova at $z = 0.43$ by the Supernova Cosmology Project [46]. The figure shows a 0.06 deg^2 portion of a CCD image of the sky taken with a ground-based telescope. Most objects are distant galaxies. The panels labeled “3 weeks before” and “supernova discovery” show zooms of the large image taken at three-week intervals. The “supernova discovery” panel shows that one of the galaxies has an apparent increase in luminosity and a slight change in shape due to a supernova explosion during the time between the two exposures. In the “difference” panel, the numerically subtracted image shows only the supernova. The final panel shows the same region of the sky taken with the Hubble Space Telescope. Because it does not suffer from the blurring effects of the Earth’s atmosphere, the HST image shows the supernova explosion took place at the edge of its host galaxy. Courtesy of Saul Perlmutter

luminosity relation. This strategy, of course, must use a LMC distance determined using other objects, e.g. clump giants or eclipsing binaries (Exercise 2.8).

The brightest “stars” are supernova explosions (Fig. 2.6) that can occur in sufficiently heavy stars at the end of their lives. It is believed that there are two fundamental types of supernovae. The first type consists of “core collapse” supernovae. The progenitor of such supernovae are massive stars ($M > 5M_{\odot}$) in which the core has burned all the way to ^{56}Fe . As soon as the core has grown to a mass of one

Chandrasekhar mass⁴ ($\sim 1.4M_{\odot}$), the core collapses to form a neutron star. It is believed that this mechanism is responsible for supernovae classified spectroscopically as types SNI_II, SNI_Ib, and SNI_Ic. The negative gravitational energy of the resulting neutron star⁵ is compensated by the radiation of neutrinos of total energy $\sim 10^{56}\text{J}$. Neutrinos are radiated rather than photons because their long scattering length makes it much easier for them to escape through the surrounding envelope of stellar matter than for photons. Since the neutron star mass is always close to the Chandrasekhar mass, these supernovae are “neutrino standard candles.” This hypothesis was confirmed by the detection of neutrinos from SN1987a in the LMC. The total energy radiated as photons is at most $\sim 10^{43}\text{J}$, much less than the integrated neutrino luminosity. The photon luminosity is mostly powered by the radioactive decay of ^{56}Ni and other radioactive nuclei in the material outside the collapsed core. This is seen in the photon luminosity as a function of time for SN1987a, shown in Fig. 2.7.

The second and brighter type of supernova, SNIa, are the most useful for cosmology. They are believed to be the thermonuclear explosion of carbon–oxygen white dwarfs that are pushed beyond the Chandrasekhar mass by matter falling on them from a binary partner. When the limit is reached, the star starts to implode and the resulting temperature increase results in the explosive burning of the carbon/oxygen to nuclei near ^{56}Ni . Most of the $\sim 10^{44}\text{J}$ of nuclear energy released is

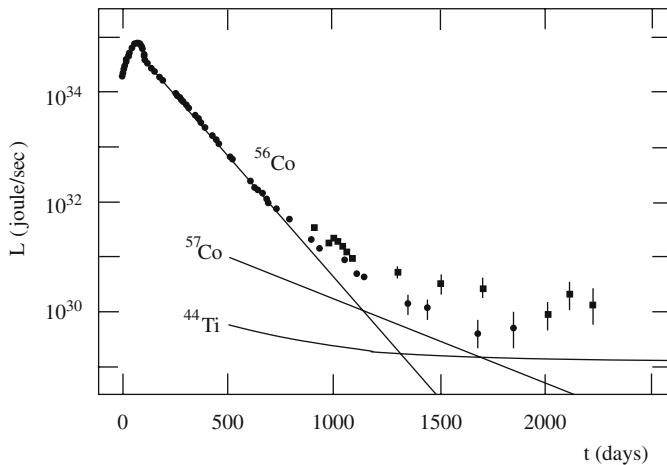


Fig. 2.7 The light curve (total luminosity versus time) for the type II supernova SN1987a in the Large Magellanic Cloud [47]. The curves show the energy released by the decay of the radioactive nuclei believed to be responsible for the photon luminosity

⁴ The Chandrasekhar mass $\sim (\hbar c/G)^{3/2}/m_p^2$ is the largest mass object that can be supported by pressure of a degenerate electron gas against gravitational collapse.

⁵ The negative of the gravitational binding energy of the neutron star is $\sim GM^2/r \sim 10^{56}\text{J}$ for a Chandrasekhar mass neutron star of radius $R \sim 3\text{ km}$.

converted to the kinetic energy of the nuclei resulting in the complete disintegration of the star. As with core-collapse supernovae, the photon luminosity is mostly powered by the radioactive decay of ^{56}Ni to ^{56}Co and then to ^{56}Fe . Since the total mass of ^{56}Ni is generally between 40% and 70% of the Chandrasekhar mass, type Ia supernovae are nearly “photon standard candles” making them very useful for distance determinations. They are bright enough to be seen at redshifts near unity. Their only problem is that they are rare, roughly one explosion per galaxy per century remaining visible over a period of a month.

The “light curve” (flux versus time) of a typical SNIa is shown in Fig. 2.8. As illustrated in Fig. 2.9, the luminosities at maximum light are correlated with the color at maximum light and the event time duration. An empirical relation is

$$M_B \sim -19.2 - 1.52(s - 1) + 1.57c \quad (2.7)$$

where $c = (B - V) + 0.057$ at maximum light and the “stretch” s can be roughly defined as the event time duration relative to the mean time duration of SNIa. This relation is the “brighter-bluer” and “brighter-slower” relation. While not completely understood it has plausible explanations. The brighter-bluer relation could reflect absorption in the atmosphere of the supernova or in the interstellar medium of the host galaxy. The brighter-slower relation could reflect the amount of ^{56}Ni produced since increased production implies both increased luminosity and increased opacity, resulting in a longer time scale for photon escape.

Note that Fig. 2.9 includes 6 SNIa that occurred in galaxies or galaxy clusters with Cepheid distances. These six supernovae calibrate the luminosity of type Ia supernovae for their use as standard candles.

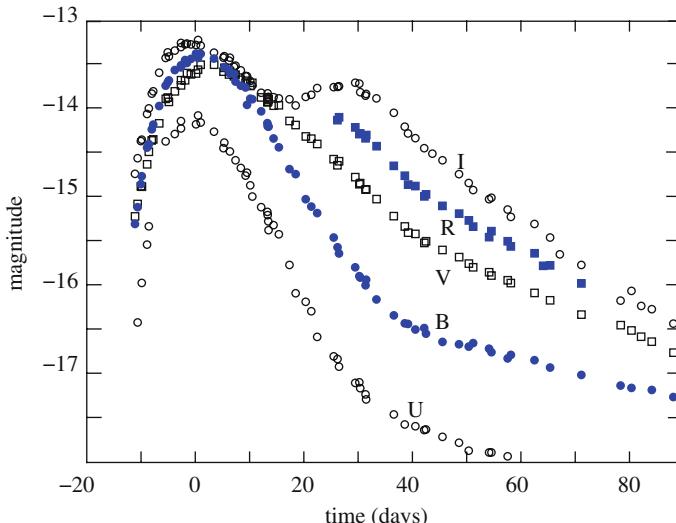


Fig. 2.8 The light curve (apparent magnitude versus time) for a typical type Ia supernova (sn2005cf [48]). The conventional (Vega-based) magnitudes in the Johnson–Cousin bands (UBVRI, as marked) have been offset by (0.70,-0.15,-0.01,-0.18,-0.43) to correspond to AB magnitudes reflecting the true flux. For clarity, the R-band curve has been shown only for $t > 30$ days

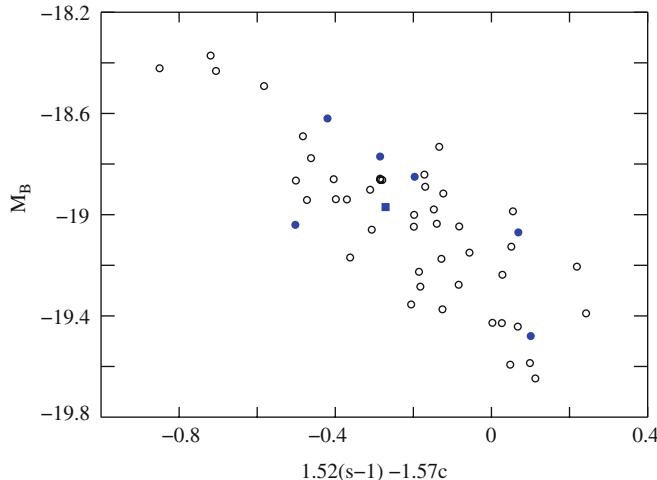


Fig. 2.9 The B-band absolute magnitude, M_B , vs. $1.52(s - 1) - 1.57c$, where s is the “stretch” and $c = (B - V) + 0.057$ at maximum light. The *open circles* are a collection of nearby ($0.01 < z < 0.03$) supernovae [2] for which the distance needed to convert apparent magnitudes to absolute magnitudes was calculated from the redshift assuming $h_{70} = 1$. The *filled circles* are supernovae in galaxies with Cepheid-determined distances [49]. The *filled square* is sn2005cf (Fig. 2.8). The figure illustrates the “brighter-bluer” and “brighter-slower” relations for type Ia supernovae

The most luminous known objects are “quasars” or “QSOs” (quasi-stellar objects). The luminosities can be up to 100 times that of a bright galaxy. They are believed to consist of a massive black hole $M > 10^6 M_\odot$ surrounded by gas clouds. If this is true, their source of energy would be the accretion of the surrounding gas by the hole during which the increasing gravitational binding energy of the accreted matter is accompanied by radiation of photons. The spectrum of a typical quasar is shown in Fig. 6.8. The spectrum is non-thermal with a prominent peak due to “Lyman- α ” emission by atomic hydrogen ($n = 2 \rightarrow n = 1$) in the surrounding gas. Being so bright, quasars can be found at very high redshifts, the highest to date being $z \sim 6$.

Quasars are important for cosmology for two reasons. First, being among the oldest identified objects, they provide important clues for theories of structure formation. In particular, structure formation must proceed sufficiently rapidly to have produced such objects by $z \sim 6$. The large quasar masses also suggest that they may have played an important role in the formation of the first galaxies, as also suggested by the fact that many galaxies now have central black holes of mass $\sim 10^6 - 10^7 M_\odot$ [50].

Second, quasar spectra show absorption lines due to elements in intervening clouds of gas so they serve as an important probe of the intergalactic medium. The spectrum in Fig. 6.8 exhibits a “Lyman- α ” forest of absorption lines on the blueward side of the quasar’s own Lyman- α emission. As photons travel to us from the quasar, their energies are degraded by the cosmological redshift. When passing through a cloud of redshift z_{cloud} , a photon will be resonant at Lyman- α if the photons original

energy E_1 satisfies $(z_{\text{quasar}}+1)/(z_{\text{cloud}}+1) = E_1/E_{Ly-\alpha}$, where $E_{Ly-\alpha}$ is the energy of the Lyman- α transition. The forest of absorption lines thus corresponds to a series of clouds between us and the quasar with each cloud absorbing just those photons whose energies have been degraded by the correct amount. The observed number density of such clouds places important constraints on structure formation theory and permits one to determine the quantity and chemical state of the intervening hydrogen. In particular, we will see in Chap. 5 that the fact that any of the photons blueward of Lyman- α reach us indicates that most of the hydrogen between us and high-redshift quasars is ionized. This fact is confirmed by the polarization of the CMB (Chap. 7).

2.2 Galaxies

Galaxies are clusters of stars, gas, and dark matter, the largest of which contain 10^{11} order of stars. They have a variety of morphologies, loosely classified as “ellipticals,” “spirals,” and “irregulars.”⁶ The Milky Way is a typical spiral galaxy and the characteristics of the solar neighborhood (Table 2.2) give an indication of the environment of a typical galactic disk 8 kpc from its galactic center. We note especially

Table 2.2 The sources of mass and luminosity in the solar neighborhood, 8 kpc from the galactic center [51, 18]. The dark matter is thought to have two components. The first is confined to the galactic disk and has a density that is estimated from the movement of stars perpendicular to the galactic plane. The second, “halo,” component is believed to be roughly spherical in shape. Its total mass is deduced from the galactic rotation curve

Component	Mass density ($M_\odot \text{ pc}^{-3}$)	Luminosity density ($L_\odot \text{ pc}^{-3}$)
Visible stars	0.044	0.067
Dead stars	0.028	0
Gas	0.042	0
Dark matter (disk)	< 0.07	0
Dark matter (halo)	0.003–0.017 (0.1–0.7 GeV cm $^{-3}$)	0

⁶ Elliptical and spiral galaxies are also referred to as “early-type” and “late-type” galaxies. This historical classification is misleading because elliptical galaxies are believed to be in a *later* stage of evolution, being the result of mergers of smaller (spiral) galaxies. A handy way or remembering the correspondence is to note the “elliptical” and “early-type” both begin with the letter “e” while “spiral” and the German word for “late” both begin with the letter “s.”

the mass-to-light ratio, $M/L \sim 2.5M_\odot/L_\odot$, which is, within a factor two or so, typical for the visible parts of most galaxies.

As for a star, the only directly observable properties of a galaxy are its position in the sky, its photon flux, and the shape of its spectrum (Fig. 2.10). The position and widths of spectral lines in the spectrum and their variation over the surface of a galaxy can be used to determine the galactic redshift as well as velocity dispersion of the stars. Luminosities can be determined only if the distance to the galaxy is known, as is the case for a few galaxies nearer than ~ 50 Mpc through the observation of Cepheid variable stars in the galaxy. The resulting luminosities have calibrated empirical relations between luminosities and velocity dispersions, the Faber–Jackson or Fundamental Plane relation for elliptical galaxies and the Tully–Fisher relation for spiral galaxies (Fig. 2.11). Once calibrated, these relations can be used to measure distances out to $R \sim 200$ Mpc.

Of course, the distance to any galaxy can be deduced from the redshift of its spectrum if one assumes a value of H_0 . Over the last decade, the large redshift surveys of galaxies, SDSS and 2dFRS, have produced $\sim 10^6$ redshifts, giving a rather complete picture of galaxies out to $z \sim 0.2$. Figure 2.12 shows the luminosity distribution of SDSS galaxies at $z \sim 0.1$. The number of galaxies per unit volume and per unit luminosity interval is well described by the “Schechter” distribution:

$$\frac{dn_{\text{gal}}}{dVdL} = \frac{\phi_*}{L_*} \left(\frac{L_*}{L} \right)^\alpha \exp(-L/L_*) \quad \alpha \sim 1 , \quad (2.8)$$

where ϕ_* , L_* , and α are constants. The distribution is particularly well measured at $z \sim 0.1$ by SDSS (Fig. 2.12) who gives, for luminosities measured near $\lambda \sim 560$ nm

$$\begin{aligned} \phi_* &= (0.511 \pm 0.016) \times 10^{-2} h_{70}^3 \text{Mpc}^{-3} \\ L_* &= (2.45 \pm 0.02) h_{70}^{-2} \times 10^{10} L_\odot . \quad \alpha = 1.05 \pm 0.001 . \end{aligned} \quad (2.9)$$

The factors of h_{70} comes from the use of Hubble’s law to determine galactic distances, $R = H_0^{-1}z$. The logarithmic divergence in the luminosity distribution at small L makes the total number density of galaxies ill-defined. However, the low-luminosity galaxies produce little light so the total light output per unit volume is (fortunately) finite. For SDSS galaxies, the integral of the Schechter distribution gives:

$$J = \int_0^\infty dL \phi_* \exp(-L/L_*) \sim L_* \phi_* = 1.29 h_{70} \times 10^8 L_\odot \text{Mpc}^{-3} \quad (2.10)$$

While the number of galaxies is not well defined, 95% of the light comes from galaxies brighter than $\sim 10^9 L_\odot$. The integral then gives the number density of bright galaxies:

$$n_{\text{gal}} = \int_{10^9}^\infty dL \frac{\phi_*}{L} \exp(-L/L_*) = 0.015 h_{70}^3 \text{Mpc}^{-3} . \quad (2.11)$$

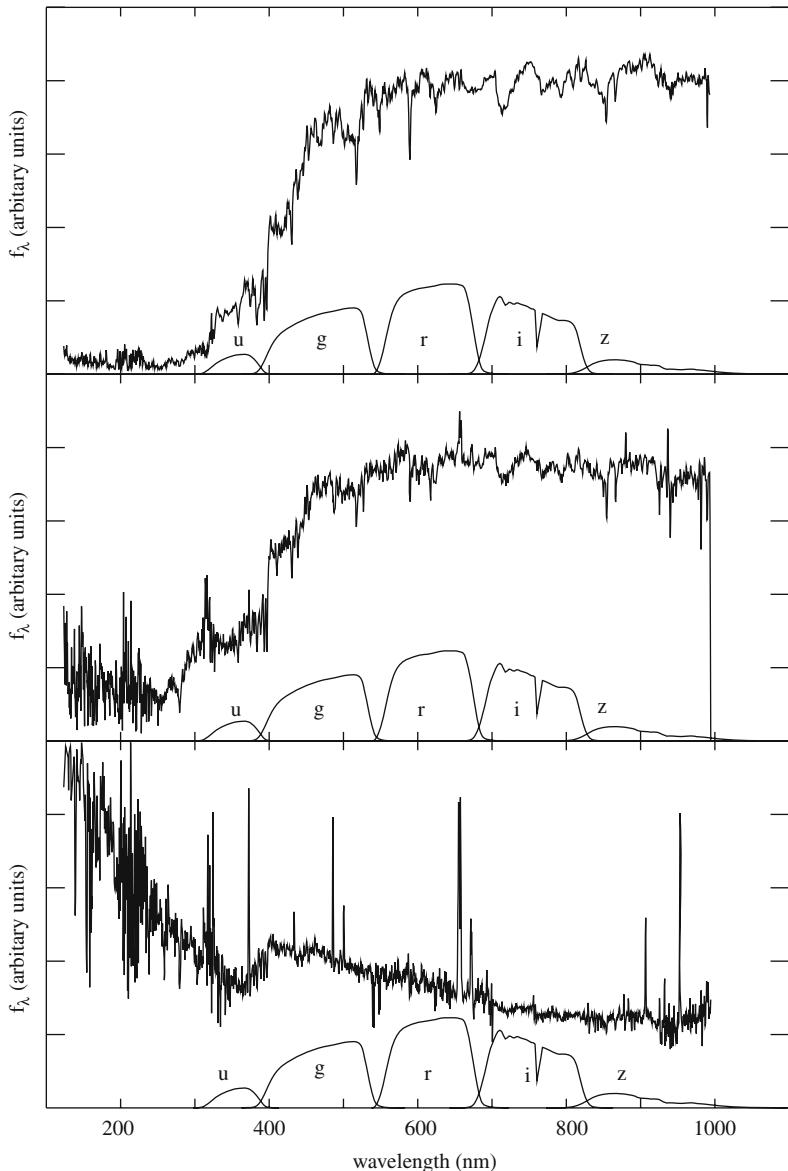


Fig. 2.10 Typical spectra (energy per unit wavelength) of elliptic (*top*), spiral (*middle*), and star-forming (*bottom*) galaxies [52]. Being the sum of its stellar spectra, galactic spectra show the same spectral lines as those in Fig. 2.1. Galactic emission lines (present in the lower spectrum) are due to hot interstellar gas that is excited by ultraviolet photons from massive stars or QSOs within the galaxy. Also shown are the bandpasses of the SDSS filters [39]

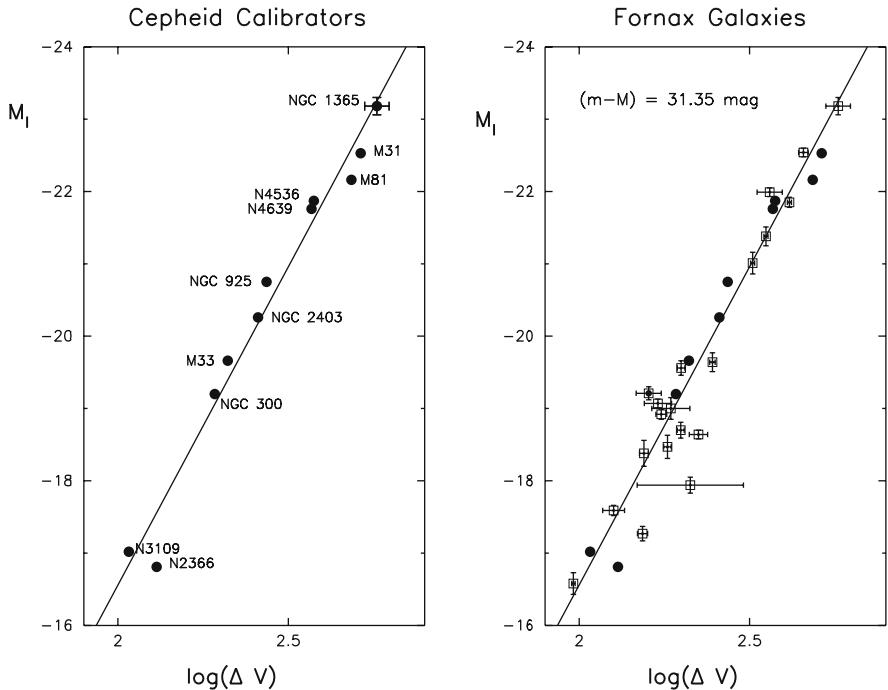


Fig. 2.11 The I-band absolute magnitude of spiral galaxies with Cepheid distances [53]. The magnitude is a linear function of $\log \Delta v$ where Δv is the velocity dispersion in km s^{-1}

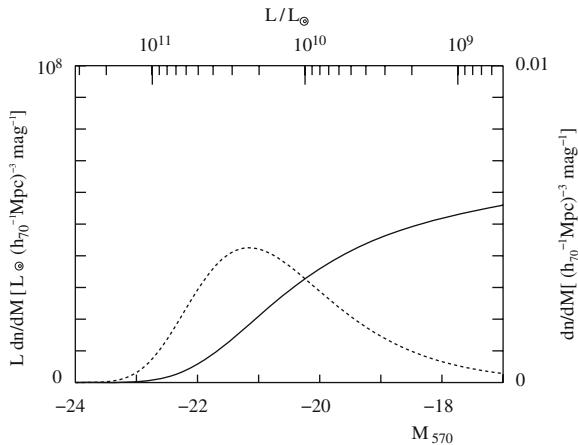


Fig. 2.12 The number density (solid line, right scale) and luminosity density (dashed line, left scale) of SDSS galaxies at redshift $z \sim 0.1$ [54]. The lower scale gives the absolute magnitude at $\lambda \sim 560 \text{ nm}$ ($M_{570\odot} = 4.76$) while the upper scale give the galactic luminosity in units of L_\odot

The total mass in stars for a given galaxy can be deduced from the galaxy's spectrum with the help of stellar evolution models. For SDSS galaxies [55], a galaxy with $L = L_*$ has on average a total stellar mass of $6 \times 10^{10} M_\odot$ corresponding to $M/L = 6/2.45 = 2.45$. Multiplying this by J from (2.10) and dividing by the critical density gives the contribution of stars to the universal density

$$\Omega_{stars} = 0.0023 . \quad (2.12)$$

This is a factor 6 greater than the contribution of atomic hydrogen gas (HI) contained in galaxies deduced from galactic 21 cm emission [56]

$$\Omega_{HI} = 3.5 \times 10^{-4} . \quad (2.13)$$

There is a comparable amount of molecular hydrogen.

The visible parts of galaxies bathe in an extensive “halo” of dark matter. The density profile of these halos (and of clusters of galaxies) is often described by a “NFW” profile proposed by Navarro, Frenk and White [57] on the basis of n -body simulations of galaxy formation:

$$\rho(r) \propto \frac{1}{r(1+r/r_s)^2} , \quad (2.14)$$

where r_s parameterizes the size of the halo. The density has a singularity at $r = 0$ though this does not lead to a singularity in the potential or mass. The density falls like r^{-3} for $r \gg r_s$ so the integrated mass of the halo rises logarithmically with r and the total halo mass is undefined by the density profile. However, the halo mass can be defined as the mass within a radius where objects are effectively bound. Structure formation theory informs us that within this radius the mean density is of ~ 200 times the mean universal density (Eq. 7.17).

Because of its simplicity, a distribution of mass that is often used as a first approximation for galaxy clusters and galaxy halos is that for an “isothermal sphere”:

$$G\rho(r) = \frac{\sigma_v^2}{2\pi r^2} , \quad (2.15)$$

where σ_v is the one-dimensional velocity dispersion of objects bound in the potential. The r -independent rotational velocity for circular orbits is $\sqrt{2}\sigma_v$. Having an r dependence that is intermediate between the small and large r dependence of the NFW profile, the isothermal profile gives results that over a large range of r resemble those of more realistic profiles.

Galactic mass profiles can be derived from the velocities of stars or gas clouds orbiting a galaxy (Exercise 2.7). From Newtonian dynamics, the circular velocity $v(r)$ at a distance r from a galactic center is given by $v(r)^2 \sim GM(r)/r$ where $M(r)$

is the mass within r . Figure 2.13 shows the rotation curve, $v(r)$, of a spiral galaxy. The velocity is that of clouds of atomic hydrogen determined by the Doppler shift of the 21 cm hyperfine line of the hydrogen ground state. The rotation curve is flat for distances beyond the visible radius of the galaxy, indicating $M(r) \propto r$. Since the rotation curve stays flat out to the last measured point, $r \sim 30 h_{70}^{-1}$ kpc, one can deduce only a lower limit on the galactic mass or mass-to-light ratio: $M/L > 20 h_{70} M_\odot/L_\odot$. The factor of h_{70} comes from the use of Hubble's law to convert observed angular sizes $\Delta\theta$ to physical sizes $r = R/\Delta\theta = H_0^{-1}z/\Delta\theta$.

The study of weak gravitational lensing of background galaxies ($z > 0.5$) by foreground galaxies ($z < 0.2$) has overcome some of the limitations of rotation curve measurements. As discussed in detail in Sect. 3.8, the trajectories of light from

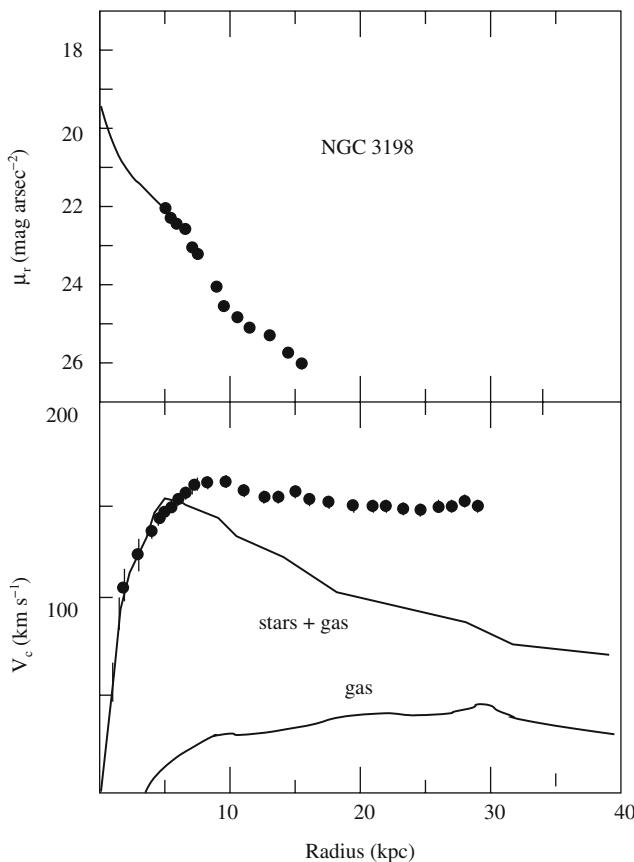


Fig. 2.13 The rotation curve of the galaxy NGC3198 [58]. The *upper panel* shows the luminosity distribution indicating that most of the luminous matter is concentrated at distances less than 5 kpc from the center. The *lower panel* shows the measured rotation curve. The curve remains flat far beyond the luminous radius indicating the presence of dark matter far from the center. The curve labeled “stars and gas” shows the rotation curve that would be expected if the mass distribution followed the luminosity distribution with a mass-to-light ratio of $3.8M_\odot/L_\odot$. The horizontal scale supposes $H_0 = 75 \text{ km s}^{-1}\text{Mpc}^{-1}$

background galaxies are bent in the gravitational field of a foreground galaxy. This results in a small distortion of the shape of the background galaxies that can be used to deduce the mass of the foreground galaxy (Eq. 3.113). For a given background galaxy, the effect is not measurable because the unperturbed shape is not known. However, if the shapes of many background galaxies are measured, one can find systematic stretching of the galaxies in the direction tangent to circles centered on the foreground galaxy. Since the effect of generally less than 1% for a given galaxy, it can only be seen by averaging over many foreground galaxies. This has been done with SDSS galaxies [55] yielding the density profiles shown in Fig. 2.14. The total galactic mass can be found by integrating the profile out to a radius where the mean density is a factor ~ 200 greater than the mean universal density. The data in Fig. 2.14 imply that a galaxy with $L = L_*$ is surrounded by a halo of mass $1.4 \times 10^{12} M_\odot$ [55], corresponding to a mass-to-light ratio

$$M/L = 140/2.45 \sim 60 . \quad (2.16)$$

Multiplying this by J (2.10) and dividing by the critical density gives the contribution of galaxies to the universal density

$$\Omega_{\text{gal}} = 0.054 . \quad (2.17)$$

Since $\Omega_M \sim 0.27$, this means that $\sim 80\%$ of the mass is in intergalactic space.

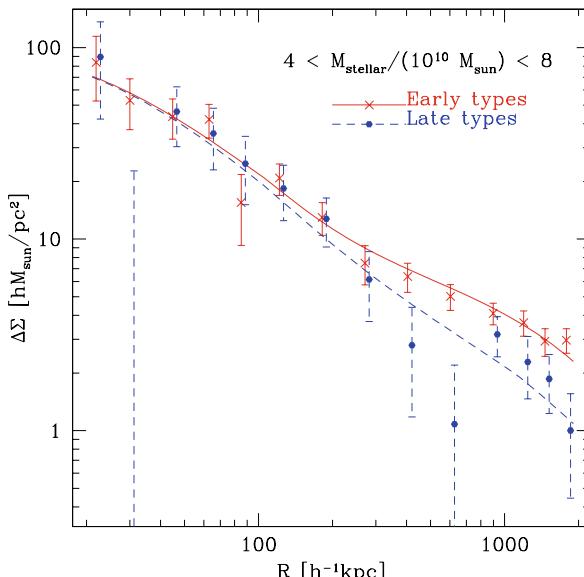


Fig. 2.14 Galactic mass surface densities (density integrated along the line of sight) as a function of distance from the galactic center as deduced from the weak lensing by SDSS galaxies of background galaxies [55]. The lenses are SDSS galaxies with $4 < M_{\text{stars}}/10^{10} M_\odot < 8$. The surface densities are deduced from the tangential shear given by Eq. (3.113). Figure courtesy of Rachel Mandelbaum

Finally, we note that one would expect that the ratio of total galactic baryonic mass to total galactic mass would be roughly the universal value, $\Omega_b/\Omega_M \sim 0.17$. For an L_* of mass $1.4 \times 10^{12} M_\odot$ this would imply a baryonic mass of $2 \times 10^{11} M_\odot$ to be compared with the estimated stellar mass $6 \times 10^{10} M_\odot$. This means that within galaxies, about 25% of baryonic material has been transformed into stars.

2.3 Galaxy Clusters

Galaxies are often gravitationally bound in groups. Our own galaxy is a member of the “Local Group” containing ~ 30 small galaxies plus three large spirals, the Milky Way, M31 (Andromeda), and M101. The largest known bound structures are “rich clusters” of galaxies that can contain thousands of galaxies and up to $10^{15} h_{70}^{-1} M_\odot$ in volumes of a few Mpc³. This mass corresponds to the mean mass contained in a sphere of radius $\sim 10 h_{70}^{-1}$ Mpc so regions within 1 Mpc of the centers of rich clusters have over-densities of order 1000. The nearest rich cluster is the Virgo Cluster at a distance of ~ 20 Mpc. The number density of clusters as a function of their mass is shown in Fig. 2.15. There are $\sim 10^{-5}$ clusters per $(h^{-1}\text{Mpc})^3$ with

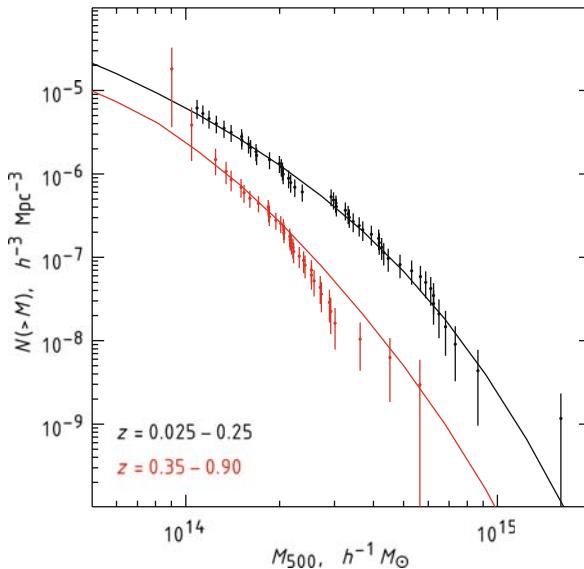


Fig. 2.15 The number density of clusters with masses greater than M_{500} as a function of M_{500} [100] as determined by Chandra X-ray observations. (M_{500} is the mass contained within the radius where the mean cluster density is a factor 500 greater than the universal density.) The data is shown for two redshift ranges. There are a factor ~ 3 more clusters at low redshift than at high redshift meaning that clusters are still in the construction process. The evolution of the number of clusters with redshift is consistent with that expected for a Λ CDM universe with $\Omega_M \sim 0.034 \pm 0.08$. (The ratio between the high- and low-redshift densities would be ~ 20 in a $\Omega_M = 1$ universe.)

a mean mass $\sim 10^{14} h^{-1} M_{\odot}$ which allows us to estimate the total mass contained in such large clusters:

$$\Omega_{\text{cluster}} \sim \frac{10^9 h^2 M_{\odot} \text{Mpc}^{-3}}{3H_0^2/8\pi G} \sim 0.003 \quad (2.18)$$

Comparing with Ω_{gal} (2.17), we see that $\sim 5\%$ of galaxies are in large clusters.

Clusters like that in Fig. 2.16 are rather ill-defined when viewed as simple collections of galaxies. They become much more distinct when observed through their X-ray emission. The X-rays are produced through bremsstrahlung by electrons in the ionized intergalactic gas. It is believed that this gas constitutes the majority of the baryonic mass of clusters.

Cluster masses can be most easily estimated by measuring the velocity dispersion of the member galaxies and then applying the virial theorem (Exercise 2.6). More modern techniques use the X-ray temperature (Exercise 2.10) or the shapes of background galaxies that are deformed by the gravitational lensing action of the cluster (Sect. 3.8). The most massive clusters have mass-to-light ratios of order 200

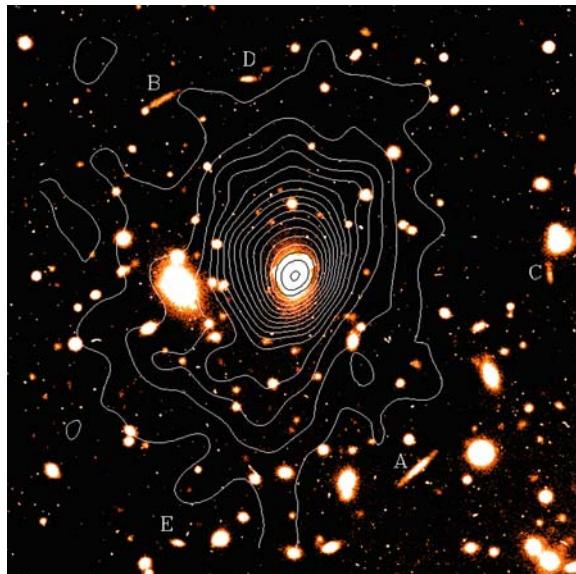


Fig. 2.16 The galaxy cluster RXJ1347.5-1145 [59]. The contours show the level of emission of X-rays by the ionized intergalactic gas in the cluster. The cluster is a “gravitational lens” for galaxies behind it. The images of distant lensed galaxies are in the arc-shaped objects marked with the letters A, B, C, D, and E. Courtesy of S. Schindler

$$\frac{M}{L} \sim 200 h_{70} M_\odot / L_\odot . \quad (2.19)$$

The mass of the gas in a cluster can be estimated from the observed X-ray flux (Exercise 2.10). It is an order of magnitude less than the total mass [99]:

$$\frac{M_{\text{gas}}}{M_{\text{total}}} \sim 0.12 h_{70}^{-3/2} \Rightarrow \frac{M_{\text{baryons}}}{M_{\text{total}}} \sim 0.16 h_{70}^{-3/2} . \quad (2.20)$$

where the second form includes the mass of stars and gas in individual cluster galaxies. This factor of 6 between total mass and baryonic mass is perhaps the best evidence for non-baryonic dark matter.

An important prediction of CDM models is that the dark matter is made up of particles that have very weak non-gravitational interactions. This was confirmed by observations of the “Bullet cluster” [60] shown in Fig. 2.17. Superimposed on the optical and X-ray images are mass contours as derived from gravitational lensing of background galaxies. One sees that the Bullet cluster actually consists of two clusters that recently passed through each other. The CDM and the galaxies have only gravitational interactions so these two components more or less retain their form after passing through the gravitational potential of the other cluster. On the other hand, the ionized gas consists of particles that scatter on each other. This causes “frictional” forces on the gas that slows it down with respect to the CDM. The result is a displacement of the gas and dark matter, as clearly demonstrated in the figure. This constitutes “visual proof” of the existence of non-collisional dark matter.

2.4 Large-Scale Structure

At scales above that of galaxy clusters, large redshift surveys have shown that galaxies appear to be grouped along filamentary “walls” sometimes surrounding great “voids” containing few visible galaxies. Figure 2.18 shows a “slice” of the sky as seen in by SDSS [61]. The largest walls or voids have sizes of order $100 h_{70}^{-1}$ Mpc. These structures are not bound and, at the present epoch, are still participating in the universal expansion though at a reduced rate because of their self-gravity. As discussed in Sect. 1.2.6, these density fluctuation are unlikely to form bound structures in the future because of the acceleration of the expansion caused by dark energy.

At scales > 100 Mpc the density is relatively uniform (Fig. 7.4). The characterization of the small fluctuations above these scales gives important clues about the matter content of the universe. This will be discussed in Chap. 7.

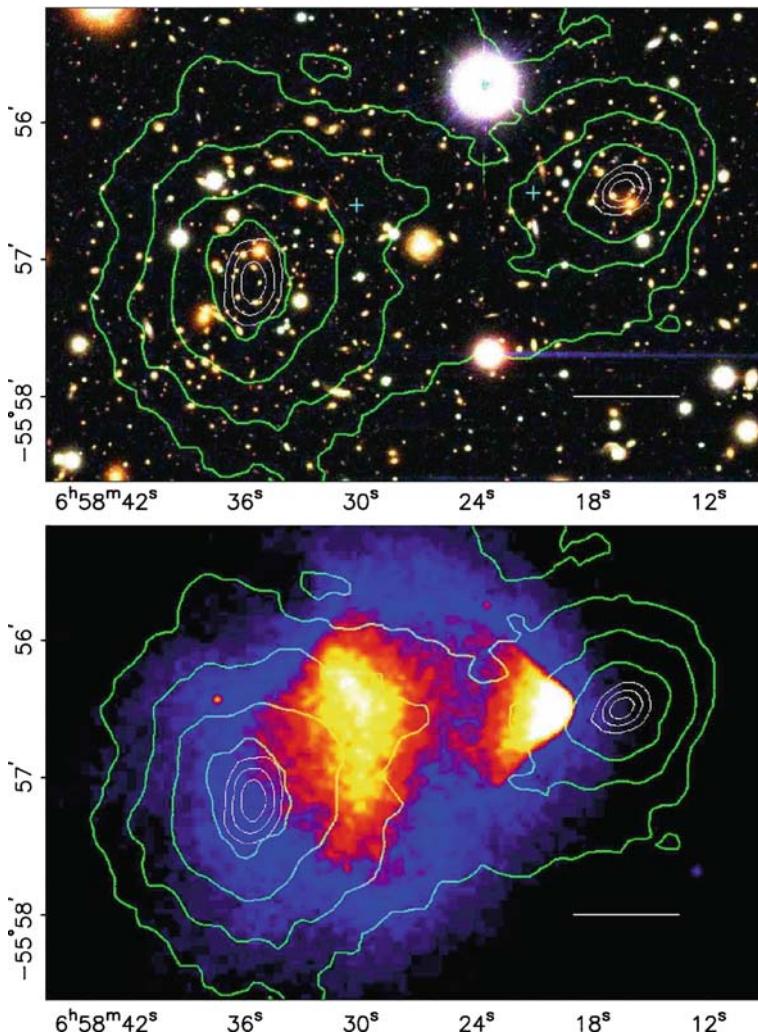


Fig. 2.17 Optical (top) and X-ray (bottom) images of the “Bullet Cluster” [60]. Mass contours as derived from a lensing analysis are shown. The cluster actually consists of two mass concentrations (shown by the lensing mass contours) with two concentrations of gas (shown by the X-ray image) displaced from the centers of mass. The interpretation is that the two clusters recently passed through each other causing the collisional gas to be displaced with respect to the non-collisional dark matter and galaxies. This provides “visual proof” of the existence of weakly interacting dark matter

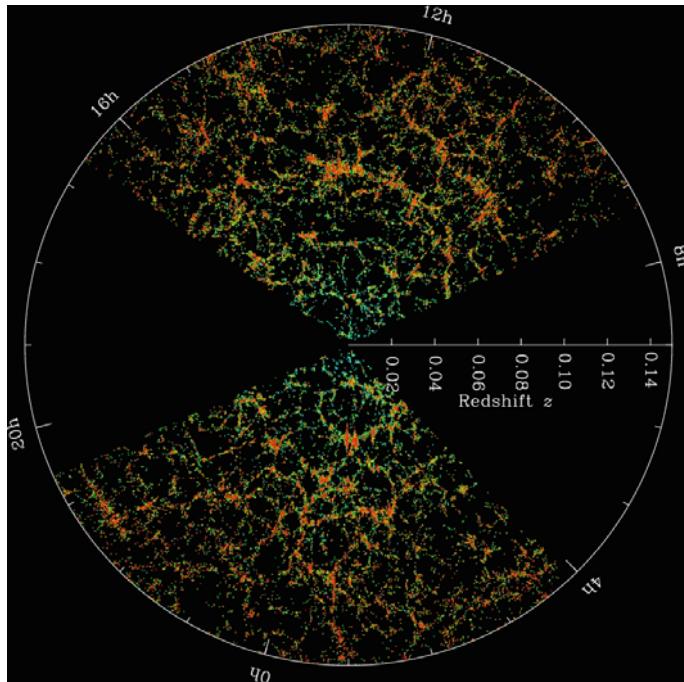


Fig. 2.18 Slices through the SDSS three-dimensional map of the distribution of galaxies. Each slice covers about 7 hours (105 deg.) in right ascension and 2.5 deg. in declination. Our galaxy is positioned at the intersection of the northern and southern slices. The radial coordinate is given as the redshift, z . “Walls” and “voids” are seen up to sizes of $\Delta z \sim 0.02 \Rightarrow \Delta R \sim\sim 86 h_{70}^{-1}$ Mpc. Courtesy of the Sloan Digital Sky Survey

2.5 Dark Matter

It is clear that there will always be doubts about the Λ CDM cosmological model as long the dark matter has not been identified. Here, we review efforts to detect the two favored non-baryonic candidates, WIMPs and axions. We also review limits on the numbers of dark astrophysical objects (MACHOs) in the galactic halo, which appear to have eliminated most such objects as candidates for galactic dark matter. Finally, we discuss the possibility that dark matter consists of cold gas.

2.5.1 WIMPs

Since the 1980s, WIMPs (weakly interacting massive particles) have been the standard cold dark matter (CDM) candidate. Supersymmetric extensions of the standard model of particle physics (invented to solve problems unrelated to dark matter) predict the existence of such particles. The fact that they have not yet been seen at accelerators suggests that they must have a mass $m_\chi > 30$ GeV [62]. The particles would

have been thermally produced in the early universe (Chap. 6) yielding a cosmological abundance inversely proportional to their annihilation cross-section. Supersymmetric models contain many free parameters yielding relic densities within a few orders of magnitude on either side of the critical density.

Today, WIMPs would be expected to inhabit the halos of spiral galaxies like our own. From the galactic rotation velocity, one can estimate the local density to be about 0.3 GeV cm^{-3} [17]. The orbital velocities of objects trapped in the Galaxy are of order 250 km s^{-1} so the local WIMP flux is of order $10^7 \times (1 \text{ GeV}/m\chi) \text{ cm}^{-2} \text{ s}^{-1}$.

Goodman and Witten [63] suggested that these WIMPs could be detected via the observation of nuclei recoiling from WIMP-nucleus elastic scatters. Galactic WIMPs with masses in the GeV range have kinetic energies in the keV range so we can also expect nuclear recoils in the keV range. The rate is proportional to the elastic WIMP-nucleus scattering cross-section which depends on the parameters of the particle physics model. Typical values of the supersymmetric WIMP-nucleon cross-section are of order 10^{-44} cm^2 , corresponding to a very weak interaction.

WIMP scatters can be observed with “calorimetric” techniques (Fig. 2.19). Unfortunately, it is difficult to distinguish WIMP events from events due to beta or gamma radioactivity (also shown in the figure). Statistically, a signal from WIMPs can be isolated through the expected $\sim 5\%$ seasonal modulation of the event rate [17, 64]. This modulation is due to the fact that while the Solar System moves through the (isotropic) WIMP gas, the Earth’s motion around the Sun alternately adds or subtracts from the WIMP-detector velocity. Alternatively, certain detectors can distinguish nuclear recoils from the Compton-electron background, e.g. hybrid cryogenic detectors that detect two types of excitations, phonons and ionization or

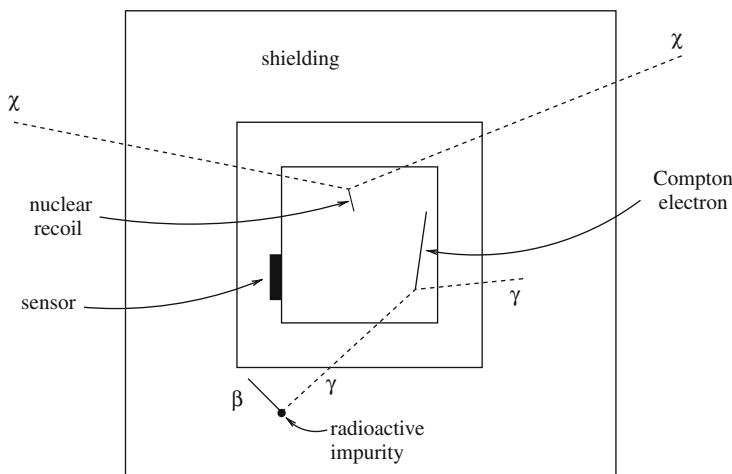


Fig. 2.19 A generic “calorimetric” WIMP detector surrounded by its shielding. The galactic WIMP χ enters the detector, scatters off a nucleus, and leaves the detector. The recoiling nucleus creates secondary excitations (e.g., scintillation light, ionization, phonons) that can be detected by the sensor. Also shown is a background event due to the ambient radioactivity yielding a Compton electron in the detector

phonons and scintillation. The ratio of the two signals is different for nuclear recoils and Compton electrons, allowing background rejection.

The present generation of experiments using hybrid cryogenic calorimeters [65] and xenon ionization–scintillator calorimeters [66] have backgrounds that make them sensitive to the supersymmetric WIMP candidates with the highest scattering cross-sections. The present limits are shown in Fig. 2.20. Particles with $m_\chi \sim 100\text{GeV}$ must have cross-sections on nucleons less than $\sim 10^{-43}\text{cm}^2$ to have escaped detection. The cross-section limits are very weak for particles with $m < 1\text{GeV}$ because such light particles create nuclear recoils with energies too small to be detected.

We note that the limits shown in Fig. 2.20 are in nominal conflict with an experiment using NaI detectors [67]. This experiment reported an annual modulation of the event rate similar to what one would expect for WIMPs. The source of the disagreement is not understood.

Besides direct detection, it is possible to detect WIMPs “indirectly” through the detection of particles produced in present-day WIMP-antiWIMP annihilation. While WIMP annihilation ceased in the early universe because of the universal expansion, it started up again once the WIMPs became gravitationally bound in galactic halos. These annihilations are a source of cosmic-ray photons, electron/positrons and proton/antiprotons and the observed fluxes are sometimes interpreted as being due to WIMP annihilation [68] but it has so far proved impossible to rule out astrophysical sources for the particles.

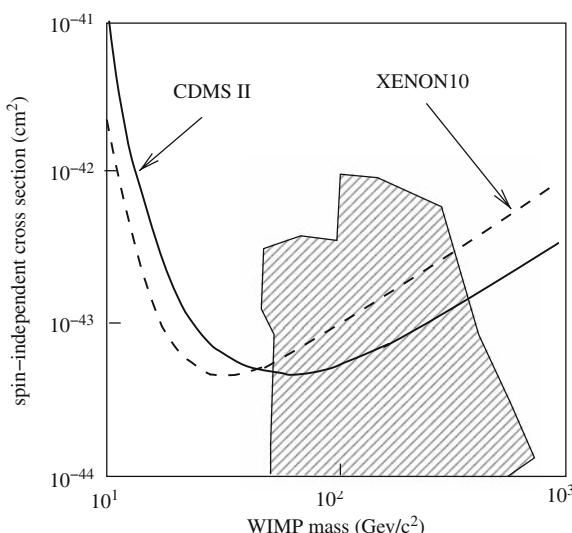


Fig. 2.20 Current limits on WIMP mass and WIMP-nucleon scattering cross-section from direct-detection experiments using xenon [66] (dashed line) and germanium [65] (solid line). The region above and to the right of the curves are excluded. The shaded regions show the region of mass-cross-section space favored by supersymmetric models

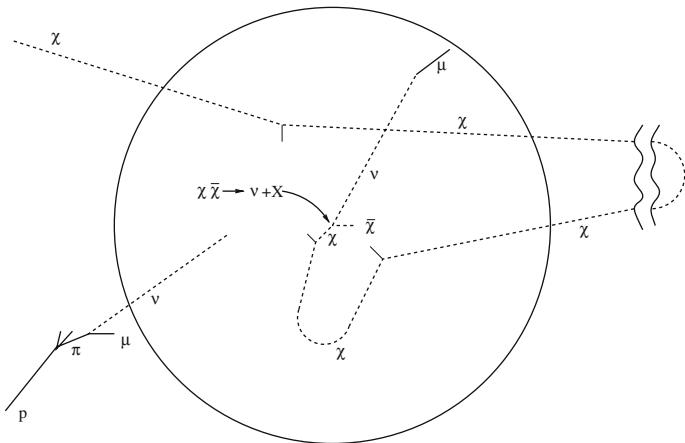


Fig. 2.21 The capture of a WIMP χ in the Earth. If the WIMP loses sufficient energy in a collision with a nucleus, the WIMP's velocity will drop below the escape velocity placing it in an orbit intersecting the Earth. The subsequent collisions will eventually thermalize the WIMP in the center of the Earth. An annihilation with a thermalized anti-WIMP may lead to the production of neutrinos that can be detected at the surface. As shown, the background for such events comes from neutrinos produced by decays of pions and kaons produced by cosmic rays in the Earth's atmosphere

The annihilation rate is further enhanced inside material objects like the Sun or Earth. This is because it is possible for WIMPs to be trapped in such objects if, while traversing the object, the WIMP suffers an elastic collision with a nucleus (Fig. 2.21). If the scatter results in a WIMP velocity below the object's escape velocity, the WIMP will find itself in an orbit that passes through the object. After repeated collisions the WIMP will be thermalized in the core. In the case of supersymmetric dark matter, the trapping rate in the Sun is sufficiently high that the concentration of WIMPs reaches a steady state where trapping is balanced by either annihilation (for high-mass WIMPs) or by evaporation (for low-mass WIMPs). For the Sun, the dividing line between low mass and high mass is $\sim 3 \text{ GeV}$ so accelerator limits ($m > 30 \text{ GeV}$) suggest that WIMPs should be annihilating inside the Sun.

The only annihilation products that can be seen emerging from the Sun or Earth are, of course, neutrinos. The flux of such neutrinos can be calculated for a given WIMP candidate and the flux compared with that observed in underground detectors. The observed flux is entirely understood as being due to the decay of cosmic-ray pions and kaons in the Earth's atmosphere (Fig. 2.21). Certain supersymmetric WIMPs would give a higher flux and are thus excluded [69, 70].

Because the observed flux of neutrinos is due to an unavoidable background, the only improvements in the limits from these techniques would come from the observation of a small excess of neutrinos coming from the direction of the Sun or center of the Earth. The most reasonable possibility is to search for upward-going muons coming from ν_μ interactions in the rock below a detector. Calculations [71] indicate that a 1 km^2 detector with a muon energy threshold of $\sim 10 \text{ GeV}$ would be needed

to observe a statistically significant solar signal for typical supersymmetric dark matter. Efforts in this direction are underway by instrumenting the Mediterranean [72] or the Antarctic Continental Glacier [73] to observe Cerenkov light produced by muons.

2.5.2 Axions

Axions [74] are hypothetical light scalar particles invented to prevent CP violation in the strong interactions.⁷ They would have been produced in the early universe via both thermal and non-thermal mechanisms and might produce near-critical relic densities if they have masses in the range $m_a \sim 10^{-5}$ eV to $\sim 10^{-3}$ eV. It was also recently emphasized [29] that axions with smaller masses, $m_a \sim 10^{-8}$ eV, could give a near critical density in some models that require the application of anthropic selection.

Axions act as cold dark matter and should be present in the galactic halo. The most popular detection scheme for galactic axions is based on the expectation [75] that axions can “convert” to a photon of frequency $\nu = m_a c^2 / h$ in the presence of a magnetic field. If a microwave cavity is tuned to this frequency, the axions will cause an excess power to be absorbed (compared to neighboring frequencies). If the halo is dominated by axions, the predicted power is small, about 10^{-21} W for a cavity of volume 3 m^3 and a magnetic field of 10 T. Since the axion mass is not known, it is necessary to scan over the range of interesting frequencies. Pilot experiments [76, 77] have produced limits on the local axion density about a factor 30 above the expected density. Experiments are now in progress to search for axions of masses in the range $m_a \sim 10^{-5}$ eV to $\sim 10^{-3}$ eV at the required level of sensitivity [78].

2.5.3 MACHOs

Unless the current estimates of the baryon density, $\Omega_b = 0.0456 \pm 0.0015$ are incorrect, baryons cannot account for all of the dark matter. Nevertheless, baryons could account for *galactic* dark matter if they are in a form that neither absorbs nor emits light in significant quantities. The various possibilities have been reviewed in [79]. The simplest way to hide baryons is to place them in compact objects that either do not burn (e.g., brown dwarfs) or have ceased to burn (e.g., white dwarfs, neutron stars, black holes). Such dark objects in a galactic halo are called MACHOs for MAssive Compact Halo Objects.

Brown dwarfs have masses $< 0.07M_\odot$ making them too cool to burn hydrogen. They were originally the favored MACHO candidates because they completely avoid constraints based on the production of background light or pollution

⁷ Such a violation would produce a permanent nucleon electric dipole moment in violation of experimental limits [18].

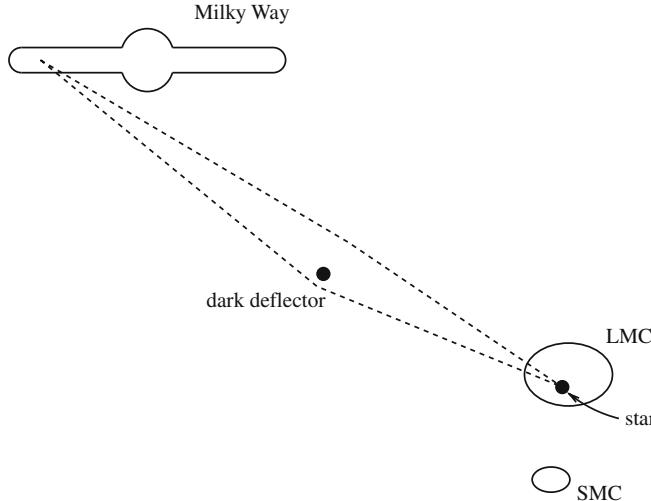


Fig. 2.22 A schematic of the lensing of a star in the Large Magellanic Cloud (LMC) by an unseen object in the galactic halo. While the two images cannot be easily resolved, the combined light from the two images gives a transient amplification of the light from the star as the unseen object passes near the line-of-sight. The light curve for a point source is shown in Fig. 2.23

of the interstellar medium with heavy elements through supernova explosions [80]. MACHOs in the form of black holes could also avoid the Ω_b constraint if they are “primordial” black holes that were produced before the epoch of nucleosynthesis.

Paczynski [81] suggested that MACHOs could be detected through their gravitational lensing of visible background stars in the Large Magellanic Cloud (LMC) (Fig. 2.22). This small galaxy is at a distance of 50 kpc from Earth.

The theory of gravitational lensing will be presented in Sect. 3.8. As the MACHO approaches the line-of-sight to the background star, two images are formed. It turns out that in the case of lensing by stellar objects in the galactic halo, the angle separating the two images is small (<1 milliarcsec). This type of gravitational lensing is therefore referred to as “microlensing.” Earth-bound telescopes cannot resolve the two images because atmospheric turbulence smears images so that stellar objects have angular sizes of order 1 arcsec. The only observable effect is therefore a transient increase of the total observed light as the MACHO moves toward and then away from the line-of-sight. The amplification is

$$A = \frac{u^2 + 2}{u\sqrt{u^2 + 4}}, \quad (2.21)$$

where u is the distance of closest approach of the (undeflected) line-of-sight to the deflector in units of the “Einstein radius” $R_E = \sqrt{4GMx(1-x)/c^2}$ where L is the observer–source distance, Lx is the observer–deflector distance, and M is the MACHO mass.

The amplification is greater than 1.34 when the distance to the line-of-sight is less than R_E . This amplification corresponds to a reasonable observational threshold since photometry can “easily” be done to better than 10% accuracy. At a given moment, the probability, P , of a given star being amplified by more than a factor 1.34 is just the probability that its undeflected light passes within one Einstein radius of a MACHO:

$$P \sim n_{\text{MACHO}} L \pi R_E^2 , \quad (2.22)$$

where n_{MACHO} is the mean number density of MACHOs between us and the LMC and L is the distance to the LMC. If the entire halo is comprised of MACHOs, the density of MACHOs is roughly $n_{\text{MACHO}} \sim M_{\text{halo}} / (ML^3)$ where M_{halo} is the total halo mass out to the position of the LMC. Using the expression for the Einstein radius, we find that P is independent of M and determined only by the velocity of the LMC:

$$P \sim \frac{GM_{\text{halo}}}{Lc^2} \sim \frac{v_{\text{LMC}}^2}{c^2} . \quad (2.23)$$

The LMC is believed to orbit the galaxy with $v_{\text{LMC}} \sim 200 \text{ km s}^{-1}$ (corresponding to a flat rotation curve out to the position of the LMC). In this case, P is of order 10^{-6} . More detailed calculations give $P = 0.5 \times 10^{-6}$ [82].

Since the observer, star, and deflector are in relative motion, a sizable amplification lasts only as long as the undeflected light beam remains within the Einstein radius. The light curve for a star near the center of the Milky Way lensed by a star in the Milky Way disk is shown in Fig. 2.23. The time scale of the amplification is the time t_E for the deflecting object to cross one Einstein radius with respect to the observer and source. For the lensing of stars in the LMC by objects in our halo, the relative speeds are of order 200 km s^{-1} and the position of the deflector is roughly midway between the observer and the source ($x \sim 0.5$). The mean t_E is then

$$t_E \sim \frac{R_E}{200 \text{ km s}^{-1}} \sim 75 \text{ days} \sqrt{\frac{M}{M_\odot}} . \quad (2.24)$$

The observed t_E distribution can therefore be used to estimate the mass of the MACHOs if one assumes that they are in the galactic halo.

Three groups, the MACHO, EROS and OGLE collaborations have published results of searches for events in the directions of the LMC and the SMC (the neighboring Small Magellanic Cloud). The limits are shown in Fig. 2.24. The lack of events with $t_E < 15$ days allowed the two groups to exclude as the dominant halo component objects with masses in the range $10^{-7} M_\odot < M < 10^{-1} M_\odot$ [84]. These limits exclude as a major halo component brown dwarfs of masses $\sim 0.07 M_\odot$. Furthermore, the EROS collaboration observed no events with $t_E < 400$ days and this excludes MACHOs with masses less than $\sim 30 M_\odot$. The MACHO collaboration has, however, observed 13 events of mean duration ~ 50 days [85]. If interpreted

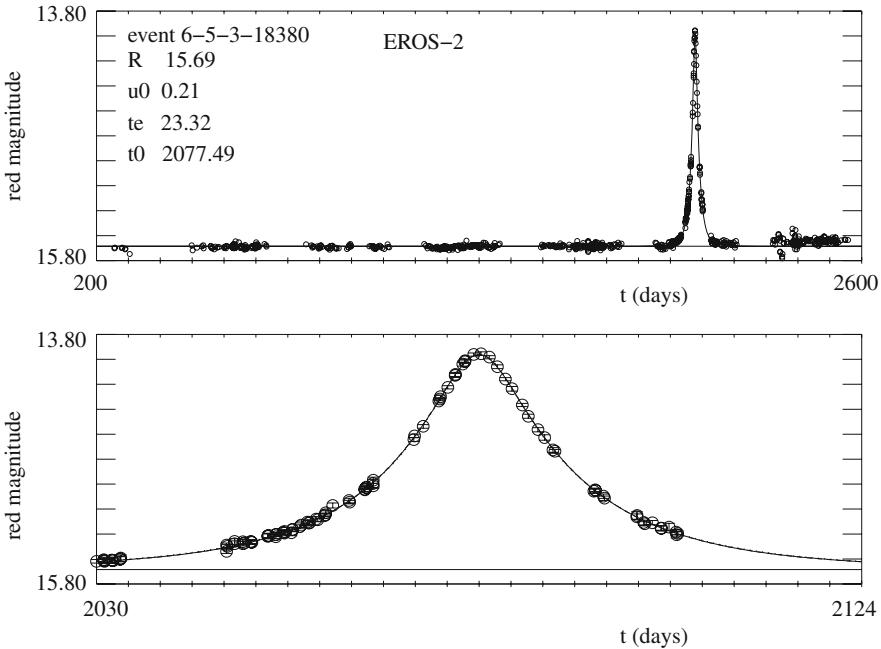


Fig. 2.23 A microlensing event observed by the EROS collaboration [83]. The lensed star is near the center of the Milky Way and the lens is a faint star in the disk of the Milky Way

as being due to dark lenses in the galactic halo, the rate corresponds to a fraction $f = 0.16$ of the total halo mass being comprised of MACHOs. The range of observed t_E correspond to halo objects of mass $\sim 0.4M_\odot$. However, the results of the EROS group appears to rule out this possibility, suggesting that the events observed by the MACHO collaboration are due to lensing by stars in the LMC itself. (The MACHO group monitored mostly very dense regions of the LMC where such so-called ‘self-lensing’ may have a high rate.)

Microlensing searches for dark objects are also being performed for the nearby spiral galaxy M31 [92]. Events have been observed but there is no consensus on the amount of MACHO dark matter that they represent.

The microlensing limits toward the Magellanic clouds appear to rule out MACHOs as the dominant dark matter for MACHO masses less than $\sim 30M_\odot$. MACHOs with higher masses are mostly ruled out as dark matter candidates because they would disrupt bound systems as they pass through the disk of the Milky Way. The various limits are shown in Fig. 2.24.

2.5.4 Cold Gas

A second way to hide baryons in galactic halos is to place them in small clouds of cold gas comprised of primordial helium and *molecular* hydrogen [12]. The

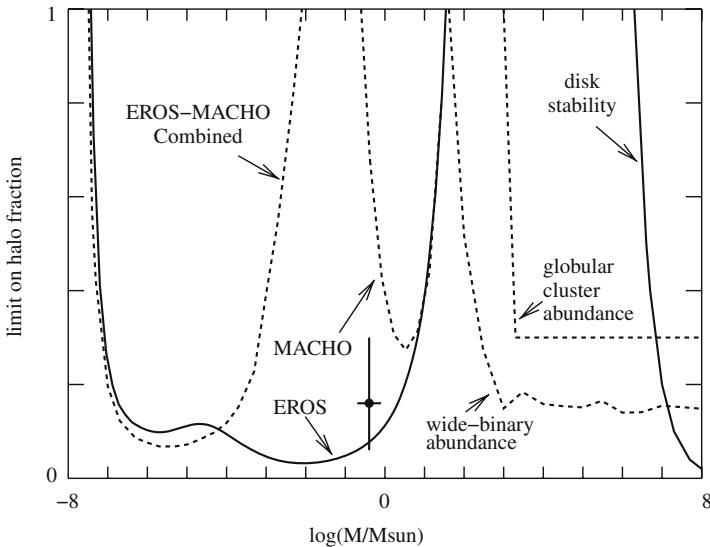


Fig. 2.24 Upper limits on the contribution of MACHOs to the halo of the Milky way as a function of MACHO mass. The curves labeled EROS [86], EROS-MACHO [84], and MACHO [87] are limits deduced from the lack of microlensing events toward the LMC. The curves labeled “wide-binary abundance” (controversial) [88], “globular cluster abundance” [89], and “disk stability” [90] are deduced from the lack of disruption of bound structures by passing MACHOs. The cross at a halo fraction of ~ 0.16 for a MACHO mass of $\sim 0.4M_{\odot}$ corresponds to the microlensing events seen by the MACHO collaboration [85] as corrected for variable star contamination [91]. The strict EROS limit at this mass suggests that the MACHO collaboration events are not due to lensing by MACHOs in the Milky way halo but, rather, by faint stars in the LMC itself

hydrogen must be molecular in order to escape detection via 21 cm emission by atomic hydrogen. The gas must be in clouds because a spatially uniform gas would lead to unobserved absorption of extragalactic sources at molecular transitions [93]. Clouds of a sufficiently high density would be sufficiently rare that most lines-of-sight would have no such absorption.

While this proposal is very efficient in hiding the gas, the plausibility of producing such quantities of molecules is controversial. In galactic disks, molecules are believed to be produced primarily on the surfaces of dust grains and this would not be possible in a primordial mixture of gas.

Limits of the quantity of cold molecular gas clouds near the Milky Way disk have been obtained from limits on the flux of high-energy photons that would be produced by cosmic-ray interactions in the clouds [94]. Limits on the amount in the halo are more difficult to obtain. Under certain conditions, molecular clouds should be observable in microlensing surveys, either in our Galaxy by using the Magellanic Clouds [95] or in galaxy clusters by using background quasars [96].

2.6 The Cosmological Parameters

2.6.1 H_0

The current universal expansion rate, H_0 , plays several roles in cosmology. Its inverse, the Hubble time t_H , is the time scale of the expansion giving the order of magnitude of the time elapsed since the beginning of the present epoch of classical expansion after the end of the inflationary epoch:

$$t_0 = H_0^{-1} f(\Omega_M, \Omega_\Lambda) , \quad (2.25)$$

where the function $f(\Omega_M, \Omega_\Lambda)$ takes into account acceleration or deceleration and will be calculated in Chap. 5. The Hubble distance, $d_H = c H_0^{-1}$, gives the scale of the present classical horizon and the relation between the distances and the redshifts of nearby galaxies

$$R = c H_0^{-1} z \quad z \ll 1 . \quad (2.26)$$

H_0 determines the critical density

$$\rho_c = \frac{3H_0^2}{8\pi G} . \quad (2.27)$$

In a universe with $\Omega_\Lambda = 0$, the critical density determines the dividing line between universes that will continue to expand eternally and universes that will eventually contract.

Finally, as we will see in the next section, H_0 enters into measurements of universal densities in various ways, and will therefore be needed to compare densities. For example, the density of photons is directly measured giving $\rho_\gamma \propto \Omega_\gamma H_0^2$. The theory of primordial nucleosynthesis allows us to determine the baryon-to-photon ratio, $\eta \equiv n_b/n_\gamma$, yielding an estimate of $\rho_b \sim m_p \eta \rho_\gamma \propto \Omega_b H_0^2$. On the other hand, observations of BAO and high-redshift supernovae give directly an estimate of Ω_M and Ω_Λ . Conclusions about the relative quantities of photons, baryons, dark matter, and dark energy therefore depend on H_0 . This fact makes the expected anisotropies of the CMB depend on the value of H_0 . The precise value given in Table 1.1, $H_0 = 70.5 \pm 1.3 \text{ km s}^{-1}\text{Mpc}^{-1}$ is based on CMB studies discussed in Chap. 7

While less precise than the estimates based on CMB anisotropies, the value of H_0 measured locally from Hubble's law remains an essential consistency test of cosmology. Hubble's law is

$$v = H_0 R , \quad (2.28)$$

where v is the recession velocity of a galaxy (easily determined from the redshift of the galaxy's spectral lines) and R is the distance of the galaxy. The law applies in the velocity range $10^{-2}c < v < 10^{-1}c$, the lower limit necessary to neglect random peculiar velocities and the upper limit to neglect relativistic corrections that depend on q_0 . Galaxies in this redshift range are called "Hubble flow" galaxies.

The Hubble Key project [9] on the Hubble Space Telescope (HST) provided a breakthrough in the measurement of H_0 giving a value with a precision of $\sim 10\%$:

$$H_0 = (72 \pm 8) \text{ km sec}^{-1} \text{ Mpc}^{-1}. \quad (2.29)$$

Reported values of H_0 over the previous decade had spanned a range over nearly a factor 2 from $H_0 \sim 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ to $H_0 \sim 90 \text{ km s}^{-1} \text{ Mpc}^{-1}$. The Hubble Key project "Hubble diagram" on log-log scale is shown in Fig. 2.25

The primary difficulty in determining H_0 comes from the difficulty in determining the distances to galaxies. Very loosely speaking the various methods can be separated into "astrophysical methods" that are calibrated with the "distance ladder" and "physical methods" giving directly galactic distances.

Most of the points in Fig. 2.25 use the Hubble Key Project distance ladder summarized in Table 2.3. There are four steps on this ladder: the Milky Way, the large Magellanic Cloud, other nearby galaxies ($100 \text{ kpc} < R < 30 \text{ Mpc}$), and Hubble flow galaxies ($60 \text{ Mpc} < R < 400 \text{ Mpc}$). Only objects on the last step can be used to measure H_0 . Generally speaking, the distance, R , to an object on a given step is deduced from its observed light flux, $f = L/4\pi R^2$, where L is the known

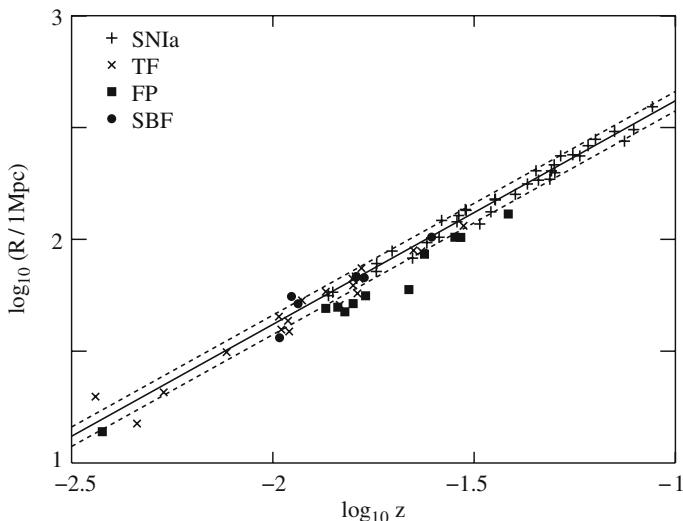


Fig. 2.25 The Hubble diagram of the Hubble Key Project [9]. The plot shows distances to galaxies vs. their redshift with distances determined by the techniques as marked

Table 2.3 The distance ladder used by the Hubble Key Project [9]. The four steps are (1) nearby Milky Way stars measured by parallax; (2) the Large Magellanic Cloud measured mostly by standard-candle stars calibrated locally by parallax (Fig. 2.3); (3) nearby galaxies measured by Cepheids variable stars calibrated in the Large Magellanic Cloud (Fig. 2.5); and (4) “secondary” distance indicators calibrated in nearby galaxies

Objects	Method
Milky Way stars ($R < 200$ pc)	Parallax (Hipparcos)
Large Magellanic Cloud (~ 1000 Cepheids)	Cepheid variables..... $\mu = 18.57 \pm 0.14$ Eclipsing variables..... 18.33 ± 0.05 SN1987a rings..... 18.47 ± 0.08 Red Giants..... 18.64 ± 0.05 Red Clump..... 18.27 ± 0.11 RR Lyrae variables..... 18.30 ± 0.13 Mira variables..... 18.54 ± 0.04 average..... $\mu = 18.5 \pm 0.1$ $R_{\text{LMC}} = 10\text{pc } 10^{\mu/5} = (50.1 \pm 2.5)\text{kpc}$
31 galaxies ($0.1\text{Mpc} < R < 30\text{ Mpc}$) (6 SNIa hosts)	Cepheid variables
Galaxies ($60\text{Mpc} < R < 400\text{ Mpc}$)	SNIA..... $H_0 = 71 \pm 2 \pm 6$ Tully-Fisher..... $H_0 = 71 \pm 3 \pm 7$ Fundamental Plane.... $H_0 = 82 \pm 6 \pm 9$ Surface Fluctuations... $H_0 = 70 \pm 5 \pm 6$ SNII..... $H_0 = 72 \pm 9 \pm 7$ Average..... $H_0 = 72 \pm 8$

luminosity of the object. The luminosity is derived from similar objects at a lower step of *known distance*. The Hubble Key Project ladder proceeds as follows:

- 1: Distances to Milky Way objects are derived from parallax measurements of the Hipparcos satellite (Fig. 2.3) allowing one to deduce the luminosities of various classes of stars.
- 2: Observation of these classes of stars in the Large Magellanic Cloud (LMC) allows one to deduce the distance to the LMC. This LMC contains many bright Cepheid periodic variable stars that have a well-defined luminosity as a function of their period (Fig. 2.5) and the known distance to the LMC allows one to measure this period-luminosity relation.
- 3: Because of its excellent angular resolution, allowing it to resolve individual stars in nearby galaxies, the HST can detect Cepheid variable stars in galaxies within ~ 50 Mpc, thus determining their distance. The galaxies on the third step are then used to calibrate various “secondary” distance indicators, the most important being the luminosity of type Ia supernovae (Fig. 2.9) and

the luminosity of spiral galaxies as a function of their velocity dispersion (Tully-Fisher relation, Fig 2.11).

- 4: The known luminosities of supernovae or galaxies are then used to determine the distances to galaxies more distant than 50 Mpc, thus determining H_0 .

The uncertainty in the quoted value (2.29) includes the cumulative uncertainties in each step of the ladder, e.g., the 5% uncertainty in the distance to the Large Magellanic Cloud.

Three “physical” methods permit one, in principle, to bypass the distance ladder. The first is the “expanding photosphere method” (EPM) applied to type II supernovae (Sect. 2.1). After the initial explosion, the surface (photosphere) of the remnant expands at a velocity $v \sim 10^{-2}c$, as illustrated in Fig. 2.26. The velocity can be determined empirically from the Doppler shifts of the lines in the supernova spectrum. Knowledge of the moment of the explosion t_{exp} then allows one to calculate the physical diameter of the photosphere $D = 2v(t - t_{\text{exp}})$. If the angular size $\Delta\theta$ were known, the distance to the supernova could be determined directly via $R = 2v(t - t_{\text{exp}})/\Delta\theta$. For extragalactic supernovae, the angular size is too small to be directly measured but it can be estimated from the measured luminous flux by using Stefan’s law (2.1), appropriately modified for the non-blackness of the photosphere. This technique has given a value of $H_0 = 73 \pm 15 \text{ km s}^{-1}\text{Mpc}^{-1}$ [97]. The Hubble Key project calibrated this method in their distance ladder (Table 2.3) and deduced a value of $H_0 = 72 \pm 9 \pm 7$.

The second physical method uses the time delay between two images of quasars that are gravitationally lensed by foreground galaxy clusters. Several examples of gravitationally lensed quasars have been found where intrinsic variability of the quasar permits the measurement of the time delay. The light-curve (flux vs. time) for one is shown in Fig. 3.23. Since the optical paths are proportional to the distance scale and therefore to c/H_0 it is not surprising that the time delay between the two images is proportional to c/H_0 . The constant of proportionality will be calculated in Sect. 3.8. It depends on the angular separation of the two images and on the mass distribution of the lens. The mass distribution can be estimated from the distribution of gravitational arcs due to lensed background galaxies. Knowledge of the distribu-

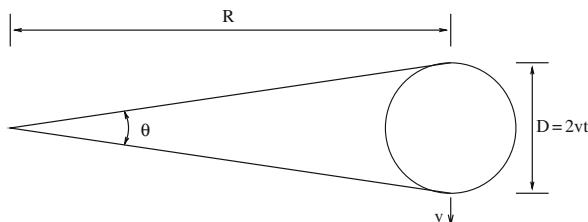


Fig. 2.26 The photosphere of a supernova expanding at a velocity v . The velocity can be deduced from the position and shape of the spectral lines. Knowledge of the time of explosion t_{exp} allows one to deduce the diameter $D = 2v(t - t_{\text{exp}})$. The angle θ can be estimated using Stefan’s law, allowing one to estimate the distance R

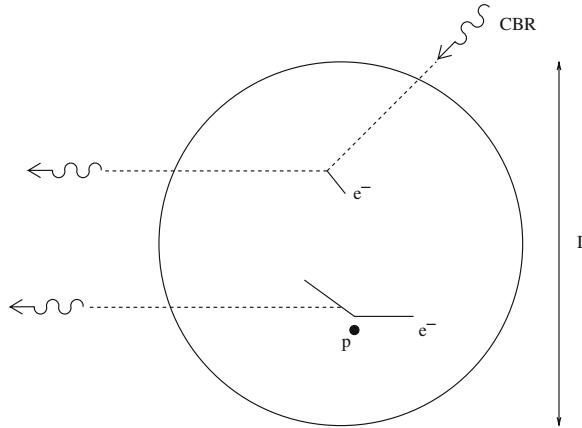


Fig. 2.27 A cluster of galaxies of diameter D . CMB photons can scatter on the free electrons in the hot ionized cluster gas increasing the mean energy of the CMB photons coming from the direction of the cluster (SZ effect). The ionized gas also emits X-rays via thermal bremsstrahlung of electrons scattering on protons and nuclei

tion limits the precision of this method. The Hubble Key Project [9] suggests that the measurements appear to be “converging” to about $H_0 \sim 65 \text{ km s}^{-1} \text{Mpc}^{-1}$.

The last physical method uses the Sunyaev–Zel’dovich (SZ) effect by which the CMB is heated as it Compton scatters on the hot ionized gas of a galaxy cluster (Fig. 2.27). The CMB spectrum is thus deformed in the direction of the cluster in proportion to the probability to scatter in the cluster:

$$P \sim \langle n_e \rangle \sigma_T D = \langle n_e \rangle \sigma_T R \Delta\theta , \quad (2.30)$$

where $\langle n_e \rangle$ is the mean electron density in the cluster, σ_T is the Thomson cross-section, and $D = R \Delta\theta$ is the diameter of the cluster at a distance R and subtending an angle $\Delta\theta$. For a cluster, $\Delta\theta$ can be directly observed so if P is measured from the SZ effect, we need only an estimation of $\langle n_e \rangle$ to determine R .

The electron density can be estimated from the cluster X-ray luminosity which is (Exercise 2.10)

$$L_X \sim c^2 \langle n_e^2 \rangle \alpha \sigma_T \sqrt{m_e T} D^3 , \quad (2.31)$$

where α is the fine structure constant. It follows that

$$\langle n_e^2 \rangle \sim \frac{4\pi f_x}{c\alpha\sigma_T\Delta\theta^3 R^3} . \quad (2.32)$$

Combining (2.30) and (2.32), we find the distance to the cluster:

$$R \sim \frac{P^2 c \alpha \Delta \theta \sqrt{m_e T}}{4\pi f_x \sigma_T} \frac{\langle n_e^2 \rangle}{\langle n_e \rangle^2}. \quad (2.33)$$

Everything on the right can be measured except for $\langle n_e^2 \rangle / \langle n_e \rangle^2$. This last factor is of order unity if the gas is relatively uniform. Any non-uniformity would result in an overestimation of R and, therefore, an underestimation of H_0 . It is also necessary to suppose that the cluster is spherical, though this hypothesis can be eliminated by averaging measurements of several clusters. The compilation of the Hubble Key Project [9] lists measured values in the range $40 \text{ km s}^{-1} \text{Mpc}^{-1} < H_0 < 80 \text{ km s}^{-1} \text{Mpc}^{-1}$. An increase in the number of studied clusters will enable astronomers to better estimate and control the uncertainties associated with this promising method.

2.6.2 ρ_s and Ω_s

The densities listed in Table 1.1 were derived from WMAP measurements of CMB anisotropies combined with measurements of type Ia supernovae and of baryon acoustic oscillations (BAO). In this section, we give a summary of these methods and a few others that have somewhat lower precision but serve as important consistency checks. The major methods are listed in Table 2.4.

The only directly measured cosmological density is the photon density with the COBE measurement [5] of the CMB temperature giving:

$$\rho_\gamma = (2.61 \pm 0.01) \times 10^5 \text{ eV m}^{-3} \Rightarrow \Omega_\gamma = 5.16 h_{70}^{-2} \times 10^{-5}. \quad (2.34)$$

Theoretical arguments based on the thermodynamics of the early universe allow one to deduce the densities of neutrinos and baryons:

- Neutrinos. The thermodynamic calculations of Chap. 6 will give $n_\nu = (3/11)n_\gamma$ for each species. The present neutrino density then depends on the neutrino masses with (1.36) for effectively massless neutrinos ($m_\nu \ll T_\gamma(t_0) = 2.349 \times 10^{-4} \text{ eV}$) and (1.37) for massive neutrinos. We note that the same type of calculation will give the number density of any species of weakly interacting massive particle if there is no particle–antiparticle asymmetry.
- Baryons. The theory of primordial nucleosynthesis predicts the abundances of the light elements as a function of the baryon–photon ratio, $\eta = n_b/n_\gamma$. The observed abundances [10] imply $\eta = (5 \pm 1) \times 10^{-10}$ which gives $\Omega_b = (0.04 \pm 0.01) h_{70}^{-2}$.

Knowing Ω_b one can deduce Ω_M by measuring the total mass to baryonic mass in objects that are believed to have a representative mix of baryons and CDM. This is expected to be nearly true for large clusters of galaxies. The baryons in these galaxies is mostly in the form of hot ionized gas and the mass in this form can be

Table 2.4 Some methods of determining the cosmological density parameters. Only ρ_γ is directly measured. Knowledge of H_0 allows one to deduce Ω_γ followed by Ω_V , Ω_b , and Ω_M . So-called “geometrical” methods using standard candles or rulers allow one to directly deduce Ω_M (BAO), $\Omega_M - \Omega_\Lambda$ (SNIa) and Ω_T (peak positions in CMB power spectrum). The details of the CMB power spectrum allows one to deduce Ω_Λ , Ω_M , and Ω_b if one knows H_0 . Finally, measurements of the abundance of galaxy clusters and weak gravitational lensing can be used to determine the product of Ω_M and σ_8 , the relative mass fluctuations of the scale of $8h^{-1}\text{Mpc}$

Quantity	Method	Reference
$\rho_\gamma \propto \Omega_\gamma H_0^2$	CMB density	
Ω_γ	CMB density + H_0	
Ω_V/Ω_γ	primordial thermal equilibrium + m_V	Sect. 6.4
Ω_b/Ω_γ	primordial nucleosynthesis	Sect. 6.5
Ω_b/Ω_M	galaxy cluster baryon fraction	Exercise 2.6
$\Omega_M H_0$	Large Scale Structure power spectrum	Sect. 7.2
Ω_M	Baryon Acoustic Oscillations (BAO) cluster number evolution	Sect. 5.3 Sect. 2.3
$\Omega_M - \Omega_\Lambda$	SNIa Hubble diagram	Sect. 5.2
Ω_T	CMB anisotropies + BAO	Sect. 7.4.2
$\Omega_\Lambda, \Omega_M, \Omega_b$	CMB anisotropies + H_0	Sect. 7.4.2
$\Omega_M \sigma_8$	galaxy cluster abundance large-scale weak lensing	Sect. 7.2

deduced from the X-ray luminosity. The fraction of the total mass of the largest galaxy clusters that is comprised of hot gas is measured to be $f_{\text{gas}} \sim 0.12 h_{70}^{-3/2}$ [98, 99]. This gives $\Omega_M \sim \Omega_b/0.12 \sim 0.37$. Corrections taking into account the amount of baryons in stars bring Ω_M down to $\Omega_M \sim 0.28 \pm 0.06$ [99].

The cluster number density measurements (Fig. 2.15) can give Ω_M if the density can be measured as a function of time or equivalently of redshift. The evolution of the number of clusters depends on Ω_M because structure formation slows at $\hat{a} = \Omega_M/(1 - \Omega_M)$ for $\Omega_\Lambda = 0$ and at $\hat{a} = (\Omega_M/(1 - \Omega_M))^{1/3}$ for $\Omega_T = 1$. The small amount of evolution shown in Fig. 2.15 indicates [100] $\Omega_M = 0.34 \pm 0.08$.

At super-cluster scales, Ω_M is one of the parameters that determine the shape of the “power spectrum” of density fluctuations (Fig. 7.4). The large-scale structure of the universe can be characterized by Fourier decomposing the density contrast into modes of comoving wavelength $\lambda = 2\pi\hat{a}(t)/k$ where k is the fixed wavenumber. In structure formation theories with cold dark matter and primordial adiabatic density

fluctuations, short-wavelength modes have amplitudes that are suppressed because these modes oscillated as acoustic waves during the radiation epoch whereas the amplitude of long-wavelength modes grew during both radiation and matter epochs. The separation between short and long wavelengths corresponds to the Hubble distance at the moment of matter-radiation equality, $a(t_{\text{eq}}) = a_0 1.68 \Omega_\gamma / \Omega_M$ and to a present wavelength of $\lambda_{\text{eq}} \sim 600 h_{70}^{-1} \text{Mpc} \times 0.27 / (\Omega_M h_{70})$.

The most important methods for measuring Ω_M and Ω_A are “geometrical” methods that use type Ia supernovae as standard candles and the sound horizon at recombination as a standard ruler seen in the matter and CMB fluctuation spectra. These techniques are described in Chap. 5.

Finally, we note that Ω_M can be deduced from the abundance of galaxy clusters and from the gravitational lensing galaxies by large-scale structure if one uses information of the mass fluctuation spectrum (Chap. 7).

Exercises

2.1 The luminosity of a typical galaxy is $\sim 2h_{70}^{-2} \times 10^{10} L_\odot$ and the mean energy of stellar photons is $\sim 2 \text{ eV}$. What is the photon flux (in $\text{m}^{-2} \text{s}^{-1}$) of a galaxy of redshift z ($z \ll 1 \Rightarrow R \sim zd_H$)?

Compare the photon flux from the nearest large galaxies ($R \sim 1 \text{ Mpc}$) with the photon flux from the nearest stars ($R \sim 1 \text{ pc}$). (This calculation should explain why most objects visible to the naked eye are stars.)

2.2 The luminosity density of the universe is $\sim 1.2 h_{70} 10^8 L_\odot \text{Mpc}^{-3}$. Supposing that stellar light output has been relatively constant since the formation of the first stars about one Hubble time ago, estimate the number of photons ($E \sim 2 \text{ eV}$) that have been produced by stars. Compare the number of stellar photons with the number of CMB photons. (This problem will be treated more rigorously in Exercise 5.12.)

Stellar energy is mostly produced by the fusion of hydrogen to helium $4\text{p} \rightarrow^4 \text{He} + 2\text{e}^+ + 2\nu_e$. This transformation occurs through a series of reactions in stellar cores that liberate a total of $\sim 25 \text{ MeV}$. After thermalization, the energy emerges from stellar surfaces in the form of starlight. Estimate the number of protons (per Mpc^3) that have been transformed into helium over the last Hubble time. Compare this number with the number of protons available $n_b \sim \Omega_b \rho_c / m_p$.

2.3 Estimate the contribution to the universal photon mean free path of the following processes:

- Thomson scattering of photons on free electrons of number density $n_e \sim n_b$.
- Absorption by stars of number density $n_{\text{stars}} \sim \Omega_{\text{stars}} \rho_c / M_\odot$, $\Omega_{\text{stars}} \sim 0.0023$ and cross-section $\sim \pi R_\odot^2$.

- Absorption by dust in galaxies with $n_{\text{gal}} \sim 0.005 \text{ Mpc}^{-3}$ and cross-section $\sim \epsilon \pi R_{\text{gal}}^2$ where $R_{\text{gal}} \sim 10 \text{ kpc}$ and the fraction of visible light absorbed when passing through a galaxy is $\epsilon \sim 0.1$.

Compare these distances with d_H (\sim the distance of the most distant visible objects). Is the universe “transparent”? (Section 5.7 will treat this problem more rigorously.)

2.4 Supposing that we can only measure redshifts, angles, and photon fluxes, explain the factors of h_{70} in (2.9), (2.10), and (2.11) as well as the absence of such factors in (2.12).

2.5 By comparing the apparent magnitudes of LMC Cepheids (Fig. 2.5) to the apparent magnitudes of Cepheids in the galaxy NGC 1365 (Fig. 2.28), estimate the ratio of the NGC 1365 distance to the LMC distance. If the LMC distance is taken to be $50 \pm 5 \text{ kpc}$, what is the distance to NGC 1365.

NGC 1365 is a member of the Fornax galaxy cluster. The recession velocity of this cluster is 1441 km s^{-1} . Estimate H_0 .

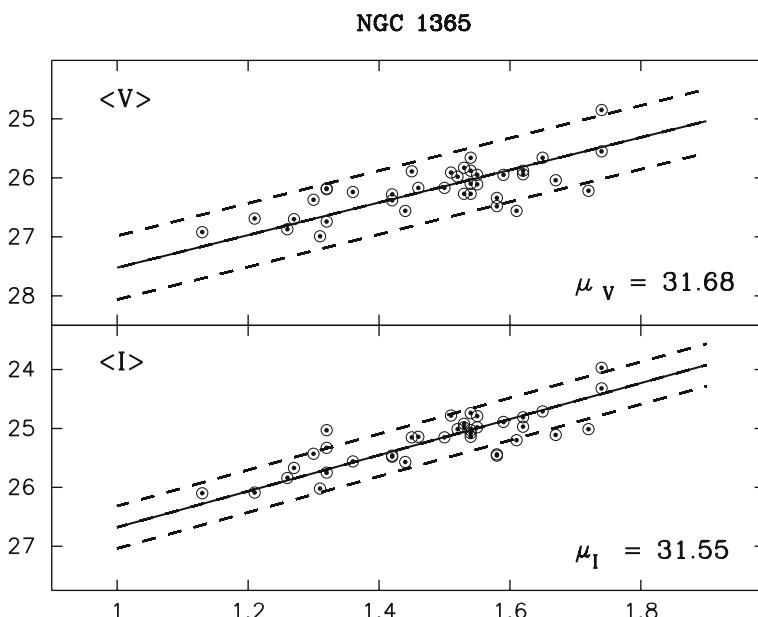


Fig. 2.28 The apparent magnitudes in the I and V bands of Cepheids in the galaxy NGC 1365 [53]. The magnitude is a linear function of $\log P$ where P is the Cepheid period in days

2.6 Abell-496 is a galaxy cluster whose properties were studied in [101].

- (a) The recession velocity of A496 is 9885 km s^{-1} . What is its distance as a function of h_{70} ?
- (b) Figure 2.29 shows the angular distribution of the brightest galaxies in A496. Estimate the radius of the cluster.
- (c) Figure 2.30 shows the distribution of recession velocities in the direction of A496. The accumulation near 9885 km s^{-1} corresponds to the cluster members. The width of this accumulation suggests that the line-of-sight velocity dispersion of A496 is $\Delta v \sim 715 \text{ km s}^{-1}$. Use the virial theorem to estimate the cluster mass: ($GM/\Delta r \sim \Delta v^2$). A detailed study in [101] gives $M_{\text{vir}} = (5.1 \pm 0.8)h_{70}^{-1}10^{14}M_{\odot}$.
- (d) The flux of visible light from A496 indicates a total luminosity of $L = 2.0 h_{70}^{-2} \times 10^{12} L_{\odot}$. By assuming that M_{vir}/L is equal to the universal value ρ_M/J_0 with the universal luminosity density given by (2.10), estimate ρ_M and Ω_M .
- (e) The X-ray spectrum indicates that the temperature of the intergalactic gas in A496 is $4 \pm 1 \text{ keV}$. The X-ray flux allows one (Exercise 2.10) to estimate the total mass and the total mass of intergalactic gas. For A496, the total mass within $0.7 h_{70}^{-1} \text{ Mpc}$ of the cluster center is $(1.7 \pm 0.4)h_{70}^{-1} \times 10^{14}M_{\odot}$ and the total gas mass within the same radius is $(1.4 \pm 0.5)h_{70}^{-5/2} \times 10^{13}M_{\odot}$. Assuming that the ratio between baryonic mass and total mass of A496 is equal to the universal value, estimate Ω_b/Ω_M .

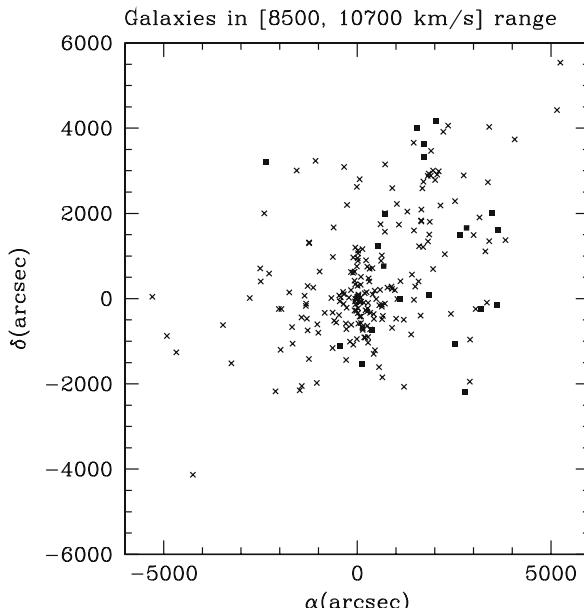


Fig. 2.29 The angular distribution of bright galaxies in A496 [101]

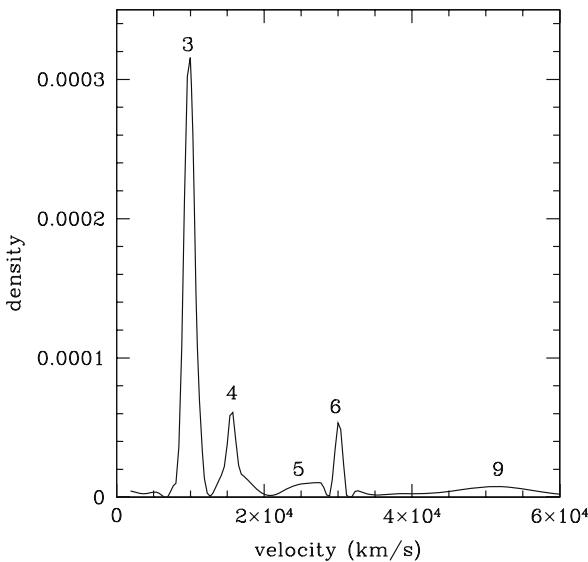


Fig. 2.30 The distribution of recession velocities in the direction of the A496 [101]. The accumulation of 274 galaxies around 9885 km s^{-1} corresponds to the cluster members

2.7 Figure 2.31 shows the iso-recession velocity curves of the galaxy NGC 5033 deduced from the Doppler shift of the 21 cm line of atomic hydrogen. The curves are superimposed on an optical image of the galaxy.

- (a) What is the redshift of NGC 5033. By neglecting its peculiar velocity, estimate its distance as a function of h_{70} .
- (b) The visible angular radius of NGC 5033 is about 3 arcmin. What is the visible radius as a function of h_{70} ?
- (c) What is the rotation velocity far from the galactic center? Take into account the galaxy inclination by supposing that the galaxy would appear to be circular if viewed face-on.
- (d) Estimate the mass of NGC 5033 that is within 6 arcmin of the galactic center (in units of M_\odot and as a function of h_{70}).
- (e) NGC 5033 has an apparent magnitude in the V band of 10.1. What is its absolute magnitude and its luminosity (in units of $L_{\odot V}$) as a function of h_{70} ? What is its mass-to-light ratio?

2.8 The most reliable distance indicators out to $\sim 50 \text{ Mpc}$ are Cepheid variable stars. The most reliable method of calibrating the luminosity-period relation of Cepheids is to use the large number of Cepheids observed in the Large Magellanic Cloud (LMC). This method of calibration obviously requires a knowledge of the LMC distance.

One of the most direct measurements of the LMC distance uses “eclipsing binaries.” Such systems consist of two orbiting stars whose orbital plane is oriented such

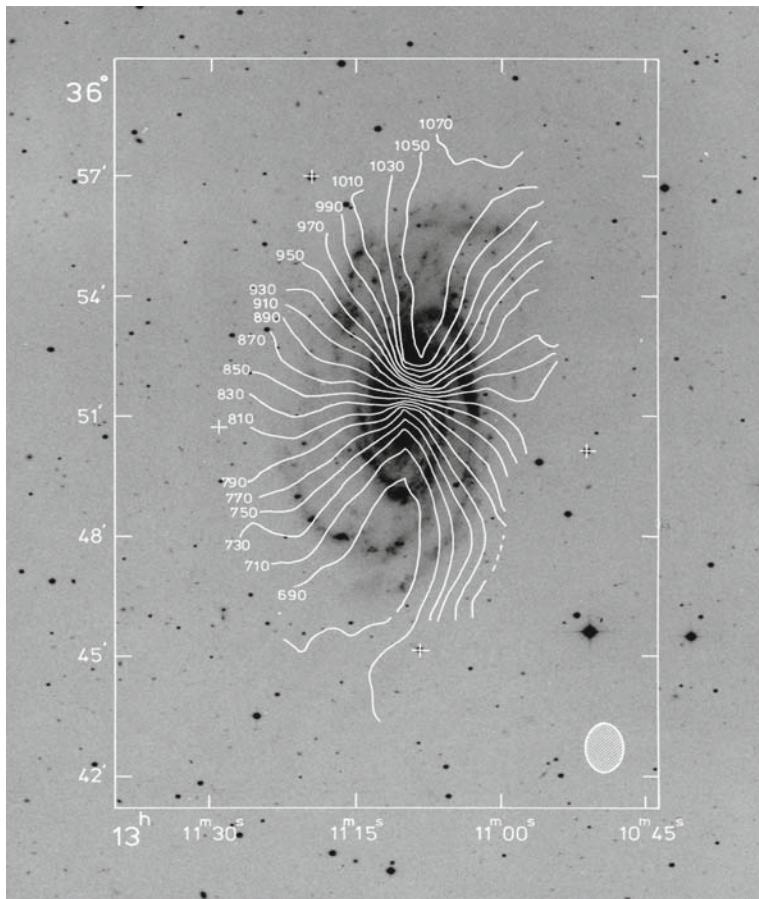


Fig. 2.31 The iso-recession-velocity curves of the galaxy NGC 5033 deduced from the Doppler shift of the 21 cm line of atomic hydrogen [102]. The curves are superimposed on an optical image of the galaxy. The angular scale of the greater dimension is in arcmin. Courtesy of A. Bosma

that, viewed from Earth, the two stars periodically eclipse each other. For eclipsing binaries at the distance of the LMC, the two stars generally have an angular separation that is so small that the two stars cannot be optically resolved. Rather, they appear as a single star with a periodic luminosity due to the periodic eclipses.

Figure 2.32 shows the “light curve” (apparent magnitude versus time) of the binary system HV2274 in the LMC [103]. Two eclipses are present with a period of 5.726 days. The magnitude change of 0.75 during the eclipses corresponds to a factor of two in flux indicating a total eclipse of two stars of equal luminosities and radii.

The spectral lines of the two stars do not coincide because of the Doppler shift due to their orbital motion. It is therefore possible to determine independently the

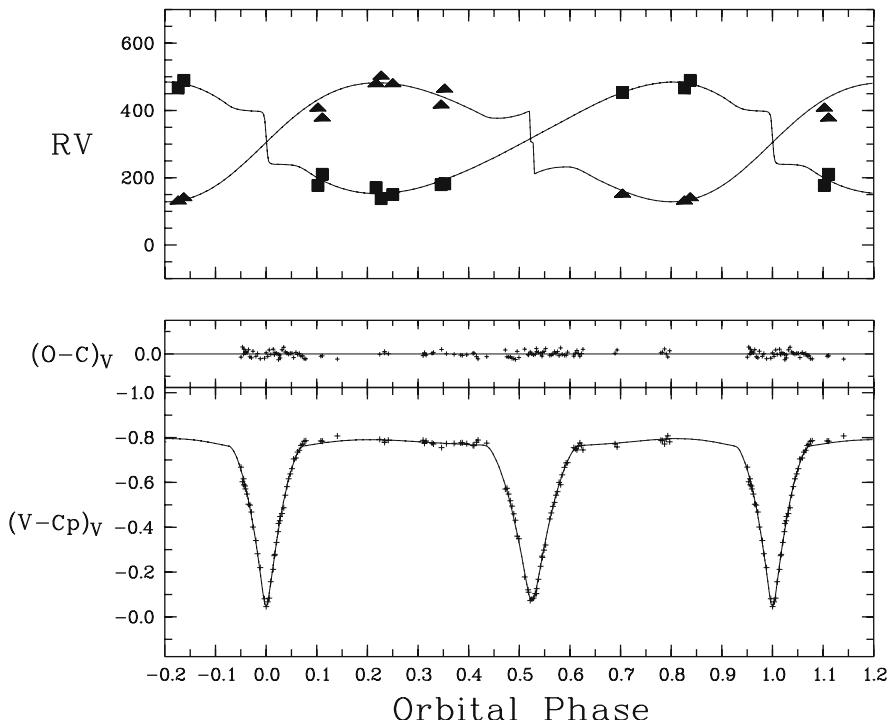


Fig. 2.32 The binary system HV2274 in the LMC [103]. The *upper panel* shows the recession velocity of the two stellar components as a function of the orbital phase (period = 5.726 days). The *lower panel* shows the light curve (apparent magnitude versus time)

recession velocities of the two stars. The two velocities as a function of time are also shown in Fig. 2.32.

- (a) Estimate the orbital velocity of the two stars and the radius of the (circular) orbit.
- (b) Supposing that the two stars have equal masses, estimate their mass (in units of M_{\odot}).
- (c) Use the durations of the eclipses to estimate the common radius of the two stars (in units of R_{\odot}).
- (d) The surface temperature of the two stars is $\sim 23\,000$ K. The measured flux indicates, via (2.1), an angular size of $D/R = 9.48 \times 10^{-12}$. Estimate the distance R to HV2274. After a small correction for the relative distance between HV2274 and the center of the LMC, the authors of [103] give an LMC distance of 45.77 ± 1.6 kpc.
- (e) Figure 2.5 shows the apparent magnitudes of LMC Cepheids as a function of their periods. Using the distance to the LMC, transform the apparent magnitudes into absolute magnitudes. Compare these magnitudes with those of Cepheids with distances determined by parallax [44], $M_V = -2.81 \log P - 1.43 \pm 0.16$

(period P in days). Compare the value of H_0 that would be estimated using Hipparcos Cepheids with that using LMC Cepheids.

2.9 It is perhaps surprising that the luminosity of a star can be estimated theoretically without knowing the nuclear reactions that power the star. To see how this can be done, we consider a sphere of radius R containing N_p protons and N_p electrons in the form of an ideal ionized gas.

- (a) if the sphere has a uniform density and is in hydrostatic equilibrium with a mean pressure P and volume V , show that

$$3PV = -E_g \sim (3/5) \frac{Gm_p^2 N_p^2}{R}, \quad (2.35)$$

where E_g is the total gravitational energy of the sphere and m_p is the proton mass.

The numerical factor (3/5) in (2.35) applies only to a sphere of uniform density. This is not the case for a star but a nonuniform distribution would simply give a different numerical factor. For the rest of this exercise we will ignore all numerical factors.

Applying the ideal gas law to (2.35), we can estimate the mean temperature T in the star:

$$T \sim \frac{Gm_p^2 N_p}{R}. \quad (2.36)$$

- (b) Supposing that the sphere contains photons in thermal equilibrium at the temperature T , show that the total number of photons inside the star is

$$N_\gamma \sim N_p^3 \left(\frac{m_p}{m_{pl}} \right)^6, \quad (2.37)$$

where $m_{pl} = \sqrt{\hbar c/G} = 1.2 \times 10^{19}$ GeV is the Planck mass. Compare N_γ with N_p for the sun ($N_p \sim 10^{57}$).

The photons diffuse in the star before escaping at the surface. The number of collisions in this random walk is of order

$$N_{col} \sim \left(\frac{R}{\lambda} \right)^2, \quad (2.38)$$

where λ is the mean free path of a photon in the star.

- (c) Show that the mean escape time for a photon is

$$\tau \sim \frac{N_p \sigma}{R c}, \quad (2.39)$$

where σ is the mean photon-particle cross-section in the star. From this, argue that the stellar luminosity is

$$L \sim N_p^3 \left(\frac{m_p}{m_{pl}} \right)^8 \frac{\hbar c^2}{\sigma}. \quad (2.40)$$

For a star like the Sun, the atoms are nearly all ionized except near the surface. It follows that $\sigma \sim \sigma_T$ (the Thomson cross-section). For $N = 10^{57}$, compare the luminosity from (2.40) with L_\odot .

A more careful management of the numerical factors multiplies the above result by $\pi^4/(5 \times 3^8) \sim 3 \times 10^{-3}$ [104]. This gives an agreement with the observed solar luminosity that is reasonable considering the approximations involved in the calculation.

We note that (2.40) predicts that a stellar luminosity is proportional to the third power of its mass, in good agreement with observations.

If the luminosity of a star depends only on its mass, one can ask what is the role of the nuclear reactions that power the star. The answer is that they allow the star to burn *longer* at a stable radius. A star begins its life as a diffuse cloud that is too cold to initiate nuclear reactions. It nevertheless radiates photons as required by (2.40). In so doing, it radiates energy and total energy conservation requires the star's radius to diminish. As the star becomes smaller, its temperature rises until nuclear reactions are ignited. At this point, a stable regime is reached where the energy radiated is replaced by the energy liberated by the nuclear reactions.

2.10 Roughly 10% of the mass of large galactic clusters is contained in ionized intergalactic gas in hydrostatic equilibrium. This gas produces photons via bremsstrahlung:



Theoretical and observed spectra are shown in Figs. 2.33 and 2.34.

Unlike the photons produced in stars, these photons escape directly from the cluster:

- (a) The largest clusters contain $\sim 10^{14} M_\odot$ of ionized hydrogen in a radius of ~ 1 Mpc. Verify that the mean free path of photons due to Thomson scattering in the cluster is greater than the cluster radius.
- (b) The large angle Rutherford scattering cross-section is $\sim \sigma_T/(v/c)^4$. Verify that the effective mean free path of electrons (scattering on protons) in the cluster is less than the cluster radius. This justifies the assumption that the electrons and protons form a thermal gas in hydrostatic equilibrium.
For non-relativistic electrons, the differential cross-section for bremsstrahlung production of photons of energy $E\gamma$ is approximately [106]

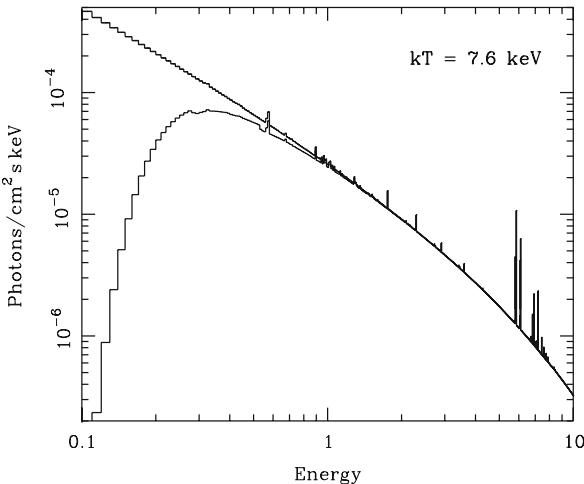


Fig. 2.33 The theoretical X-ray spectrum from a galactic cluster of temperature 7 keV. The spectrum follow the $1/E$ bremsstrahlung cross-section at low energy and then is exponentially cutoff at energies above the temperature. Recombination lines for iron are seen around 6 keV. The second curve that has a suppressed flux at low energy shows the effect of absorption in the Milky Way. Courtesy of Monique Arnaud

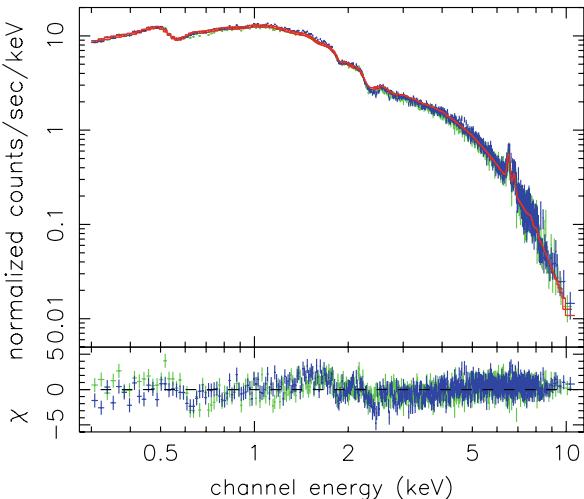


Fig. 2.34 The observed X-ray spectrum from the Coma galaxy cluster as observed by the XMM satellite [105]. The structures in the spectrum around 2 keV and 0.5 keV are due to the varying efficiency of the detection system. The fitted cluster temperature is $kT = 8.25$ keV. Courtesy of Monique Arnaud

$$\frac{d\sigma}{dE\gamma} \sim \alpha \frac{c^2}{v^2} \frac{\sigma_T}{E\gamma} E\gamma \ll (1/2)m_e v^2 , \quad (2.42)$$

where $\alpha \sim 1/137$ is the fine-structure constant, $v \ll c$ is the electron–proton relative velocity, and σ_T is the Thomson cross-section.

Using a line of reasoning that will be justified in Chap. 6, we know that the production rate per unit volume of photons is proportional to the differential cross-section (2.42), to the electron density n_e , to the proton density $n_p \sim n_e$, and to the mean electron–proton velocity:

$$\frac{dN_\gamma}{dt dV} \sim n_p^2 v \frac{d\sigma}{dE\gamma} \sim \frac{c}{v} n_p^2 \alpha c \frac{\sigma_T}{E\gamma} . \quad (2.43)$$

- (c) Integrate this expression up to a photon energy cutoff given by the temperature of the cluster T to find the total X-ray luminosity (energy/time):

$$L_x \sim n_p^2 \alpha c \sigma_T \sqrt{m_e c^2 T} D^3 , \quad (2.44)$$

where D is the diameter of the cluster.

- (d) Show that the total number of baryons, N_b , in the cluster can be estimated from the observed X-ray flux, f_x :

$$N_b^2 \sim \frac{f_x R^5 \theta^3}{\alpha c \sigma_T \sqrt{m_e c^2 T}} , \quad (2.45)$$

where θ is the observed angular diameter of the cluster and R is the distance to the cluster. This formula shows that if a cluster redshift is used to estimate the cluster’s distance, the measured total baryonic mass in the cluster scales as $h_{70}^{-5/2}$.

- (e) Modify (2.35) so that the thermal pressure supports only the baryonic mass of the cluster and thereby show that the total cluster mass in terms of the X-ray temperature is

$$M_{\text{tot}} \sim \frac{6kT R}{G m_p} \quad (2.46)$$

Chapter 3

Coordinates and Metrics

In order to study the movement of galaxies and photons, we will need to establish a coordinate system and its metric. Generally speaking, a coordinate system assigns to any event one temporal coordinate, x^0 , and three spatial coordinates, x^i , $i = 1, 2, 3$. The metric tensor $g_{\mu\nu}$ gives the square of the proper time¹ $d\tau$, i.e. the time actually measured by a clock as it changes its spatial coordinates by dx^1, dx^2, dx^3 during a coordinate time dx^0 :

$$d\tau^2 = \sum_{\mu,\nu=0}^3 g_{\mu\nu} dx^\mu dx^\nu \equiv g_{\mu\nu} dx^\mu dx^\nu , \quad (3.1)$$

where in the second form we adopt the usual convention that repeated indexes are summed over. Without loss of generality we can take $g_{\mu\nu} = g_{\nu\mu}$. The utility of the metric comes from the fact that, in the absence of non-gravitational forces, massive test particles follow trajectories that maximize their elapsed proper time and massless particles follow trajectories of vanishing proper time. Knowing the metric will therefore allow us to calculate the trajectories of galaxies and photons.

Throughout this book, we will use only “ruler free” relativity where distances are measured by photon flight times. A simple coordinate system can then be constructed by equipping an observer adrift (in free-fall) in intergalactic space with a clock and a radar emitter. As illustrated in Fig. 3.1, the observer can measure the time and distance of an event in the usual way. The photon is emitted at t_a (as measured by the clock), reflected at the event in question, and then detected by the observer at t_b (as measured by the clock). The event can then be assigned a time coordinate T and a radial coordinate R

$$T = \frac{t_b + t_a}{2} \quad R = \frac{t_b - t_a}{2} . \quad (3.2)$$

¹ The universal use of the term “proper time” is unfortunate. “A clock’s ‘proper’ time” appears to be a lazy translation of the French “*le temps ‘propre’ d’une horloge*” meaning “a clock’s own time,” which is precisely what it is.

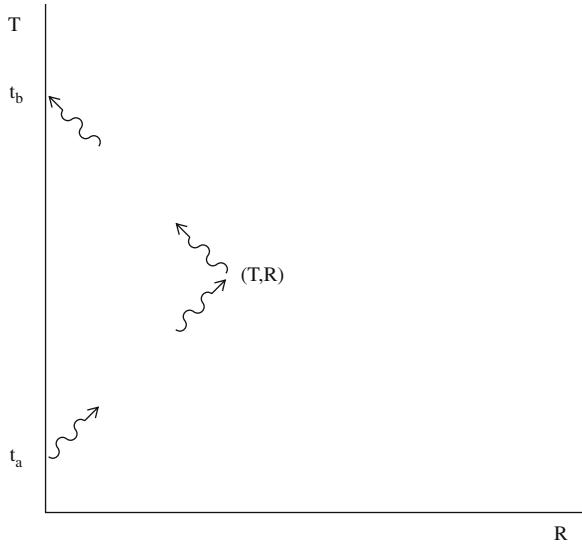


Fig. 3.1 The construction of a coordinate system by an observer equipped with a clock and a radar. The observer emits a photon at $t = t_a$ that is reflected at the event in question. The photon returns to the observer at $t = t_b$. The observer can assign to the reflection event a radial coordinate $R = (t_b - t_a)/2$ and the temporal coordinate $T = (t_b + t_a)/2$. It turns out that if the observer is in free-fall, the metric of the system (T, R) is locally Lorentzian at the origin. Unfortunately, this coordinate system cannot be used to cover the whole universe because of the existence of horizons

We will see that for events not too far from the observer, this procedure for defining coordinates gives the usual formula for proper time between events separated by (dT, dX, dY, dZ) :

$$d\tau^2 = dT^2 - dX^2 - dY^2 - dZ^2 , \quad (3.3)$$

where $X = R \sin\theta \cos\phi$, $Y = R \sin\theta \sin\phi$, $Z = R \cos\theta$ where θ and ϕ are the usual spherical angular coordinates of the event as seen by the observer. In metric language, this means

$$g_{00} = 1 \quad g_{11} = g_{22} = g_{33} = -1 \quad g_{\mu \neq \nu} = 0 . \quad (3.4)$$

While this coordinate system is very practical for nearby events, it cannot be used even in principle to describe the most distant observable cosmological events. The reason for this is that the photons that we now receive from very high redshift objects were emitted *before* those objects could have received any photons from us, i.e. we were outside the objects' horizon. Because of this horizon problem, we will need, at the very least, a network of clocks distributed around the universe.

It turns out that the most practical coordinate system for cosmology is a “co-moving” coordinate system. Such systems measure the temporal coordinate with “co-moving” clocks, i.e. clocks that participate in the universal expansion. An

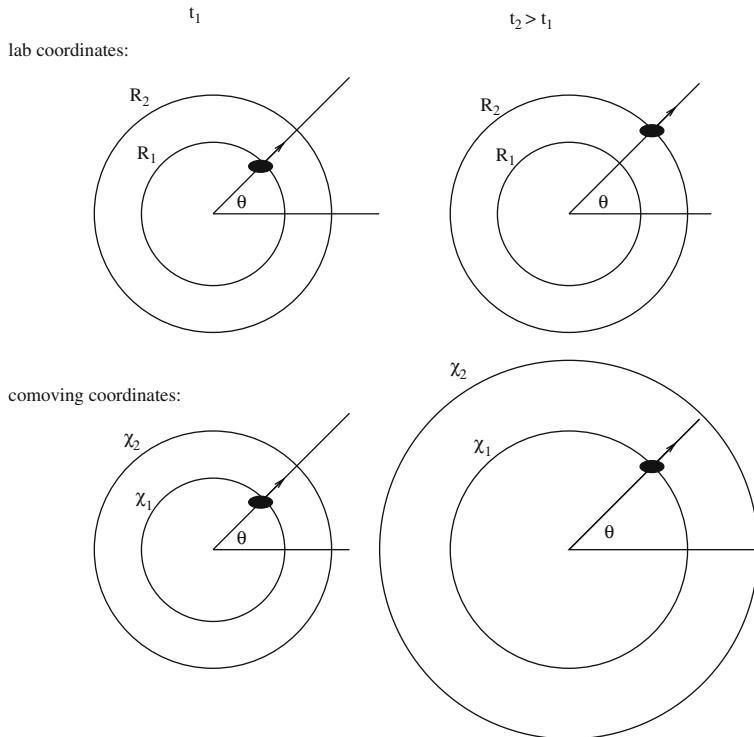


Fig. 3.2 The movement of a galaxy of negligible peculiar velocity in laboratory coordinates and in co-moving coordinates. In laboratory coordinates, the radial coordinate R evolves with time according to Hubble's Law. In co-moving coordinates, the radial coordinate χ is time independent. In the two systems, the angular coordinates of the galaxy are time independent

event that occurs in a galaxy is therefore given a temporal coordinate measured by a clock in that galaxy. Furthermore, as illustrated in Fig. 3.2, in co-moving systems, galaxies are given spatial coordinates that are time-independent. Co-moving coordinates are thus quite different from traditional “laboratory” coordinates in which the receding galaxies would have time-dependent spatial coordinates and the time coordinate would be measured by clocks fixed with respect to the origin.

We start this chapter in Sect. 3.1 by discussing in more detail how gravitation is described relativistically. Emphasis will be placed on the fact that in both special and general relativity, free massive particles follow trajectories that maximize their proper time. Such trajectories are called space-time “geodesics.” The primary difference between special and general relativity is that in the latter the metric is a function of position in a way that describes the effects of gravitation.

In Sect. 3.2 we start the real business of this chapter by giving operational definitions for our co-moving coordinates. The metric for this system will then be derived in the following two sections. As we will see, certain characteristics of the metric will be the pure consequence of the definition of the coordinates while others will be

related to the hypotheses we make about the characteristics of space-time. The first hypothesis, used in Sect. 3.3, will be that the universe is isotropic about the origin, i.e. that there is no preferred direction. The second, used in Sect. 3.4, will be that the universe is homogeneous, i.e. that there is no preferred spatial origin. These two hypotheses will lead us to the Robertson–Walker metric.

After having established the metric, we will treat the problem of particle trajectories in co-moving coordinates. In Sect. 3.5, we will attack the most important class of trajectories, that of photons traveling between distant galaxies and our position at the origin. This treatment will allow us to establish, in Sect. 3.6, the cosmological relations between the luminosities and measured fluxes and between object sizes and measured angular sizes.

In Sect. 3.7, we will treat the general problem of particle movement by deriving the geodesic equation. The geodesic equation for the Robertson–Walker metric will be found, allowing us to treat the movement of test particles in a homogeneous universe.

Finally, in Sect. 3.8 we will find the geodesic equation for a spherically symmetric gravitational field. This will allow us to treat the cosmologically important problem of gravitational lensing by galaxy clusters.

3.1 Relativity and Gravitation

We introduce special and general relativity using technology that was unknown to Einstein. We focus our attention on the “proper time,” τ , measured by ideal clocks as they move through space. To avoid debilitating abstractions, it is best to consider digital clocks that “tick”:

$$\tau \propto \text{ticks of an ideal clock} . \quad (3.5)$$

Two types of sufficiently ideal clocks have been used in experiments on relativity. The first type are atomic clocks consisting of oscillators whose frequency, ν , is locked to the frequency, $\Delta E/2\pi\hbar$, of an atomic transition between two states separated by an energy ΔE . In this case the number of ticks is the number of oscillations, N_{osc} :

$$\tau \equiv N_{\text{osc}} \frac{2\pi\hbar}{\Delta E} \quad (3.6)$$

The fundamental period of the clock, $2\pi\hbar/\Delta E$, converts ticks to proper time. The second type consists of collections of a large number, N , of any species of unstable elementary particle or nuclei. In this case the number of ticks is the number of decays, N_{dec} :

$$\tau \equiv N_{\text{dec}} \frac{T}{N} \quad N_{\text{dec}} \ll N \quad (3.7)$$

where T is the particle mean lifetime. Such “decay” clocks have the disadvantage that, since decays occur at random moments, they do not tick at a uniform rate. Proper time measurements with these clocks must be understood as averages over intervals with many decays. Another disadvantage is that their fundamental period T/N increases with time as N decreases.

Each type of clock is characterized by a period that converts ticks to time, $2\pi\hbar/\Delta E$ for atomic clocks and T/N for decay clocks. Ideal clock can be intercalibrated by choosing their periods so that the clocks give the same proper time when they follow the same trajectory. In fact, the operational definition of a sufficiently ideal clock is that it agrees sufficiently well with other sufficiently ideal clocks. An example of a non-ideal clock is a pendulum. Such clocks stop ticking when in free-fall and would not work, for instance, in a GPS satellite.

A startling prediction of relativity theory is that two ideal clocks following different trajectories that start and end at the same points in space time will, in general, not measure the same proper times. This is the famous “twin paradox.” Figure 3.3 gives two examples of this paradox that have been verified experimentally.

The first paradox was demonstrated most precisely with a beam of unstable particles (muons) at CERN [107]. Clock A is stationary,

$$x_A(t) = R \quad y_A(t) = 0 , \quad (3.8)$$

while clock B (the muons) follows a circular trajectory:

$$x_B(t) = R \cos \omega t \quad y_B(t) = R \sin \omega t . \quad (3.9)$$

The clocks compare their number of ticks between their separation ($t = 0$) and their reunion ($t = 2\pi/\omega$). The well-known result of this experiment is that the number of ticks of clock B is less than the number of ticks of clock A by a factor $\sqrt{1 - v^2/c^2}$ where $v = R\omega$. This result means that the proper time $d\tau$ measured by either clock as it moves through (dx, dy) over (coordinate) time dt is given by the formula from special relativity

$$d\tau^2 = dt^2 - (dx^2 - dy^2)/c^2 . \quad (3.10)$$

For clock A, $dx = dy = 0$ so $d\tau = dt$ while for clock B, $dx^2 + dy^2 = v^2 dt^2$ so $d\tau = dt\sqrt{1 - v^2/c^2}$, yielding the observed result. This is the “standard” twin paradox: the moving clock has fewer ticks than the stationary clock.

Before going to the second paradox, we note that the coordinates (t, x, y) in (3.10) constitute a two-dimensional example of “Lorentzian coordinates” that apply to inertial reference frames. In such coordinate systems, the proper time is given by

$$d\tau^2 = dx^0 dx^0 - (dx^1 dx^1 + dx^2 dx^2 + dx^3 dx^3) , \quad (3.11)$$

where $x^0 = t$ and $(x^1, x^2, x^3) = (x, y, z)$. This equation is equivalent to

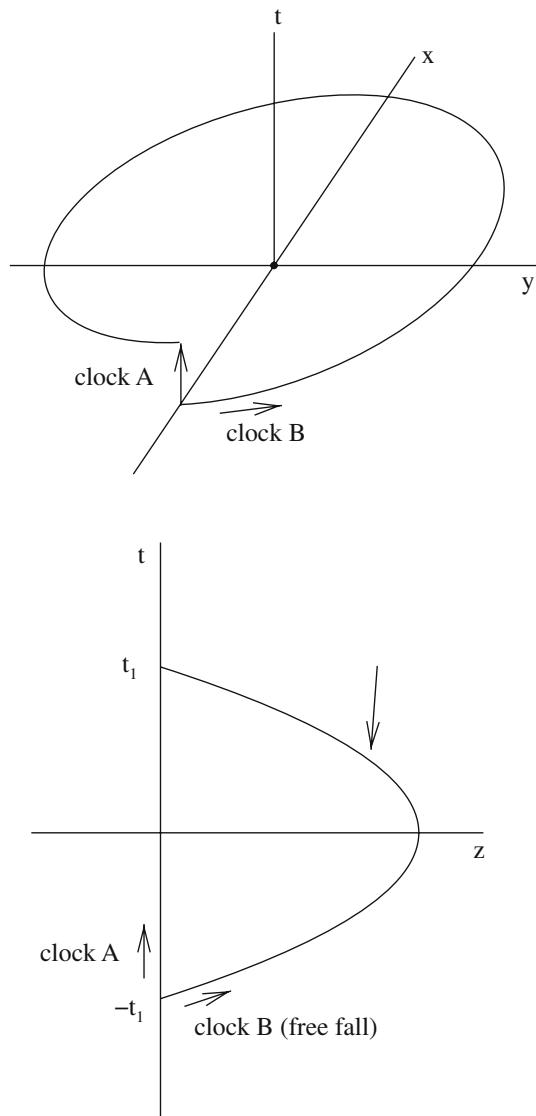


Fig. 3.3 Two twin paradoxes. The top paradox shows a stationary clock A and a moving clock B that follows a circular horizontal trajectory. The clock B can consist of an ensemble of charged unstable particles in a magnetic field. The bottom paradox shows a stationary clock A and a moving clock B that follows a parabolic free-fall trajectory in a nearly uniform gravitational field

$$d\tau^2 = \sum_{\mu,\nu=0}^3 \eta_{\mu\nu} dx^\mu dx^\nu \equiv \eta_{\mu\nu} dx^\mu dx^\nu , \quad (3.12)$$

where the metric $\eta_{\mu\nu}$ is

$$\eta_{00} = 1 \quad \eta_{11} = \eta_{22} = \eta_{33} = -1 \quad \eta_{\mu\neq\nu} = 0 . \quad (3.13)$$

A space-time where it is possible to find a system of Lorentzian coordinates is called “flat.” We note that a system of Lorentzian coordinates can be transformed into another system of Lorentzian coordinates via spatial rotations or by Lorentz boosts (Exercise 3.2).

An important advantage of using Lorentzian coordinates is that they are directly derived from clock and radar measurements, as in Fig. 3.1 (Exercise 3.13).

The second twin paradox of Fig. 3.3 was observed in the NASA experiment Gravity Probe A [108]. The experiment used two atomic clocks, one stationary at the launch pad and a second in a rocket that, after launch, followed a approximately parabolic trajectory $z(t)$:

$$z_A(t) = 0 \quad z_B(t) = g (t_1^2 - t^2) / 2 \quad (3.14)$$

where g is the gravitational acceleration. (In fact, the rocket went sufficiently high that g cannot be considered constant but we will ignore this fact.) The clocks compare their number of ticks between their separation ($t = -t_1$) and their reunion ($t = t_1$). Unlike the standard twin paradox, the number of ticks of clock B is *more* than the number of ticks of clock A, this time by a factor $1 + (g^2 t_1^2 / c^2) / 6$. A formula for proper time that predicts this result is

$$d\tau^2 = (1 + 2gz/c^2) dt^2 - dz^2 \quad (3.15)$$

For clock A, $dz = 0$ and $z = 0$ so $d\tau = dt$ while for clock B, $dz^2 = \dot{z}^2 dt^2$ so $d\tau = dt[1 + (gz(t) - \dot{z}^2/c^2)/2]$ for non-relativistic clocks: $\dot{z} \ll c$, $gz \ll c^2$. Integrating from $t = -t_1$ to $t = +t_1$ yields the factor $1 + (g^2 t_1^2 / c^2) / 6$. The gravitational factor $+gz/c^2$ in (3.15) over-compensates for the normal relativistic factor $-\dot{z}^2/c^2$.

Insight into the gravitational effect on proper time can be found by giving clock B a more general trajectory:

$$z_B(t) = \gamma (t_1^2 - t^2) / 2 \quad \gamma \neq g \quad (3.16)$$

This can be done by allowing the rocket to use its engine or by placing an electric charge on the rocket and using the Earth’s small electric field to generate the non-gravitational acceleration. Substituting the general trajectory into (3.15), we find that between $t = -t_1$ and $t = +t_1$ clock B measures a proper time

$$\tau_B = 2t_1 [1 + (2g - \gamma)\gamma (t_1^2/c^2) / 6] \quad (3.17)$$

The remarkable thing about this formula is that it is maximized for $\gamma = g$, i.e. for free-fall motion. This allows us to announce the fundamental property of trajectories in general relativity:

- In the absence of non-gravitational forces, a massive particle follows a path that maximizes its proper time compared to neighboring paths that have the same initial and final points. Such paths are called space-time “geodesics.”

In fact, this result is anticipated by the first twin paradox since the moving particles that must be accelerated by a (magnetic) force to make them reunite with the stationary clock.

A second general principle of general relativity can be illustrated by noting that the factor gz/c^2 in the proper time (3.15) can be eliminated by simply changing coordinates:

$$Z = z + gt^2/2 \quad T = t(1 + gz) \quad (3.18)$$

The spatial origin of the new system falls with gravity: $Z = 0 \Rightarrow z = -gt^2/2$. As such, (Z, T) are called “freely-falling” coordinates. For these coordinates

$$dT^2 - dZ^2 = (1 + 2gz/c^2)dt^2 - dz^2 + O(t^2, z^2, zt) \quad (3.19)$$

The first two terms on the right are just $d\tau^2$. Near the origin, $(z, t) = 0$, we have $(z, t) \sim (Z, T)$, so the proper time is

$$d\tau^2 = dT^2 - dZ^2 + O(T^2, Z^2, ZT) \quad (3.20)$$

Near $T = Z = 0$, the proper time takes nearly the special relativistic form with only quadratic corrections. One might ask if it is possible to eliminate the quadratic terms by further coordinate transformations. This turns out to be not mathematically possible. Physically, this is because while the gravitational force can be eliminated in a small region of space by adopting freely falling coordinates, widely separated particles in a non-uniform gravitational field will experience relative accelerations which are ascribed to “tidal” forces. For example, in the Earth’s gravitational field a particle will accelerate toward the center of the Earth. Two particles that are separated horizontally and then dropped will therefore be seen by an observer falling with them to accelerate toward each other (Fig. 3.4). We will see in Sect. 4.7 that these tidal effects can be taken into account by a metric that has second-order corrections to the Lorentz metric

$$g_{\mu\nu}(\tilde{x}) = \eta_{\mu\nu} + O(\tilde{x}^2) \quad (\text{freely falling coordinates}). \quad (3.21)$$

The coefficients of the quadratic terms in the metric will be constrained by the local density of energy and momentum via the Einstein field equation introduced in Chap. 4.

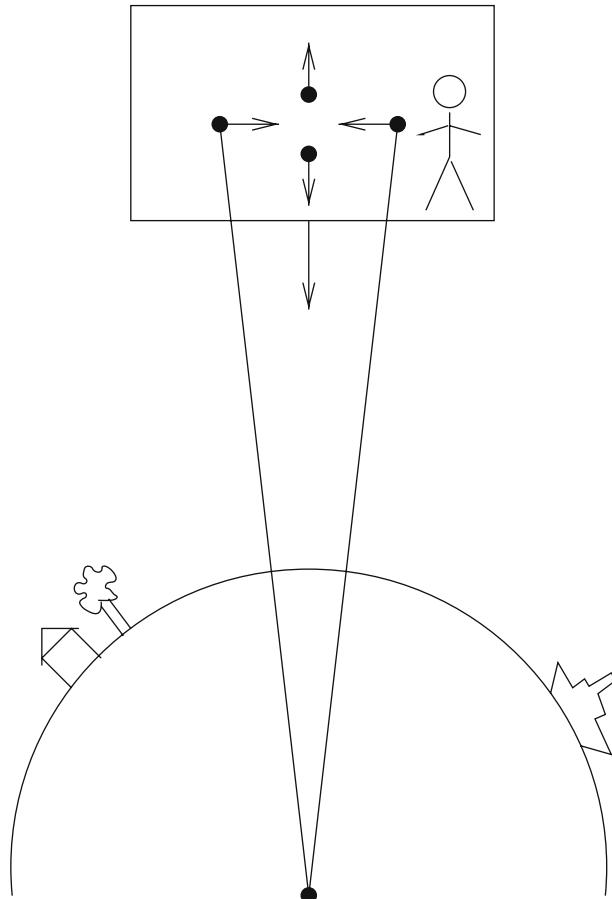


Fig. 3.4 Four particles separated by small distances in the Earth's gravitational field. All particles fall toward the center of the Earth by following paths that maximize their individual proper times compared to neighboring paths (Exercise 3.5). Because the paths are not perfectly parallel, a freely falling observer will see that the particles have small relative accelerations due to tidal effects, i.e. the non-uniformity of the gravitational field. The two particles at the same radius will appear to accelerate toward each other. Two particles at different radii will appear to accelerate away from each other. In general relativity, these relative accelerations are governed by the Riemann tensor (Sect. 4.7)

We note that the possibility of constructing freely falling coordinates is due to the equivalence principle equating gravitational and inertial masses. This principle guarantees that all particles are equally accelerated, so the acceleration can be transformed away by using freely falling coordinates. This is, of course, responsible for the “weightlessness” experienced by astronauts. A collection of particles in an infinitesimally small spatial region appear to follow the laws of special relativity. In fact, the freely falling frames are, locally, the best real examples of inertial reference frames. It follows that in an infinitesimal region, we can choose freely

falling coordinates, \tilde{x}^μ , $\mu = 0, 1, 2, 3$, centered on the region such that the metric is $g_{\mu\nu}(\tilde{x} = 0) = \eta_{\mu\nu}$. The particles in this region follow the geodesics of special relativity, i.e. straight lines.

3.2 Comoving Coordinates

In this section we will establish a comoving coordinate system that is useful for an expanding universe. For an event that occurs near a galaxy (or some imaginary observer receding according to Hubble's law), we will need a time coordinate t , two angular coordinates, (θ, ϕ) and one radial coordinate χ .

- For the spatial origin, we take (naturally) our own position.
- For the time coordinate, we will use the time measured by a clock attached to the receding galaxy or observer. Such a clock is called a “comoving” clock:

$$t = \text{time measured by comoving clock near the event.} \quad (3.22)$$

The clocks can be synchronized at the “big bang” or some other well-defined event (recombination, nucleosynthesis).

- For the angular coordinates, we take the standard polar and azimuthal angles of the event as seen on the sky with respect to some arbitrarily chosen direction

$$\theta, \phi = \text{angles seen on the sky.} \quad (3.23)$$

- The dimensionless radial coordinate, χ , is defined (Fig. 3.5) so that for two nearby points (separated in the radial direction) $d\chi$ is given by the photon flight time between the two points divided by the scale factor:

$$d\chi \equiv \frac{1}{a(t)} [\text{photon flight time over } (d\chi, d\theta = 0, d\phi = 0)]. \quad (3.24)$$

It is necessary to divide by $a(t)$ so that $d\chi$ between galaxies does not change with time.

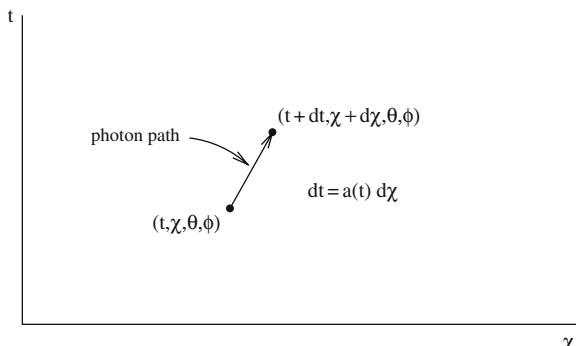
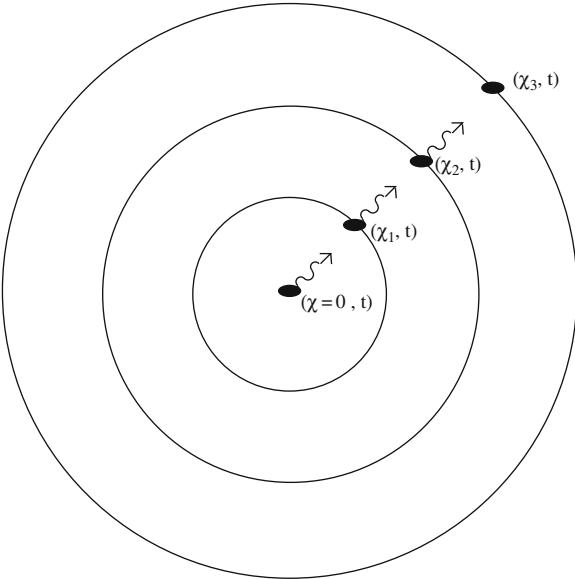


Fig. 3.5 The dimensionless radial coordinate χ is defined so that the photon flight time between two points separated in the radial direction by $d\chi$ is $dt = a(t)d\chi$

Fig. 3.6 The measurement of χ with a networks of synchronized comoving clocks and photon emitters. At the time t , the three photon emitters at $\chi = 0, \chi_1, \chi_2$ send a photon to their neighbor. The photons are detected at $(t + \Delta t_1, \chi_1)$, at $(t + \Delta t_2, \chi_2)$ and at $(t + \Delta t_3, \chi_3)$. The coordinate χ_3 is $\chi_3 = a(t)^{-1}(\Delta t_1 + \Delta t_2 + \Delta t_3)$



The coordinate separation $\Delta\chi$ between comoving observers who are not “nearby” is more difficult to determine because of the expansion of the universe. In order to measure coordinate separation sufficiently quickly so that $a(t)$ does not change significantly during the measurement, we use a network of synchronized comoving clocks and photon emitters as illustrated in Fig. 3.6. The figure shows how to measure the coordinate separation, at time t , between the origin and a distant galaxy. At the predetermined time t , each station i sends a photon to the next station and counts the time Δt_i before the arrival of the photon that was sent to him by the previous station. The coordinate separation is then just proportional to the sum of the flight times:

$$\chi = \frac{1}{a(t)} \sum_i \Delta t_i . \quad (3.25)$$

It should be clear that χ is a *gedanken* coordinate. Later, we will see how χ can be determined from redshift measurements.

Before calculating the metric for our coordinate system, we should emphasize a few points. First, the use of comoving coordinates is primarily a practical matter since, as we shall see, the motion of photons and galaxies is simple in such coordinates. This is partly because, in the limit of negligible peculiar velocities, galaxies have time-independent spatial coordinates (χ, θ, ϕ) . Of even more importance is the fact that the time coordinate t can be directly related to physical processes in galaxies. The use of clocks fixed with respect to the origin would require the use of, among other things, a factor of $1/\sqrt{1 - v^2}$ when relating the time coordinate to, e.g., reaction rates within galaxies.

A second reason for using comoving coordinates is that the homogeneity of the universe can be made manifest in comoving coordinates where all galaxies are treated equivalently (apart from the choice of spatial origin). This characteristic will be exploited in the next two sections when we determine the metric.

On the other hand, we should emphasize that comoving coordinates are not inertial coordinates (even locally) and certain results while mathematically simple to obtain will be non-intuitive. For local effects, like the propagation of photons between neighboring observers, our intuition is better served by using freely falling coordinates. For instance, the time dependence of the CMB temperature was derived in Sect. 1.2.4 in this way. As we shall see in Chap. 4, the Einstein equations for the evolution of $a(t)$ will be much simpler if we adopt freely falling coordinates.

Finally, we should emphasize that our definitions, (3.22), (3.23), and (3.24), of comoving coordinates are ambiguous if the matter distribution is inhomogeneous. Clumped matter can act as a gravitational lens and a given object can have more than one image on the sky. Our definition of the angular coordinates (θ, ϕ) is therefore ambiguous if there are two images of comparable brightness. The definition of χ is also ambiguous because more than one path can be used to determine the coordinate separation from the origin. The use of comoving clocks attached to galaxies is ambiguous since the potential wells of a galaxy can desynchronize clocks in different regions of the galaxy. We will ignore these effects because, in the present state of the universe, they are of secondary importance if we consider only large-scale phenomena.

3.3 The Metric I: Mostly Isotropy

To find the metric for comoving coordinates, we will first assume that the universe is isotropic about our position, i.e., that there is no preferred direction in our sky or orientation on our sky. We will see that this implies that the proper time is given by

$$d\tau^2 = dt^2 - a(t)^2 [d\chi^2 - r(\chi)^2(d\theta^2 + \sin^2 \theta d\phi^2)] \quad (3.26)$$

where $r(\chi)$ is an arbitrary function.

Most of the metric is a consequence of the definitions of the coordinates. The coordinate t is measured by clocks with constant spatial coordinates implying

$$d\tau = dt \quad (d\chi = d\theta = d\phi = 0). \quad (3.27)$$

This requires that $g_{tt} = 1$ as in (3.26).

The rest of the metric can be understood by considering photon trajectories from (t, χ, θ, ϕ) to $(t + dS, \chi + d\chi, \theta + d\theta, \phi + d\phi)$. Since photon trajectories have $d\tau = 0$, (3.26) implies

$$dS^2 = a(t)^2 [d\chi^2 - r(\chi)^2(d\theta^2 + \sin^2 \theta d\phi^2)]. \quad (3.28)$$

where dS can be interpreted as the spatial distance measured by photon flight time between points separated by $(d\chi, d\theta, d\phi)$.

Most of the (3.28) is required by the definitions of the coordinates and the scale parameter. The factor $a(t)$ follows from our assumption that intergalactic distances are proportional to a single space-independent function of time. The definition of the radial coordinate χ (3.24) requires $g_{\chi\chi} = a(t)^2$. The fact that the comoving clocks are synchronized implies $g_{t\chi} = g_{t\theta} = g_{t\phi} = 0$. Otherwise the photon flight time will depend on the flight direction.

The factor $(d\theta^2 + \sin^2 \theta d\phi^2)$ in (3.28) means that the distance between points on the sky (at the same χ) is proportional to their angular separation. Any other choice of $g_{\theta\theta}$, $g_{\phi\phi}$, and $g_{\theta\phi}$ would lead to photon flight times that depend on the orientation of the trajectory on the sky, violating the hypothesis of isotropy. The same would be true if $g_{\chi\theta}$ or $g_{\chi\phi}$ did not vanish.

Isotropy does not place any constraints on the function $r(\chi)$, and to do so we need to make further assumptions. If space is “flat,” the photon flight times will be proportional to the distance to the origin as in Euclidean geometry:

$$r(\chi) = \chi \quad (\text{flat space}) . \quad (3.29)$$

A space that is “locally flat” has only quadratic corrections to the Euclidean relationship:

$$r(\chi) = \chi \left[1 - \frac{b}{6} \chi^2 + \dots \right] \quad (3.30)$$

$$\Rightarrow \chi(r) = r \left[1 + \frac{b}{6} r^2 + \dots \right], \quad (3.31)$$

where b is a constant defined by this relationship. In a locally flat space, the Euclidean relation between radial and non-radial distances applies if the radial distance is sufficiently small, $|b\chi^2| \ll 1$, i.e. $a\chi \ll a(t)/\sqrt{|b|}$. In Chap. 4, we will see that $a(t)/\sqrt{|b|}$ is the order of or greater than the Hubble distance H_0^{-1} , which explains why space appears to be rather Euclidean.

It will be useful for certain purposes to use the function $r(\chi)$ itself as the dimensionless radial coordinate. Using $dr = d\chi/\chi'(r)$, we find that in the (t, r, θ, ϕ) coordinate system, the spatial metric takes the form

$$ds^2 = a^2 [f(r)dr^2 + r^2(d\theta^2 + \sin^2 \theta d\phi^2)] , \quad (3.32)$$

where

$$f(r) = \chi'(r)^2 = 1 + br^2 + \dots . \quad (3.33)$$

The nature of the two radial coordinates can be seen in the spatial metric. The coordinate χ is directly related to the distance R between the origin and the surface $\chi = \text{constant}$:

$$R(\chi, t) = a(t)\chi . \quad (3.34)$$

The coordinate r is more closely related to non-radial distances through (3.28) which implies that the angular separation $\Delta\Theta$ between two comoving objects at the same χ but separated by a distance ΔS is, at time t

$$\Delta\Theta = \frac{\Delta S}{a(t)r} . \quad (3.35)$$

The Euclidean relation $\Delta S = R\Delta\Theta$ is respected only if $\chi = r$:

$$\frac{\Delta S}{R\Delta\Theta} = \frac{r}{\chi} = 1 - \frac{b}{6}r^2 + \dots = 1 - \frac{R^2}{6a^2/b} + \dots . \quad (3.36)$$

The Euclidean relationship is a good approximation if R is sufficiently small; $R^2 \ll |a^2/b|$.

The non-radial distances determine the area A of the surface $\chi = \text{constant}$.

$$A(r, t) = \int [a(t)r d\theta] [a(t)r \sin\theta d\phi] = 4\pi r^2 a(t)^2 \quad (3.37)$$

$$= 4\pi a^2 \chi^2 \left[1 - \left(\frac{b}{3} \right) \chi^2 + \dots \right] . \quad (3.38)$$

Once again, the Euclidean relation $A = 4\pi R^2$ is respected only if $\chi = r$:

$$\frac{A}{4\pi R^2} = \frac{r^2}{\chi^2} = 1 - \frac{b}{3}r^2 + \dots = 1 - \frac{R^2}{3a^2/b} + \dots . \quad (3.39)$$

Equation (3.37) assumes that on the $r = \text{constant}$ surface, the lines of constant θ are perpendicular to the lines of constant ϕ . This is already implicit in the way we combined $d\theta$ and $d\phi$ in (3.28) to form dS . It implies that there are no distortions of the shape of an object, e.g. an appropriately oriented cubic star would appear on the sky as a square (Fig. 3.7). The lack of distortion is a characteristic of an isotropic universe since there is no preferred direction in which the images can be distorted.

Lacking direct geometrical measurements at large scale, we can only restrict the function $r(\chi)$ by making further hypotheses. We might be tempted to impose flatness, $r(\chi) = \chi$, but, as we will see in Chap. 4, this is equivalent to assuming that the universe has either a critical density or zero density. A less restrictive hypothesis is to suppose that the universe is homogeneous over the volume of interest, i.e. that all comoving observers see the same geometrical relationships, like (3.36) and (3.39). We will see that this hypothesis will restrict $r(\chi)$ to be one of three possible forms.

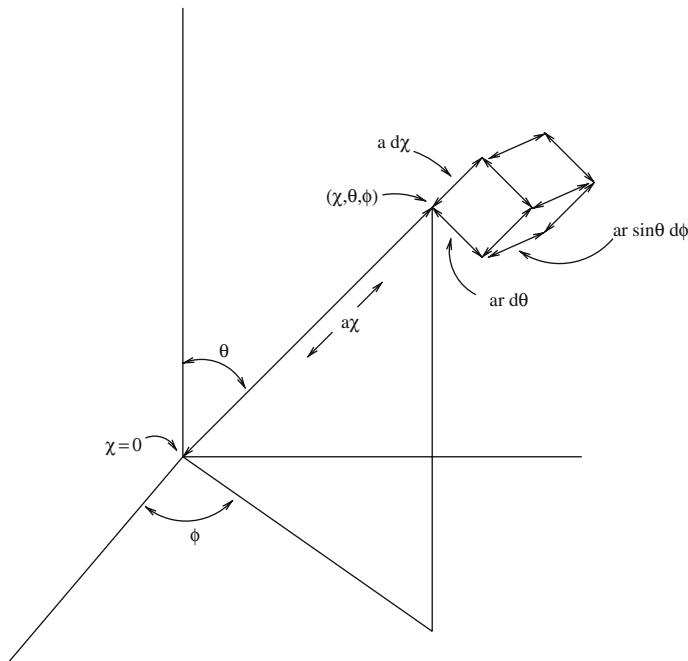


Fig. 3.7 A small cube of size dS at the position (χ, θ, ϕ) . If the universe is isotropic about the origin, the image of the cube seen on the sky from the origin will not be distorted because there is no preferred direction along which the image can be stretched. The cube in the figure will therefore appear as a square on the sky with angular size $d\theta = \sin\theta d\phi = dS/ar$. If, in addition to being isotropic, the universe is “Euclidean,” the angular size of the cube will be inversely proportional to its distance from the origin, i.e. $r(\chi) = \chi$. If the universe is homogeneous though not necessarily Euclidean, the function $r(\chi)$ is independent of the choice of origin

3.4 The Metric II: Mostly Homogeneity

The coordinates (χ, θ, ϕ) or (r, θ, ϕ) refer to the origin at $r = 0$. If the spatial metric (3.32) describes a homogeneous universe, we should be able to find coordinates centered on any point in which the metric takes the same form with the same function $f(r)$ or $r(\chi)$. If we can always find such coordinates, then we will know that geometrical relationships like (3.36) and (3.39) apply everywhere.

It is difficult to visualize curved spaces in three (spatial) dimensions. To develop some intuition, it is useful to consider surfaces in two dimensions. Points on such surfaces are described by one radial coordinate, χ , and one angular coordinate θ . Following the arguments of the previous section, a surface that is isotropic about the origin, $\chi = 0$, must have dS given by

$$dS^2 = a^2 [d\chi^2 + r(\chi)^2 d\theta^2] \quad (3.40)$$

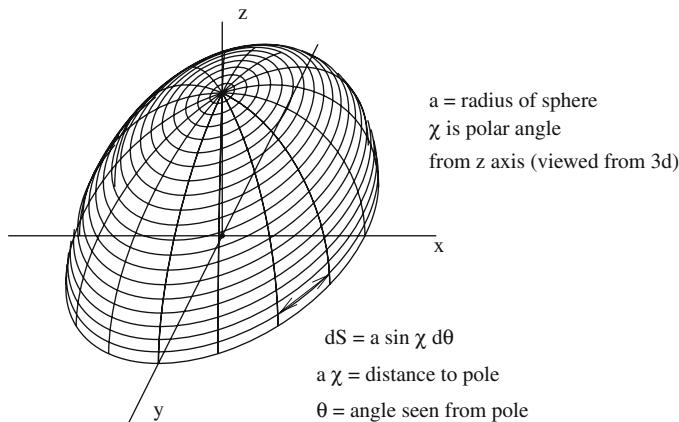


Fig. 3.8 The surface of a sphere of radius a . Points on the sphere are labeled by the coordinates (χ, θ)

The choice $r(\chi) = \chi$ corresponds to a flat surface. Such a surface is obviously homogeneous since choice of any point as the origin will lead to the same spatial metric. The choice $r(\chi) = \sin \chi$ corresponds to the surface of a sphere of radius a , illustrated in Fig. 3.8. Once again, this surface is obviously homogeneous since any point can serve as the origin.

A surface that is not homogeneous is shown in Fig. 3.9. It corresponds to $r(\chi) = \sin \chi - 0.5 \sin^2 \chi$. It is isotropic only about the point $\chi = 0$.

It can be shown in general that, for homogeneous two-dimensional surfaces or three-dimensional volumes, there are only three possibilities for $r(\chi)$: χ , $\sin \chi$, and $\sinh \chi$. To explicitly demonstrate this, we will construct a coordinate system (x, y, z) centered on the point $(\chi_0, \theta = 0)$ as in Fig. 3.10. We will find that it will be sufficient to consider a small region centered on this point.

If we use our intuition about flat space, we would expect that a good choice of coordinates would be

Fig. 3.9 An oblong surface.
 Points on the surface are
 labeled by the coordinates
 (χ, θ)

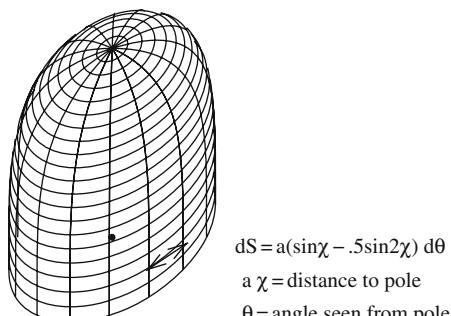
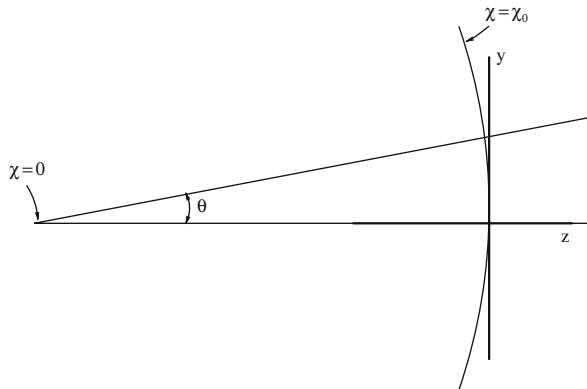


Fig. 3.10 Coordinates x, y, z centered on the point $(\chi_0, \theta = 0)$ (x is perpendicular to the paper). If the paper is flat,

$$z = \chi \cos \theta - \chi_0 \sim \chi - \chi_0 - \chi \theta^2/2$$


$$x = r \sin \theta \cos \phi, \quad y = r \sin \theta \sin \phi \quad (3.41)$$

$$z = \chi \cos \theta - \chi_0 = \chi - \chi_0 - \frac{\chi \theta^2}{2} \quad \theta \ll 1. \quad (3.42)$$

The θ^2 term corrects to first approximation for the fact that the lines of constant θ have an angle θ with the z axis in flat space. In curved space, the lines of constant θ have an angle $\alpha = \theta r'(\chi)$ with respect to the z axis (Fig. 3.11). Because of this, the following coordinate gives a simpler metric:

$$z = \chi - \chi_0 - \frac{r \theta^2}{2} r'(\chi) \quad \theta \ll 1, \quad (3.43)$$

which is the same as (3.42) if $r = \chi$ (flat space). [To understand why we chose this definition of z , we note that since the surfaces of constant χ are perpendicular to the lines of constant θ , the angle between the lines of constant χ and the y axis is also $\alpha = \theta r'(\chi)$. For $\theta = \text{constant}$ we therefore expect $dy = \alpha d\chi$ and for $\chi = \text{constant}$ we expect $dz = \alpha r d\theta$. These two conditions are satisfied by the transformations (3.43) and (3.41). In particular, we note that if $r'(\chi) = 0$, then $z = \chi - \chi_0$ as expected since the lines $\theta = \text{constant}$ are parallel to the z axis in this case.]

Using the transformations (3.41) and (3.43) the spatial metric (3.32) takes the following form near the point $x = y = z = 0$ (Exercise 3.11):

$$\begin{aligned} dS^2 = & a^2 [dz^2(1 - \beta x^2 - \beta y^2) \\ & + dy^2(1 - \alpha y^2) + dx^2(1 - \alpha x^2) - 2\alpha xy dx dy], \end{aligned} \quad (3.44)$$

where the coefficients α and β are functions of r_0 :

$$\alpha(r_0) = r_0^{-2} (f(r_0)^{-1} - 1) \quad (3.45)$$

$$\beta(r_0) = \frac{f'(r_0)}{2r_0 f^2(r_0)}. \quad (3.46)$$

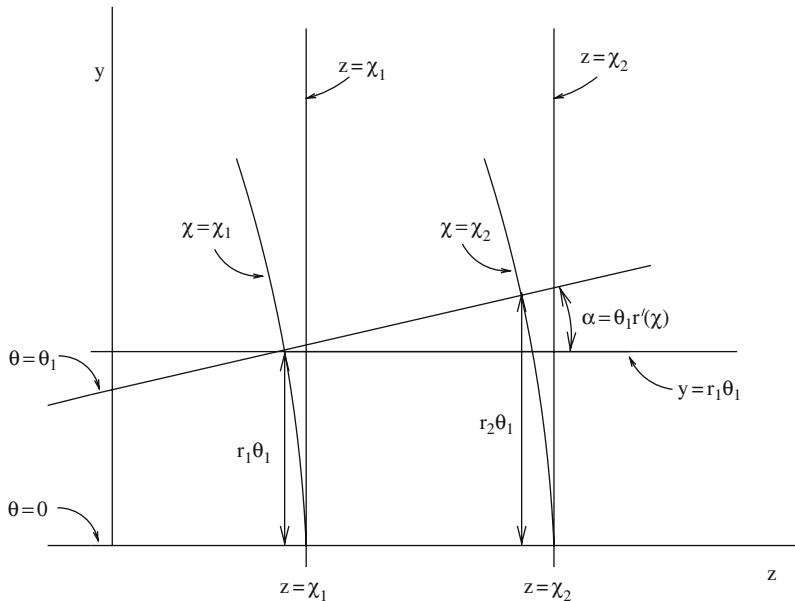


Fig. 3.11 A zoom of Fig. 3.10 around the point ($x = y = z = 0$). The angle θ_1 is small so we use the first-order transverse distance, $r\theta$. There is an angle $\alpha = \theta r'(\chi)$ between lines of constant θ and the z axis. Lines of constant χ are perpendicular to lines of constant θ and there is thus the same angle α between the lines of constant χ and the y axis

We note with some satisfaction that our careful choice of coordinates has resulted in a metric that is locally Euclidean at the origin, i.e. $dS^2 = a^2(dx^2 + dy^2 + dz^2)$ up to second-order corrections.

If the universe is homogeneous, the metric (3.44) should not depend on the choice of origin: $\alpha(r) = \text{constant}$ and $\beta(r) = \text{constant}'$. The solution of these two equations is

$$f(r) = \frac{1}{1 - br^2}, \quad (3.47)$$

where b is an arbitrary constant. Homogeneous spaces are thus defined by the value of b , with $b = 0$ corresponding to a flat space.

While (3.47) is the solution to our problem, we note that the local metric (3.44) at $r = r_0$ does not look like the original metric (3.32) at $r = 0$. To completely convince ourselves that the original metric (3.32) with $f(r)$ given by (3.47) represents a homogeneous universe, it would be good to find a coordinate transformation that turns the local metric (3.44) into (3.32). This is done in Exercise 3.12.

The complete metric for a homogeneous and isotropic universe now takes the form

$$d\tau^2 = dt^2 - a^2 \left[\frac{dr^2}{1-br^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right]. \quad (3.48)$$

It is possible to absorb the constant b into the definition of the scale factor $a(t)$ by defining $a' = a/\sqrt{|b|}$ and $k = b/|b|$. After the transformation $r' = \sqrt{|b|}r$ and dropping the primes, we find the standard form of the Robertson–Walker metric:

$$d\tau^2 = dt^2 - a^2 \left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right], \quad (3.49)$$

where $k = 0, \pm 1$. In the (χ, θ, ϕ) system, the metric is

$$d\tau^2 = dt^2 - a^2 \begin{bmatrix} \sin^2 \chi & & \\ d\chi^2 + \frac{\chi^2}{\sinh^2 \chi} (d\theta^2 + \sin^2 \theta d\phi^2) & & \\ & & \end{bmatrix} \quad k = \begin{array}{c} +1 \\ 0 \\ -1 \end{array}. \quad (3.50)$$

The relation between the two radial coordinates is

$$\begin{aligned} \sin \chi &= \chi - \chi^3/6 + \dots & k &= +1 \\ r &= \chi & k &= 0 \\ \sinh \chi &= \chi + \chi^3/6 + \dots & k &= -1 \end{aligned} \quad (3.51)$$

The spatial metric is Euclidean for $r \ll 1$ ($\chi \ll 1$).

We see that there are three types of homogeneous and isotropic universes corresponding to $k = 1, 0, -1$. In Chap. 4 we will see that these three cases correspond to $\Omega_T > 1$, $\Omega_T = 1$, and $\Omega_T < 1$.

It is difficult to visualize three-dimensional curved spaces but we will try to develop some intuition about them. For the cases $k = 0$ and $k = -1$, lines of constant (θ, ϕ) diverge because $r'(\chi) > 0$ everywhere. This means that there is no obvious “topological” difference between the two cases. In fact, the differences in the angle–distance relation depend on our choice of clocks used to make radial distance measurements. With the use of comoving clocks, the spatial geometry for $k = -1$ is non-Euclidean. On the other hand, we will show in Sect. 4.2 that if we choose to use laboratory clocks, the spatial geometry can be quite different. In particular, the case $k = -1$ with $a(t) \propto t$ represents a *flat* space-time with $\Omega_T = 0$.

The case $k = 1$ is more interesting because the distance between lines of constant (θ, ϕ) is proportional to $\sin \chi$. The lines therefore diverge and then converge to meet at $\chi = \pi$. The point at $\chi = \pi$ is called the antipode of the origin.² The situation is the three-dimensional analog of the (two-dimensional) surface of a sphere of radius $a(t)$ embedded in three dimensions. For each point on the surface there is only one point at a distance $a\pi$. The metric on the surface of a sphere has a form that is very similar to the Robertson–Walker metric for $k = 1$ (Exercise 3.6). It is perhaps not

² We will see in Chap. 5 that if $k = 1$, our antipode is most likely beyond our horizon but, for $\Omega_M = 1$, will become visible at the moment of maximal expansion.

surprising that a $k = 1$ universe can be considered to be the “surface” of a 3-sphere embedded in four Euclidean dimensions (Exercise 3.7). Readers wishing to imagine what we might look like to four-dimensional observers are encouraged to complete Exercise 3.6.

It is often said that a $k = 1$ universe has a finite volume corresponding to the space $\chi < \pi$. In fact, the metric does not exclude the possibility that there are points more distant from the origin than πa . It only tells us that the space $\chi > \pi$ is connected to our space by a single point (the origin’s antipode) making it extremely difficult to go there.

For $k = 1$, the volume of the universe ($\chi < \pi$) can be calculated:

$$V = a^3 \int_0^\pi d\chi \int_0^\pi d\theta \int_0^{2\pi} d\phi \sin^2 \chi \sin \theta = 2\pi^2 a^3, \quad (3.52)$$

where we have used the fact that lines of constant θ , ϕ , and χ are perpendicular.

Finally, it is important to remember that it is possible that the universe is inhomogeneous on very large scales and that the Robertson–Walker metric is only a useful approximation near our position. It is therefore dangerous to make “cosmic” conclusions based on its form.

3.5 Photon Propagation

According to general relativity, particles follow space-time geodesics, i.e., trajectories that are local maxima of the proper time. The general method for finding geodesics will be derived in Sect. 3.7. In this section we will treat only the most important trajectories in cosmology, those of photons traveling from distant galaxies to us at the origin. This problem is especially simple because one needs only to impose $d\tau = 0$ along the trajectory.

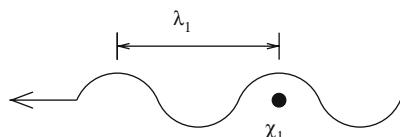
We consider a photon (Fig. 3.12) emitted at $(t_1, \chi_1, \theta_1, \phi_1)$ and then detected at $(t_0, \chi = 0)$. Supposing that the photon follows a path of constant θ and ϕ (to be justified in Sect. 3.7), the metric (3.50) implies

$$\left| \frac{d\chi}{dt} \right| = \frac{1}{a(t)}. \quad (3.53)$$

The time derivative of χ increases for decreasing $a(t)$ because the physical distance, dS , between comoving points decreases.

Fig. 3.12 A photon is emitted at (χ_1, t_1) with a wavelength λ_1 . It will be later detected by us at $(\chi = 0, t_0)$ with a wavelength λ_0

\bullet
 $\chi=0$



Integrating over the photon's path, we find a relation between (t_1, χ_1) and t_0 :

$$\chi_1 = \int_{t_1}^{t_0} \frac{dt}{a(t)} . \quad (3.54)$$

This equation defines the “light cone” associated with the point $(t_0, \chi = 0)$, i.e. the time t_1 at which we see galaxies at χ_1 . We can evaluate χ_1 if we know $a(t)$ as determined by the Friedmann equation, and we will do this in detail in Chap. 5. In Fig. 3.13 we show the photon trajectories for three values of the $(\Omega_M, \Omega_\Lambda)$ using the function $a(t)$ shown in Fig. 1.9.

For nearby galaxies, we can neglect the variation of $a(t)$ during the photon flight, in which case we find

$$a_0 \chi_1 \sim t_0 - t_1 \quad t_0 - t_1 \ll t_H . \quad (3.55)$$

This expression simply says that the flight-time is equal to the distance.

The expansion of the universe means that photon flight times between a given galaxy and us increase with time. Consider a galaxy that emits two photons, one at (t_1, χ_1) and a second one later at $(t_1 + \Delta t_1, \chi_1)$. The two photons will be detected by us at $(t_0, \chi = 0)$ and $(t_0 + \Delta t_0, \chi = 0)$. Equation (3.54) applies to both photons so we have

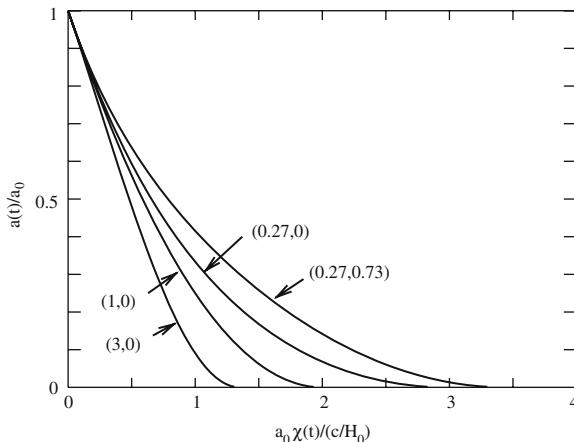


Fig. 3.13 The trajectory $\chi(t)$ of a photon that arrives at $\chi = 0$ at t_0 as calculated according to (3.53). The time (vertical) axis is parameterized by $\hat{a}(t)$. The trajectory is shown for four combinations of $(\Omega_M, \Omega_\Lambda)$ that range from a highly decelerated universe $(\Omega_M, \Omega_\Lambda) = (3, 0) \Rightarrow q_0 = 3/2$ to an accelerated (Λ CDM) universe $(\Omega_M, \Omega_\Lambda) = (0.27, 0.73) \Rightarrow q_0 = -0.595$. The figure demonstrates an important correlation between the present distance of a photon source of a given redshift and the deceleration parameter. For example, consider a photon emitted at $\hat{a}_1 = 0.5$ corresponding to a redshift $(z+1=2)$. We see in the figure that the present distance of the source is $a_0 \chi_1 \sim 0.5 H_0^{-1}$ for $(\Omega_M = 3, \Omega_\Lambda = 0)$, whereas the present distance is $a_0 \chi_1 \sim 0.8 H_0^{-1}$ for $(\Omega_M = 0.27, \Omega_\Lambda = 0.73)$.

$$\chi_1 = \int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_{t_1 + \Delta t_1}^{t_0 + \Delta t_0} \frac{dt}{a(t)} = \int_{t_1}^{t_0} - \int_{t_1}^{t_1 + \Delta t_1} + \int_{t_0}^{t_0 + \Delta t_0} .$$

Neglecting the expansion of the universe during the short times Δt_1 and Δt_0 we find

$$\frac{\Delta t_0}{\Delta t_1} = \frac{a_0}{a_1} . \quad (3.56)$$

This implies that the time duration of a pulse of photons will be greater at the moment of detection than at the moment of emission. This effect has been confirmed by measurements of the apparent durations of supernova explosions at high redshift [109]. The physical cause of this “cosmological time dilation” is simply that, in order to reach the observer, each photon has a distance to travel that is greater than the distance traveled by the preceding photon.

The relation (3.54) also determines the redshift of a galaxy as a function of χ_1 . We consider again our photon emitted in a galaxy at $(t_1, \chi_1, \theta_1, \phi_1)$ with wavelength λ_1 . We can treat the photon as a classical electromagnetic wave with each crest of the wave traveling at the speed of light. We can therefore apply (3.54) to two consecutive crests, the first crest leaving χ_1 at t_1 and the second crest leaving at $t_1 + \lambda_1/c$. The two crests are detected, respectively, at $(t_0, \chi = 0)$ and $(t_0 + \lambda_0/c, \chi = 0)$. Equation (3.54) gives

$$\chi_1 = \int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_{t_1 + \lambda_1}^{t_0 + \lambda_0} \frac{dt}{a(t)} = \int_{t_1}^{t_0} - \int_{t_1}^{t_1 + \lambda_1} + \int_{t_0}^{t_0 + \lambda_0} .$$

Neglecting the variation of $a(t)$ over one period λ/c , we find an elegant expression for the redshift

$$z + 1 \equiv \frac{\lambda_0}{\lambda_1} = \frac{a_0}{a_1} . \quad (3.57)$$

We see that the photon wavelength “expands” by the same factor as the universe itself. This confirms the law we derived in Sect. 1.2.4.

The integral for χ_1 can be evaluated for nearby sources by using the Taylor expansion (1.98) for $a(t)$:

$$a(t) = a_0 \left[1 + \frac{t - t_0}{t_H} - \frac{q_0}{2} \left(\frac{t - t_0}{t_H} \right)^2 + \dots \right] .$$

Substituting the first two terms into (3.57) we find

$$z \sim H_0 (t_0 - t_1) \sim H_0 \chi_1 a_0 \quad (t_0 - t_1) \ll t_H . \quad (3.58)$$

This is simply the Hubble law (1.15) since $z \sim v/c$ for $z \ll 1$.

Using (3.57), we can now find an expression for χ_1 as a function of redshift by changing variables in (3.54): $t \rightarrow a(t) \rightarrow z' = a_0/a(t)$

$$\chi_1(z) = \int_{a_0/(1+z)}^{a_0} \frac{da}{a\dot{a}} = a_0^{-1} \int_0^z \frac{dz'}{\dot{a}/a} . \quad (3.59)$$

This expression is more useful than (3.54) because the redshift of a galaxy can be measured whereas the time of emission t_1 cannot. The Taylor expansion for $a(t)$ gives

$$\frac{\dot{a}}{a} = H_0 \left[1 - (1 + q_0) \frac{t - t_0}{t_H} \right] = H_0 \left[1 + (1 + q_0)z' \right] . \quad (3.60)$$

Substituting this into (3.59) we find

$$a_0 \chi_1(z) = H_0^{-1} z \left[1 - \frac{1 + q_0}{2} z + \dots \right] \quad z \ll 1 . \quad (3.61)$$

The first-order result $a_0 \chi_1(z) = H_0^{-1} z$ is Hubble's law. The correction in brackets depends on the deceleration parameter q_0 and we will use this fact in Chap. 5 to constrain the cosmological parameters. The fact that the distance $a_0 \chi_1$ is a decreasing function of q_0 has a simple explanation that is illustrated in Fig. 3.13. A given redshift corresponds to a fixed expansion factor a_0/a_1 between emission and absorption. Deceleration implies fast expansion (in the past) diminishing the time to expand from a_1 to a_0 and, therefore, the flight time. A diminished flight time implies a diminished distance $a_0 \chi_1$ between source and observer.

We note that since $r \sim \chi + O(\chi^3)$ and $\chi \propto z + O(z^2)$, (3.61) also applies to r_1 :

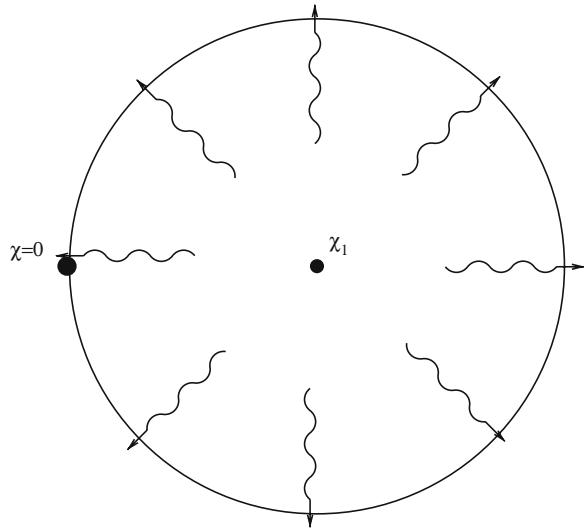
$$a_0 r_1(z) = H_0^{-1} z \left[1 - \frac{1 + q_0}{2} z + \dots \right] \quad z \ll 1 . \quad (3.62)$$

3.6 Observable Distances

We now have all the concepts necessary to calculate the photon fluxes and angular sizes necessary to interpret astronomical observations. As in the previous section, we consider a photon source of luminosity L (energy per unit time) that has a radial coordinate χ_1 . Over a time Δt_1 centered on t_1 , the source emits N photons of mean energy E_1 with $L = N E_1 / \Delta t_1$. As illustrated in Fig. 3.14, at t_0 the photons are distributed over a shell of radius $a_0 \chi_1$ centered on the source. The area of the shell's surface is $4\pi a_0^2 r_1^2$. The photons now have a mean energy $E_0 = E_1 a_1 / a_0$ and take a time $\Delta t_0 = \Delta t_1 a_0 / a_1$ to cross the shell. The energy flux is therefore

$$f = \frac{N E_0}{4\pi a_0^2 r_1^2 \Delta t_0} = \frac{N E_1 / \Delta t_1}{4\pi a_0^2 r_1^2 (1+z)^2} = \frac{L}{4\pi d_L^2} , \quad (3.63)$$

Fig. 3.14 Photons emitted at (t_1, χ_1) are, at t_0 , spread out over the surface of a sphere of radius $a_0\chi_1$ and surface area $4\pi a_0^2 r_1^2$



where this relationship defines the “luminosity distance” of the source at χ_1

$$d_L = a_0 r_1 (1 + z). \quad (3.64)$$

Equations (3.63) and (3.64) generalize the standard “ $1/R^2$ ” law of fluxes. The new law takes into account the size of the shell surrounding the source, the redshift of the photon energies, and the cosmological time dilation. For a nearby source, we can use (3.61) for r_1 yielding

$$d_L(z) = H_0^{-1} z \left[1 - \frac{1+q_0}{2} z + \dots \right] (1+z) \quad (3.65)$$

$$= H_0^{-1} z \left[1 + \frac{1-q_0}{2} z + \dots \right]. \quad (3.66)$$

As with the distance $a_0\chi_1$, and for the same reasons, the luminosity distance is a decreasing function of q_0 . At fixed redshift, objects appear brighter in a decelerating universe than in an accelerating universe.

Equation (3.63) is not directly applicable to observations because one never measures the total energy flux, f , but only the differential flux, f_ν , integrated over the band-pass of a detector. The total flux is its integral:

$$f = \int_0^\infty f_\nu(\nu_0) d\nu_0 \quad \nu_0 = c/\lambda_0. \quad (3.67)$$

The differential flux $f_\nu(\nu_0)$ is determined by the differential luminosity $L_\nu(\nu_0(1+z))$:

$$f_\nu(\nu_0) = \frac{L_\nu(\nu_0(1+z))}{4\pi a_0^2 \chi(z)^2(1+z)} \quad \int_0^\infty L_\nu(\nu_1) d\nu_1 = L \quad (3.68)$$

There is one fewer factors of $(1+z)$ in the denominator of (3.68) than in (3.63) because the flux emitted in an interval $\Delta\nu_1$ is observed in a smaller interval $\Delta\nu_1/(1+z)$. The measured integral over the pass-band, $T(\nu_0)$, is:

$$\int T(\nu_0) f_\nu(\nu_0) d\nu_0 = \frac{1}{4\pi a_0^2 \chi^2(1+z)} \int T(\nu_0) L_\nu(\nu_1) d\nu_0 . \quad (3.69)$$

Another relation we will need for astronomical observations is that between angular distance seen on the sky between two objects separated by a given distance. In Fig. 3.15 we show two points at radial coordinate χ_1 separated by a distance ΔS at t_1 . The angular separation $\Delta\Theta = (\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2)^{1/2}$ is determined by the metric (3.28) giving $\Delta S = a_1 r_1 \Delta\Theta$. A photon is emitted from each point at t_1 and later detected at $(t_0, \chi = 0)$. During their voyage to the observer at $\chi = 0$, the two photons follow paths of constant (θ, ϕ) so the photons are observed to be separated by the original angle

$$\Delta\Theta = \frac{\Delta S}{a_1 r_1} = \frac{\Delta S}{d_A} . \quad (3.70)$$

This relation defines the “angular distance”

$$d_A = a_1 r_1 = \frac{d_L}{(1+z)^2} . \quad (3.71)$$

The angular distance is necessarily smaller than the luminosity distance meaning that objects will appear larger on the sky than what would be expected from their brightness.

A final relation that we will need is that giving the difference in redshift between two objects separated by a given distance ΔS along the radial direction. The metric (3.28) gives $\Delta S = a(t_1)\Delta\chi$. For small $\Delta\chi$, (3.59) gives $a_0\Delta\chi = \Delta z/H(t_1)$ where $H = \dot{a}/a$. This gives

$$\Delta z = \frac{\Delta S}{H(t_1)^{-1}/(1+z)} . \quad (3.72)$$

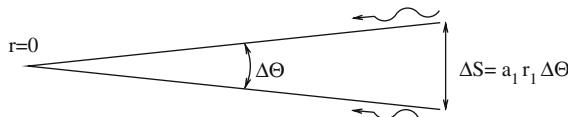


Fig. 3.15 Photons emitted at (t_1, χ_1) by an object of size $\Delta S = a_1 r_1 \Delta\Theta$. During their voyage toward the origin, the photons follow paths of constant (θ, ϕ) and will therefore be observed separated by an angle $\Delta\Theta = \Delta S/(a_1 r_1)$

Equations (3.63), (3.70), and (3.72) are the three relations that will allow us in Chap. 5 to determine the cosmological density parameters Ω_M and Ω_Λ .

3.7 The Geodesic Equation

In the absence of non-gravitational forces, a particle follows a trajectory that maximizes its proper time. Such a trajectory is called a space-time geodesic. In this section, we will find the general equation for geodesics and apply it to the Robertson-Walker metric.

Consider a four-dimensional space-time covered by a coordinate system x^μ , $\mu = 0, 1, 2, 3$ and characterized by a metric $g_{\mu\nu}$:

$$d\tau^2 = g_{\mu\nu} dx^\mu dx^\nu.$$

The trajectory of a particle can be defined either by three functions $x^i(x^0)$, $i = 1, 2, 3$ giving the spatial position as a function of coordinate time or by four functions $x^\mu(\tau)$, $\mu = 0, 1, 2, 3$ giving the particles coordinates as a function of its own clock reading. In the second case the metric imposes a constraint on the first derivatives:

$$1 = g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau}.$$

The use of τ to parameterize the trajectory does not work for massless particles so this case will require special treatment.

Consider a geodesic $x^\mu(\tau)$ between the space-time points A and B. The proper time elapsed between A and B is given by

$$\tau_{AB} = \int_A^B \sqrt{g_{\mu\nu}(x) dx^\mu dx^\nu}. \quad (3.73)$$

Consider now a non-geodesic between A and B defined by $x^\mu(\tau) + \varepsilon^\mu(\tau)$, where τ is the proper time along the geodesic $x^\mu(\tau)$ starting at the point A. This non-geodesic starts and ends at the same points so we impose $\varepsilon(0) = \varepsilon(\tau_{AB}) = 0$. The total proper time for the non-geodesic is

$$\tau_{AB}^\varepsilon = \int_A^B \sqrt{g_{\mu\nu}(x + \varepsilon)(dx^\mu + d\varepsilon^\mu)(dx^\nu + d\varepsilon^\nu)}. \quad (3.74)$$

If the functions $x^\mu(\tau)$ describe geodesics, the proper time τ_{AB} must be stationary with respect to small variations $\varepsilon(\tau)$:

$$\tau_{AB}^\varepsilon - \tau_{AB} = 0 + O(\varepsilon^2). \quad (3.75)$$

We therefore evaluate τ_{AB}^ε to order using

$$g_{\mu\nu}(x + \varepsilon) = g_{\mu\nu}(x) + \left[\frac{\partial g_{\mu\nu}}{\partial x^\lambda} \right]_x \varepsilon^\lambda . \quad (3.76)$$

Keeping only terms to order ε , the proper time along the non-geodesic is

$$\tau_{AB}^\varepsilon = \int_A^B \left[g_{\mu\nu} dx^\mu dx^\nu + \frac{\partial g_{\mu\nu}}{\partial x^\lambda} \varepsilon^\lambda dx^\mu dx^\nu + 2g_{\mu\nu} dx^\mu d\varepsilon^\nu \right]^{1/2} , \quad (3.77)$$

where we have used the symmetry of the metric to equate $g_{\mu\nu} dx^\mu d\varepsilon^\nu$ and $g_{\mu\nu} dx^\nu d\varepsilon^\mu$. We now factor out the first-order term to get

$$\tau_{AB}^\varepsilon = \int_A^B [g_{\mu\nu} dx^\mu dx^\nu]^{1/2} \left[1 + \frac{\partial g_{\mu\nu}}{\partial x^\lambda} \varepsilon^\lambda \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} + 2g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{d\varepsilon^\nu}{d\tau} \right]^{1/2} , \quad (3.78)$$

where we have used $d\tau^2 = g_{\mu\nu} dx^\nu dx^\mu$ to simplify the appearance of the second square root. Now expanding the second square root and keeping only the terms linear in ε we have

$$\begin{aligned} \tau_{AB}^\varepsilon &= \int_A^B [g_{\mu\nu} dx^\mu dx^\nu]^{1/2} \\ &+ \int_A^B [g_{\mu\nu} dx^\mu dx^\nu]^{1/2} \left[g_{\mu\nu}(x) \frac{d\varepsilon^\nu}{d\tau} \frac{dx^\mu}{d\tau} + \frac{1}{2} \frac{\partial g_{\mu\lambda}}{\partial x^\nu} \varepsilon^\nu \frac{dx^\mu}{d\tau} \frac{dx^\lambda}{d\tau} \right] . \end{aligned} \quad (3.79)$$

Once again using $d\tau^2 = g_{\mu\nu} dx^\mu dx^\nu$, this is just

$$\begin{aligned} \tau_{AB}^\varepsilon &= \tau_{AB} + \int_A^B d\tau \left[g_{\mu\nu}(x) \frac{d\varepsilon^\nu}{d\tau} \frac{dx^\mu}{d\tau} + \frac{1}{2} \frac{\partial g_{\mu\lambda}}{\partial x^\nu} \varepsilon^\nu \frac{dx^\mu}{d\tau} \frac{dx^\lambda}{d\tau} \right] \\ &= \tau_{AB} + \int_A^B d\tau \frac{d}{d\tau} \left(g_{\mu\nu}(x) \frac{dx^\mu}{d\tau} \varepsilon^\nu \right) \\ &\quad - \int_A^B d\tau \left[\varepsilon^\nu \frac{d}{d\tau} \left(g_{\mu\nu}(x) \frac{dx^\mu}{d\tau} \right) - \varepsilon^\nu \frac{1}{2} \frac{\partial g_{\mu\lambda}}{\partial x^\nu} \frac{dx^\mu}{d\tau} \frac{dx^\lambda}{d\tau} \right] \end{aligned}$$

The first integral equals zero because $\varepsilon(0) = \varepsilon(\tau_{AB}) = 0$. Because the function ε is arbitrary, the second integral equals zero only if $x^\mu(\tau)$ satisfies the equation:

$$\frac{d}{d\tau} \left(g_{\mu\nu}(x) \frac{dx^\mu}{d\tau} \right) - \frac{1}{2} \frac{\partial g_{\mu\lambda}}{\partial x^\nu} \frac{dx^\mu}{d\tau} \frac{dx^\lambda}{d\tau} = 0 \quad \nu = 0, 1, 2, 3 . \quad (3.80)$$

This equation is called the “geodesic equation.” It determines the trajectory of particles in the absence of non-gravitational forces.

For the Robertson–Walker metric

$$d\tau^2 = dt^2 - a^2(t) [d\chi^2 + r(\chi)^2(d\theta^2 + \sin^2\theta d\phi^2)] , \quad (3.81)$$

the path of a particle is defined by four functions, $t(\tau)$, $\chi(\tau)$, $\theta(\tau)$, and $\phi(\tau)$. The metric (3.81) imposes one constraint on the first derivatives of the four functions:

$$t'^2 - a^2 \chi'^2 - a^2 r^2 (\theta'^2 + \sin^2\theta \phi'^2) = 1 , \quad (3.82)$$

where $\chi' = d\chi/d\tau$, etc. The four geodesic equations (3.80) are:

$$\frac{d}{d\tau} (a^2 \chi') = a^2 \left(\chi'' + 2\frac{\dot{a}}{a} t' \chi' \right) = a^2 \frac{r}{f(r)} (\theta'^2 + \sin^2\theta \phi'^2) , \quad (3.83)$$

$$\phi'' = \frac{-2}{r} r' \phi' - \frac{2}{\tan\theta} \theta' \phi' - 2\frac{\dot{a}}{a} t' \phi' , \quad (3.84)$$

$$\theta'' = \frac{-2}{r} r' \theta' + \sin\theta \cos\theta \phi'^2 - 2\frac{\dot{a}}{a} t' \theta' , \quad (3.85)$$

$$t'' = -a\ddot{a} [\chi'^2 + r^2 (\theta'^2 + \sin^2\theta \phi'^2)] . \quad (3.86)$$

Three types of solutions of these equations are especially interesting for cosmology:

1. “Galactic motion.” The solutions $\chi = \text{constant}$, $\phi = \text{constant}$, $\theta = \text{constant}$, $t = \tau$, correspond to the expected movement of galaxies in comoving coordinates.
2. Radial motion. The initial conditions $\theta' = \phi' = 0$ imply $\theta'' = \phi'' = 0$. It follows that purely radial movement between a galaxy and the origin is a solution, as we supposed in Sect. 3.5. In the case of radial motion, the equation for χ is particularly simple:

$$\frac{d}{d\tau} (a^2 \chi') = 0 \Rightarrow a^2 \chi' = \text{constant} . \quad (3.87)$$

3. Radial motion of photons. Equations (3.83), (3.84), (3.85), (3.86), and (3.87) cannot be directly applied to photons because $d\tau = 0$ for a massless particle, implying that τ cannot parameterize their trajectories. However, we can combine (3.87) and (3.86) to derive an equation for $\chi(t)$:

$$\ddot{\chi} = -2\frac{\dot{a}}{a} \dot{\chi} + a\ddot{a} \dot{\chi}^3 , \quad (3.88)$$

where the dot means d/dt . The solution $a\dot{\chi} = 1 = c$ of this equation is consistent with the definition of $d\chi$, (3.24).

3.8 Gravitational Lensing

As an application of the general equations of motion found in the last section, we will treat the case of a spherically symmetric gravitational potential $\Phi(r)$ to calculate the gravitational lensing of distant objects [110]. Restriction to spherical symmetry will greatly simplify the formalism while still allowing us to treat the most interesting problems for cosmology. The primary results will be the light deflection angle (3.102) and the weak shear (3.113) giving the “stretch” factor of background galaxies due to a foreground galaxy.

For $\Phi(r) \ll 1$ the metric is given by³

$$dt^2 \sim dt^2 [1 + 2\Phi(r)] - [1 - 2\Phi(r)] dr^2 , \quad (3.89)$$

or

$$dt^2 \sim dt^2 [1 + 2\Phi(r)] - [1 - 2\Phi(r)] [dx^2 + dy^2 + dz^2] . \quad (3.90)$$

As always, we have to be careful in giving operational definitions of the coordinates. A clock at a fixed position can determine the temporal coordinate t by using

$$d\tau \sim dt [1 + \Phi(r)] \quad dr = 0 . \quad (3.91)$$

Spatial coordinates can be determined from photon flight times, e.g.

$$dr \sim dt [1 + 2\Phi(r)] \quad d\tau = 0 . \quad (3.92)$$

The potential Φ needed in these relations can be determined from the gravitational redshift, as illustrated in Fig. 3.16.

The most important cosmological use of this metric is the gravitational lensing of distant objects. This effect comes about from the gravitational effect on photon trajectories as illustrated in Fig. 3.17.

To calculate a photon’s trajectory, we need to solve the equations of motion for photons in a gravitational field. We start with the geodesic equation for massive particles (3.80) and apply it to the metric (3.90). For this time-independent metric, the equation for $\mu = 0$ is simple:

$$\frac{d}{d\tau} \left[g_{00} \frac{dt}{d\tau} \right] = 0 \quad \Rightarrow \quad \frac{dt}{d\tau} = \frac{A}{g_{00}} = \frac{A}{1 + 2\Phi(r)} , \quad (3.93)$$

where A is a constant (that will drop out, in the end). The equation for $i = 1, 2, 3$ is more complicated:

³ For single point mass, the metric (3.89) is a first approximation to the famous Schwarzschild metric, $g_{00} = (1 + 2\Phi)$ and $g_{rr} = -(1 + 2\Phi)^{-1}$.

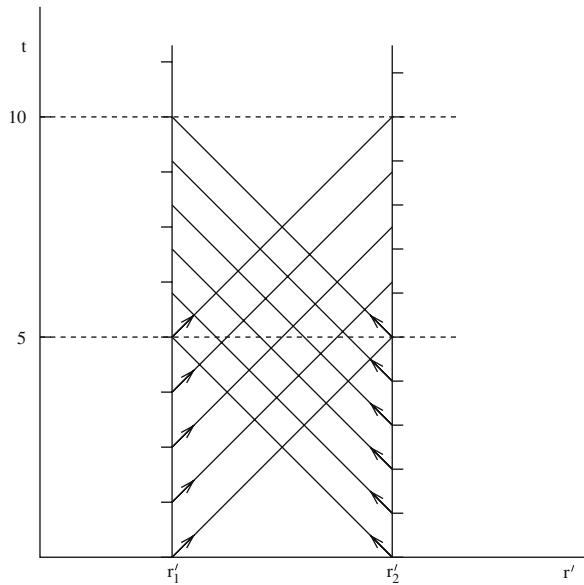


Fig. 3.16 The exchange of photons (diagonal lines) between two clocks, one at r_1 and one at r_2 in a potential $\Phi(r)$. The “ticks” of the clocks are marked by the short horizontal lines. The spatial axis gives $r' = \int^r dr(1 - 2\Phi)$ so that the photon trajectories appear as straight lines. Clock 2 is sufficiently far from the origin so that $\Phi(r_2) \sim 0$ and the clock measures the coordinate t , $dt = d\tau$. This can be seen by the fact that at $t = 5$, the clock at r_2 has ticked 5 times. Clock 1 is in the potential well and we choose $\Phi(r_1) = -0.2$. (We ignore terms of order $\Phi^2 = 0.04$!) At r_1 we have $d\tau = (1 - 0.2) dt = 0.8 dt$ and, because of this, the clock has only ticked 4 times at $t = 5$. At each tick, a clock sends a photon to the other clock. From the arrival times measured by clock 2, an observer at r_2 sees that clock 1 is running slowly (4 photons received for 5 ticks). Similarly, an observer at r_1 sees that clock 2 is running fast (5 photons received for 4 ticks). Observers at the two positions can determine the potential difference from this effect. Applied to crests of a classical electromagnetic wave, this reasoning can demonstrate the usual gravitational redshift and blueshift of individual photons, $\Delta\nu/\nu = \Delta\Phi$. Often it is said that photons “lose energy” as they climb out of potential wells. However, this demonstration shows that the gravitational redshift is more related to the differing rates of clocks in gravitational potentials. In spite of this, it is hard to avoid loosely speaking of photon energy loss, just as we often say that photons lose energy through the cosmological redshift

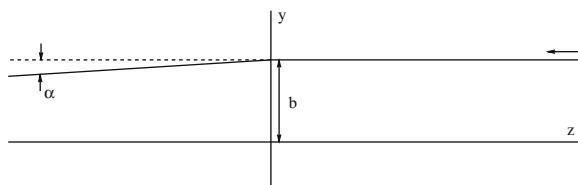


Fig. 3.17 The deflection of a photon in a gravitational potential $\Phi(r)$. The photon’s distance of closest approach (in the absence of deflection) to $r = 0$ is the “impact parameter,” b . The deflection angle α will depend on the form of the potential and, in general, on b , e.g. (3.101) or (3.103)

$$\frac{d}{d\tau} \left[g_{ii} \frac{dx^i}{d\tau} \right] - (1/2) \sum_{\mu} \frac{\partial g_{\mu\mu}}{\partial x^i} \left(\frac{dx^{\mu}}{d\tau} \right)^2 = 0 . \quad (3.94)$$

This equation does not apply to photons for which $d\tau = 0$. We can, however, use (3.93) to replace τ with t to obtain an equation applicable to photons:

$$\frac{A}{g_{00}} \frac{d}{dt} \left[g_{ii} \frac{dx^i}{dt} \frac{A}{g_{00}} \right] - (1/2) \sum_{\mu} \frac{\partial g_{\mu\mu}}{\partial x^i} \left(\frac{dx^{\mu}}{dt} \frac{A}{g_{00}} \right)^2 = 0 , \quad (3.95)$$

or

$$\frac{d}{dt} \left[\frac{dx^i}{dt} \frac{g_{ii}}{g_{00}} \right] = \frac{1}{g_{00}} \frac{\partial \Phi}{\partial x^i} \sum_{\mu} \left(\frac{dx^{\mu}}{dt} \right)^2 . \quad (3.96)$$

We consider the situation illustrated in Fig. 3.17 with the photon incident in the z direction in the yz plane. With no deflection, the trajectory would be simply

$$\frac{dz}{dt} = 1 \quad y(t) = b , \quad (3.97)$$

where b is the impact parameter. Since the scattering angle will turn out to be extremely small, we can neglect all factors in (3.96) that are proportional to dy/dt . In this case, the y equation is

$$g_{yy} \ddot{y} \sim \frac{\partial \Phi}{\partial y} \left[1 + \left(\frac{dz}{dt} \right)^2 \right] \sim 2 \frac{\partial \Phi}{\partial y} . \quad (3.98)$$

Since $\Phi \ll 1$, we can take $g_{yy} \sim -1$ giving

$$\ddot{y} \sim -2 \frac{\partial \Phi}{\partial y} . \quad (3.99)$$

This “transverse acceleration” is just twice what a simple Newtonian argument would have suggested.

The transverse velocity impulse is equal to the deflection angle, α :

$$\alpha = \Delta v_y/c = \int_{-\infty}^{\infty} \ddot{y} dt = 2 \int_{-\infty}^{\infty} \frac{\partial \Phi}{\partial y} dz . \quad (3.100)$$

For a point object of mass M we find the well-known deflection angle for photons passing near a star:

$$\frac{\partial \Phi}{\partial r} = \frac{GM}{r^2} \quad \Rightarrow \quad \alpha = \frac{4GM}{bc^2} . \quad (3.101)$$

The deflection angle is four times the gravitational potential $\Phi(b)$ at the distance of closest approach, or, equivalently, the four times the square of the circular velocity, v_c^2/c^2 , of massive particles at the distance of closest approach.

For cylindrically symmetric potentials, placing all of the mass on the z axis gives the same deflection so for such potentials we have

$$\alpha = \frac{4GM(b)}{bc^2} \quad \text{cylindrically symmetric potentials,} \quad (3.102)$$

where $M(b)$ is the mass within a distance b of the z axis.

A distribution of mass that is often used as a first approximation for galaxy clusters and galaxy halos is that for an “isothermal sphere”:

$$G\rho(r) = \frac{\sigma_v^2}{2\pi r^2} \Rightarrow GM(b) = \sigma_v^2 \int_{-\infty}^{\infty} dz \int_0^b \frac{b'db'}{z^2 + b'^2} = \pi\sigma_v^2 b^3, \quad$$

where σ_v is the one-dimensional velocity dispersion of objects bound in the potential. The rotational velocity for circular orbits is $\sqrt{2}\sigma_v$, independent of the orbital radius. The deflection angle is independent of b :

$$\alpha(b) = 4\pi \frac{\sigma_v^2}{c^2} \quad \text{isothermal sphere.} \quad (3.103)$$

To understand the phenomenology of lensing by spherically symmetric objects, it is useful to make the distinction between “strong lensing,” where the lensing center and the lensed object are well-aligned, and “weak lensing” where the position on the sky of the lensed object is well separated from the lensing center. The two configurations are illustrated in Figs. 3.18 and 3.20.

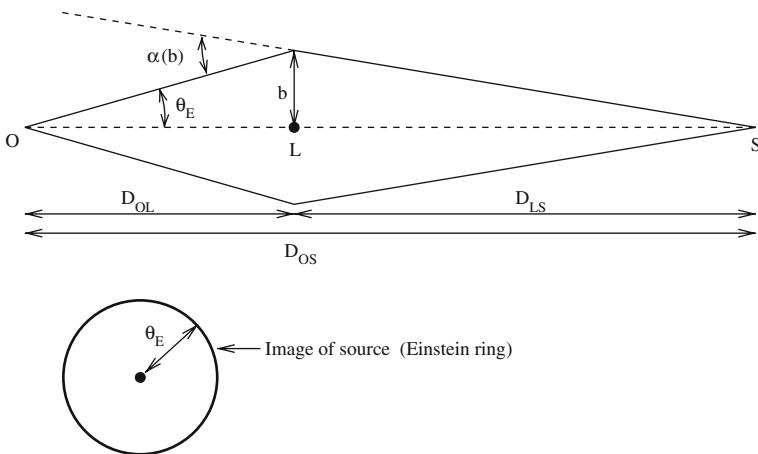


Fig. 3.18 The creation of an “Einstein ring” for a source S on the axis connecting the observer O and the lensing center L . The angles have been grossly exaggerated for clarity

Figure 3.18 shows a case of perfect alignment between the observer, lens, and source. By symmetry, the observer can see a ring around the lens center, a so-called “Einstein ring.” The angular radius of the ring θ_E can be calculated by solving for the impact parameter b in the equation

$$\alpha(b) = \frac{b}{D_{OL}} + \frac{b}{D_{LS}} \quad \theta_E \equiv \frac{b}{D_{OL}}, \quad (3.104)$$

where D_{OL} and D_{LS} are the observer–lens and lens–source distances. For an isothermal sphere, this gives

$$\theta_E = 4\pi \frac{\sigma_v^2}{c^2} \frac{D_{LS}}{D_{OS}} \quad \text{isothermal sphere.} \quad (3.105)$$

The most common available lenses are galaxies. A galaxy with a rotation velocity of 200 km s^{-1} has $(\sigma_v/c)^2 = 2 \times 10^{-7}$ and $\theta_E \sim 3 \times 10^{-6} = 0.5 \text{ asec}$ (for $D_{LS} \sim D_{OS}$). This is comparable to the angular resolution (seeing) of the best terrestrial telescopes so, seen from the Earth, the Einstein ring of a lensing galaxy cannot be separated from the lensing galaxy. This can be overcome by observing from space and Fig. 3.19 shows a collection of Einstein rings observed with the HST.

Production of a perfect Einstein ring requires perfect alignment of a circularly symmetric source with a circularly symmetric lensing potential. This is never possible though some of the rings in Fig. 3.19 are very impressive. They were found among roughly 1 million SDSS galaxies, which shows their rarity (Exercise 3.14). More common, though less spectacular, are visible arc-shaped galactic images which are formed whenever the unperturbed position of the background is separated by less than θ_E of the lens. In fact, these arcs are most easily seen using galactic clusters as lenses. Large clusters can have σ_v^2 , an order of magnitude greater than that of typical galaxies, making θ_E greater than the resolution of terrestrial telescopes so that arcs can be resolved. Examples can be seen in the cluster of Fig. 2.16.

If the unperturbed position of the background galaxy is separated by more than θ_E from the lensing center, then the image of the galaxy is only slightly perturbed. The geometry is shown in Fig. 3.20. The observed angle θ with respect to the lens center is related to the unperturbed angle θ_0 by

$$D_{OL}\theta + D_{LS}(\theta - \alpha(b)) = D_{OS}\theta_0. \quad (3.106)$$

with the implicit solution

$$\theta = \theta_0 + \frac{D_{LS}}{D_{OS}}\alpha(b = \theta/D_{OL}). \quad (3.107)$$

As illustrated in Fig. 3.20, this stretches the image of the lensed galaxy by a factor θ/θ_0 in the direction transverse to the vector separating the lensed galaxy from the lensing center. The image is also stretched in the perpendicular direction by a factor $d\theta/d\theta_0$. The total magnification factor, μ , of the galaxy is therefore

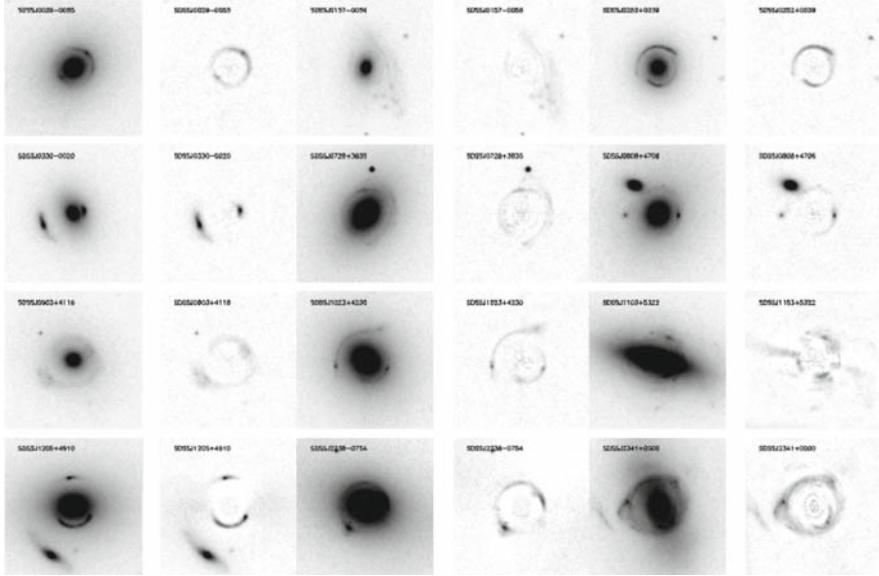


Fig. 3.19 A collection of approximate Einstein rings observed with the HST [111]. The lenses are galaxies with $z \sim 0.1$ and the rings are images of more distant galaxies that happen to lie on the Earth–Lens axis. Two frames are shown for each lens, the first with both lens and ring and the second after subtraction of the lensing galaxy. Viewed from the ground, atmospheric turbulence mixes the images of the foreground and background galaxies. These lenses were therefore discovered in SDSS spectra as mixtures of low-redshift and high-redshift galaxies. The candidate lenses were then confirmed with the HST images shown here

$$\mu = \left(\frac{\theta}{\theta_0} \right) \times \frac{d\theta}{d\theta_0} = \left(1 + \frac{D_{LS}\alpha(b)}{D_{OS}\theta_0} \right) \times \left(1 + \frac{D_{LS}}{D_{OS}} \frac{d\alpha}{d\theta_0} \right). \quad (3.108)$$

For weak lensing we can neglect the difference between θ and θ_0 when evaluating α and $d\alpha/d\theta_0$ in this expression. Using $\alpha(b) = 4GM(b)/b$ we then get

$$\begin{aligned} \mu &= \left(1 + \frac{D_{LS}D_{OL}4GM(b)}{D_{OS}b^2} \right) \times \left(1 + \frac{D_{LS}D_{OL}}{D_{OS}} \frac{d\alpha}{db} \right) \\ &\sim 1 + \frac{D_{LS}D_{OL}4GM(b)}{D_{OS}b^2} + \frac{D_{LS}D_{OL}}{D_{OS}} \frac{d\alpha}{db} \end{aligned} \quad (3.109)$$

Taking the derivative of α we find

$$\mu = 1 + \frac{D_{LS}D_{OL}}{D_{OS}} \frac{4G}{b} \frac{dM}{db} = 1 + \frac{D_{LS}D_{OL}}{D_{OS}} 8\pi G \Sigma(b), \quad (3.110)$$

where $\Sigma(b) = (2\pi b)^{-1}(dM/db)$ is the local surface density. For an isothermal sphere with a constant α , the magnification can be more directly read from (3.108) and (3.103):

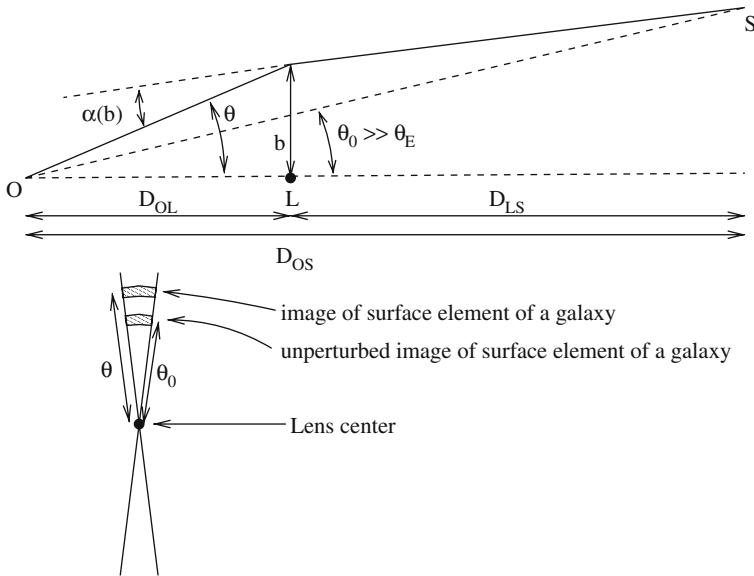


Fig. 3.20 The creation of an image of a source \$S\$ by a gravitational potential symmetric about the origin. The undeflected impact parameter is \$b_0 = D_{OL}\theta_0\$ while the actual photon path has an impact parameter \$b = D_{OL}\theta\$. The angles have been grossly exaggerated for clarity. As shown in the lower panel, because photon trajectories lie in a plane containing the \$S\$, the observer \$O\$, and the lens center \$L\$, the image of an extended object is stretched by a factor \$\theta/\theta_0\$.

$$\mu = 1 + \frac{\theta_E}{\theta} \quad \theta \gg \theta_E \quad (3.111)$$

Background galaxies appear larger if they are near foreground galaxies.

An effect that is simpler to observe than the magnification of background galaxies is the distortion of their shapes. This is characterized by the “shear,” \$\gamma\$, determined by the ratio of the tangential and radial distortions in (3.108):

$$1 + \gamma = 1 + \frac{D_{LS}D_{OL}}{D_{OS}} \left(\frac{\alpha(b)}{b} - \frac{d\alpha}{db} \right). \quad (3.112)$$

$$\gamma = \frac{D_{LS}D_{OL}}{D_{OS}} 8\pi G \left(\frac{M(b)}{\pi b^2} - \Sigma(b) \right) = \frac{D_{LS}D_{OL}}{D_{OS}} 8\pi G \Delta \Sigma(b) \quad (3.113)$$

where

$$\Delta \Sigma = \frac{M(b)}{\pi b^2} - \Sigma(b) \quad (3.114)$$

is the difference between the mean surface density inside \$b\$ and the local surface density. A measurement of galaxy distortions as in Fig. 2.14 permits one to deduce galactic masses. For an isothermal lens profile, the shear is

$$\gamma = 4\pi \frac{D_{\text{LS}}}{D_{\text{OS}}} \frac{(\sigma_v/c)^2}{\theta} = \frac{\theta_E}{\theta} \quad \theta \gg \theta_E . \quad (3.115)$$

Even when background galaxies are not lensed by a foreground cluster, the photon paths are still affected by the random gravitational potential due to mass inhomogeneities. This leads to small correlations in the apparent orientations of galaxies in the same angular region. One measurement [113] of the gravitational shear correlation function is shown in Fig. 3.21. These results have led to constraints on cosmological parameters, as discussed in Sect. 7.2.

Besides the distortion of individual images of background objects, use can be made of the fact that in certain cases there are two or more images of the background object. For lenses that are symmetrical about the line of sight, there can be an image above and below the lens as in Fig. 3.22. There are always two images for lensing by a point mass. For an isothermal sphere, there are two images if the undeflected image is within the Einstein ring: $\theta_0 < \theta_E$.

One important effect comes from the fact that the flight-times for the two images are not equal. Quasars have an intrinsic variability and comparing the light-curves (flux vs. time) of the two images determines the difference Δt of the flight-times.

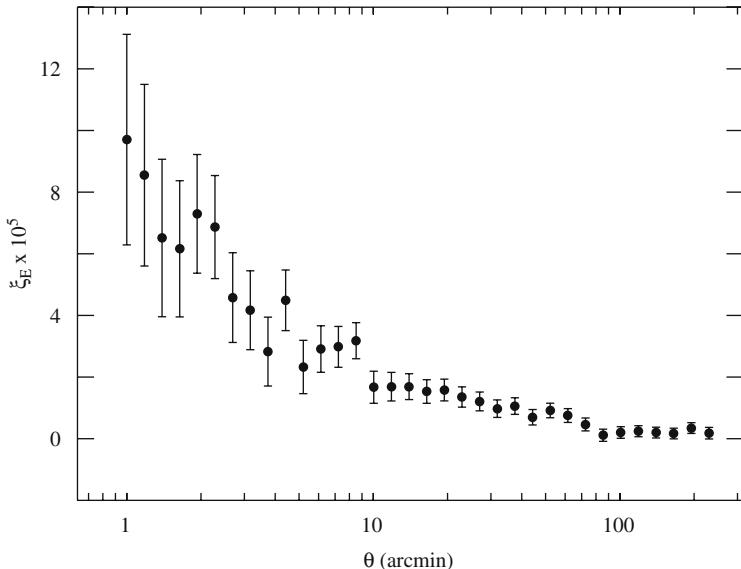


Fig. 3.21 The gravitational shear correlation function $\xi_E(\theta)$ [113] giving the mean product of gravitational shears of galaxies separated by an angle θ

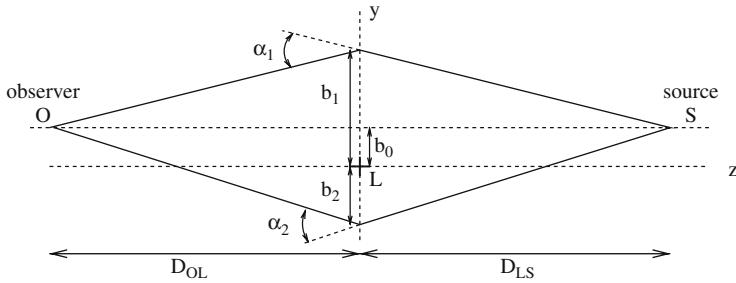


Fig. 3.22 The creation of two images of a source S by a gravitational potential symmetric about the origin. The undeflected impact parameter is b_0 while the actual photon paths have impact parameters b_1 and b_2 and deflection angles α_1 and α_2 . The angles have been grossly exaggerated for clarity

To calculate Δt , we impose $d\tau = 0$ in the metric (3.90):

$$dt \sim dz \left[1 - 2\Phi + \frac{1}{2} \left(\frac{dy}{dz} \right)^2 \right]. \quad (3.116)$$

The time delay is the difference between the integrals evaluated along the two paths. The first-order term, dz , clearly cancels leaving a “gravitational” term proportional to the potential difference and a “geometric” term proportional to the difference in $(dy/dz)^2$. For a deflection angle, α , independent of the point of impact, as is approximately the case for galaxy clusters, the geometric term vanishes leaving

$$\Delta t \sim 2 \int_{-\infty}^{\infty} dz [\Phi(y_2(z)) - (\Phi(y_1(z)))] , \quad (3.117)$$

where $y_1(z)$ and $y_2(z)$ are the trajectories of photons in the two images. We consider the nearly symmetric case, $|y_1| \sim |y_2|$. Referring to Fig. 3.22, the integral will be dominated by the region near the cluster so we can approximate:

$$\Phi(y_2(z)) - (\Phi(y_1(z)) \sim (b_1 - b_2) \frac{\partial \Phi}{\partial y} . \quad (3.118)$$

Substituting this into (3.117), we have

$$\Delta t \sim 2(b_1 - b_2) \int_{-\infty}^{\infty} dz \frac{\partial \Phi}{\partial y} . \quad (3.119)$$

The integral is simply the deflection angle, α , given by (3.100) so the time delay is

$$\begin{aligned}\Delta t &= 2D_{\text{OL}}\alpha \left(\frac{b_1}{D_{\text{OL}}} - \frac{b_2}{D_{\text{OL}}} \right) \\ &\sim 2D_{\text{OL}} \frac{b_1 + b_2}{D_{\text{OL}}} \left(1 + \frac{D_{\text{OL}}}{D_{\text{LS}}} \right) \left(\frac{b_1 - b_2}{D_{\text{OL}}} \right).\end{aligned}\quad (3.120)$$

The factor $(b_1 + b_2)/D_{\text{OL}}$ is the angular separation between the two images while the factor $(b_1 - b_2)/D_{\text{OL}}$ is the difference in the angular offsets from the lens center of the two images. Taking the angles and angle difference to be $\sim 10^{-5}$ rad and the distances to be $\sim d_{\text{H}}$ gives a $\Delta t \sim 1$ yr, as for the example in Fig. 3.23.

In (3.120), the angular separation of the two images is easily determined. The angular separation of the images from the lens center requires knowledge of the position of the center and this can be estimated, e.g., from the shear distortions of background galaxies. The ratio $D_{\text{OL}}/D_{\text{LS}}$ can be derived from the redshifts, though knowledge of $(\Omega_{\text{M}}, \Omega_{\Lambda})$ is also necessary for very distant sources. The measured time delay thus allows one to measure D_{OL} and determine H_0 . This method, discussed in Sect. 2.6.1, is limited in precision by the accuracy of the determination of the cluster center, or in the more general case, of the cluster density profile.

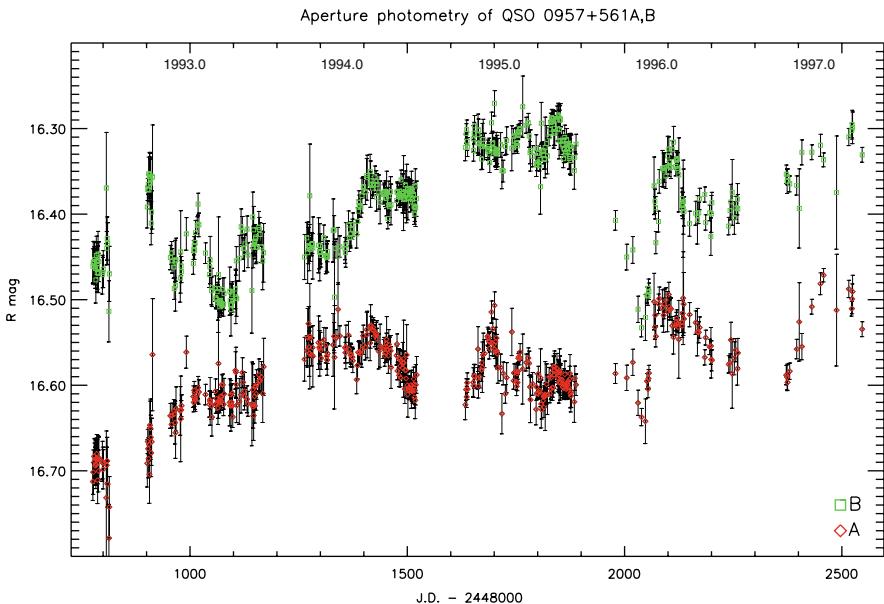


Fig. 3.23 The light curves of two images of QSO 0957+561 [112]. The quasar has a variable flux at the 10% (0.1 mag) level. The variations of the *top curve* are delayed by 424.9 ± 1.2 days from those of the *bottom curve*. For instance, the variation centered on day 1700 for the *lower curve* appears centered on day 2125 on the *top curve*

Finally, we note that because of the distortion of the lensed object, the total amount of light reaching the observer is changed. Figure 3.20 suggests that the total flux contained within a given image will be proportional to the angular size of the image since it is this size that determines the size of the “hole” in the lens plane through which photons must pass on their way from the source to the observer. (A more rigorous proof of the proportionality between the angular size and the total flux would use Liouville’s theorem.)

The ratio between the total flux in image i and the flux that would reach the observer in the absence of the lens is simply

$$A_i = \frac{b_i}{b_0} \left| \frac{db_i}{db_0} \right|. \quad (3.121)$$

The first factor is due to the stretching in the tangential direction and the second due to the stretching in the radial direction.

The amplification of the total flux has been observed of background stars by foreground stars in the Milky Way. For lensing by a point mass, the Einstein ring has an angular radius given by

$$\theta_E = \left(\frac{4GM D_{LS}}{c^2 D_{OL} D_{OS}} \right)^{1/2} = 1.4 \times 10^{-8} \left(\frac{M}{M_\odot} \frac{1 \text{ kpc}}{D_{OL} D_{OS}/D_{LS}} \right)^{1/2}$$

This angle is much smaller than the angular resolution for terrestrial telescopes so the existence of the lens can only be deduced from the amplification of the flux.

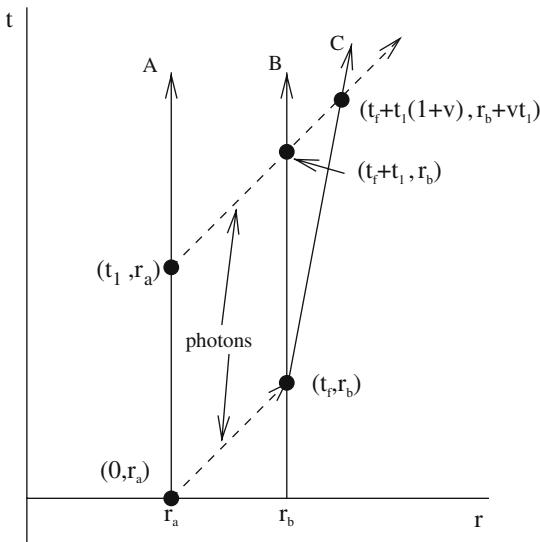
For a point-like lens, there are always two images of the background object, one on either side of the lens. It is relatively straightforward to show that the two amplifications are functions of the reduced unperturbed impact parameter $u = (b_0/D_{OL})/\theta_E$

$$A_{\pm} = \frac{u^2 + 2 \pm u\sqrt{u^2 + 4}}{2u\sqrt{u^2 + 4}} . \quad (3.122)$$

We see that for $u \gg 1$, $A_+ = 1$ and $A_- = 0$ as expected. For $u \rightarrow 0$ the amplifications become formally infinite corresponding to a point-like source being deformed into a ring. This divergence is removed by taking into account the finite size of the source giving a range of b_0 .

As reviewed in Sect. 2.5, gravitational lensing has been used to search for dark objects in our Galactic halo. Because the two images are not resolved, the effect is only seen through the total amplification given by summing the flux in the two images. This removes the \pm term giving the flux in equation (2.21). Because of the relative transverse movement of the source, lens, and observer, the total amplification will be time dependent. The background object will appear to brighten and then return to its original luminosity giving the light curve shown in Fig. 2.23.

Exercises



3.1 Consider the space around a black hole of mass M . The coordinates (t, r, θ, ϕ) can be chosen so that the metric is given by (Schwarzschild)

$$d\tau^2 = (1 - 2GM/r)dt^2 - \frac{dr^2}{1 - 2GM/r} - r^2(d\theta^2 + \sin^2\theta d\phi^2).$$

Figure above shows the space-time trajectories of three clocks, A, B, and C, and of two photons. Clocks A and B are fixed at $r_a > 2GM$ and $r_b > r_a$ while clock C recedes from the origin with a coordinate velocity $\dot{r} \equiv v$. The first photon is emitted at $(t = 0, r_a)$ and the second at (t_1, r_a) . All clocks and photons are at the same constant θ and ϕ .

- (a) Give an expression for the photon time-of-flight, t_f , from r_a to r_b . Do not bother to evaluate the integral.
- (b) What is the time, $\Delta\tau_A$, measured by clock A between the emission of the two photons?
- (c) What is the time, $\Delta\tau_B$, measured by clock B between its reception of the two photons?
- (d) Give an exact expression for $\Delta\tau_B/\Delta\tau_A$. What happens when $r_a = 2GM$?
- (e) Give an approximate expression for $\Delta\tau_B/\Delta\tau_A$ good to first order in $GM/r_a \ll 1$ and $GM/r_b \ll 1$.
- (f) Use the approximate expression to evaluate $\Delta\tau_B/\Delta\tau_A$ for $r_a = 6.4 \times 10^6$ m (the radius of the Earth), $r_b = 2.02 \times 10^7$ m (the radius of GPS satellite orbits), and $M = M_{\text{earth}} = 6.0 \times 10^{24}$ kg.

- (g) If clock C recedes slowly, $v \ll 1$ ($v \ll c$), then the second photon is received by clock C at $t \sim t_f + t_1(1 + v)$, $r \sim r_b + vt_1$. Calculate the ratio between $\Delta\tau_B$ and the time, $\Delta\tau_C$, measured by clock C between its reception of the two photons. Calculate only to first order in v and GM/r_b and assume that \dot{r}_C is constant over the trajectory. Neglect the variation of GM/r over the trajectory: $GM/r = GM/r_b$.

3.2 Consider a Lorentzian coordinate system \tilde{x} and another coordinate system x defined by

$$x^\mu = \frac{\partial x^\mu}{\partial \tilde{x}^\alpha} \tilde{x}^\alpha \quad \tilde{x}^\alpha = \frac{\partial \tilde{x}^\alpha}{\partial x^\mu} x^\mu , \quad (3.123)$$

where the transformations coefficients $\partial \tilde{x}^\alpha / \partial x^\mu$ are constants satisfying

$$\eta_{\alpha\beta} \frac{\partial \tilde{x}^\alpha}{\partial x^\mu} \frac{\partial \tilde{x}^\beta}{\partial x^\nu} = \eta_{\mu\nu} . \quad (3.124)$$

- (a) Show that the coordinates x are also Lorentzian:

$$d\tau^2 = \eta_{\alpha\beta} d\tilde{x}^\alpha d\tilde{x}^\beta = \eta_{\mu\nu} dx^\mu dx^\nu . \quad (3.125)$$

- (b) Show that spatial rotations satisfy (3.124), e.g.

$$\frac{\partial \tilde{x}^1}{\partial x^1} = \frac{\partial \tilde{x}^2}{\partial x^2} = \cos \theta \quad \frac{\partial \tilde{x}^1}{\partial x^2} = -\frac{\partial \tilde{x}^2}{\partial x^1} = \sin \theta . \quad (3.126)$$

- (c) Show that Lorentz boosts satisfy (3.124), e.g.

$$\frac{\partial \tilde{x}^0}{\partial x^0} = \frac{\partial \tilde{x}^1}{\partial x^1} = \gamma \quad \frac{\partial \tilde{x}^0}{\partial x^1} = \frac{\partial \tilde{x}^1}{\partial x^0} = \beta\gamma \quad (3.127)$$

where $\beta = v/c$ and $\gamma = (1 - \beta^2)^{-1/2}$.

3.3 By applying the geodesic equation (3.80) to the Lorentz metric (3.13), show that particles follow paths of constant velocity.

3.4 Consider the free-fall trajectory (3.14) of a slow rocket.

- (a) Calculate its proper time between $t = -t_1$ and $t = t_1$ using the proper time given by (3.15).

- (b) Calculate the proper time for the general trajectory (3.16).
 (c) Perform the same calculation for an airplane that flies for a time t at an altitude h before returning to its base.

3.5 Show that the geodesic equation (3.80) for the metric implied by (3.90) is the standard Newtonian equation, $\ddot{r} = -\nabla\Phi$.

3.6 Consider a sphere of radius a . A point on the surface can be located by the standard polar angle χ and azimuthal angle θ (Fig. 3.24).

- (a) Convince yourself that the distance dS between the point (χ, θ) and the point $(\chi + d\chi, \theta + d\theta)$ is given by

$$dS^2 = a^2 [d\chi^2 + \sin^2 \chi d\theta^2],$$

which can be compared with the metric (3.50).

- (b) Another polar coordinate r can be defined by $r = \sin \chi$. Show that dS is given by

$$dS^2 = a^2 \left[\frac{dr^2}{1 - r^2} + r^2 d\theta^2 \right],$$

which can be compared with the metric (3.49).

- (c) Consider an object of size $\Delta S \ll a$ at the point (χ, θ) . Suppose that the object emits photons that are constrained to move on the surface of the sphere. What

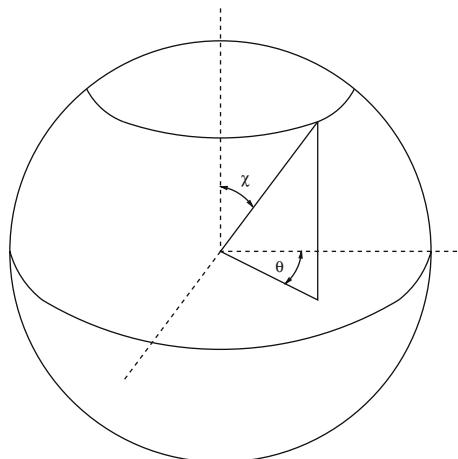


Fig. 3.24 A sphere of radius a . A point on the surface can be described by the polar angle χ and the azimuthal angle θ

will be the angular size of the object as seen by an observer at the point $\chi = 0$? How will this result be modified if the radius of the sphere is a function of time $a(t)$? Compare the results with (3.70).

- (d) Calculate the number of photons per unit length per unit time observed at $\chi = 0$. Compare the result with (3.63).

3.7 Consider a four-dimensional Euclidean space with coordinates x, y, z, w and metric $dS^2 = dx^2 + dy^2 + dz^2 + dw^2$. A “3-sphere” of radius a can be defined by the constraint $x^2 + y^2 + z^2 + w^2 = a^2$. Show that the distance dS between nearby points on the surface of the 3-sphere is given by the Robertson–Walker metric with $k = 1$ with $x = ra \sin\theta \cos\phi$, $y = ra \sin\theta \sin\phi$, and $z = ra \cos\theta$.

3.8 Consider a $k = 1$ universe. Explain why an object at $\chi = \pi$ has an angular size (seen by an observer at $\chi = 0$) of $\Delta\Omega = 4\pi$, i.e. the object covers the whole sky.

3.9 Verify (3.61).

Compare to order z^2 the following four distances between the origin and χ_1 :

1. the luminosity distance, d_L ;
2. the angular distance, d_A ;
3. the “proper” distance, $d_1 = a_1 \chi_1$, at t_1 ;
4. the “proper” distance, d_0 , at t_0 .

At what distance do the four distances start to differ by $\sim 10\%$?

3.10 An explorer leaves our galaxy at t_0 with a velocity $v \ll c$ after having used all her fuel.

- (a) Calculate $\dot{\chi}(t_0)$, $\chi'(t_0)$, and $t'(t_0)$.

- (b) Using (3.87) find an equation for $\dot{\chi}(t)$.

- (c) For $v = c/10$, what coordinate time t is necessary for the voyager to reach a galaxy at $z = 10^{-2}$. (The calculation can be done to order v/c .) Compare the travel time measured by comoving clocks with the time measured by the explorer.

- (d) For $v \ll c$, what is the redshift of the most distant galaxy that the explorer can visit in the case $a(t) \propto t$ and $a(t) \propto t^{2/3}$.

This exercise will be continued in Exercise 5.13.

3.11 (Demonstration of (3.44).) Using the transformations (3.41) and (3.43) show that

$$dx^2 + dy^2 = \sin^2\theta dr^2 + 2r \sin\theta \cos\theta dr d\theta + r^2(\sin^2\theta d\phi^2 + \cos^2\theta d\theta^2).$$

and that to order θ^2

$$dz^2 = d\chi^2 \left(1 - \frac{\theta^2}{f(r)} + \frac{rf'(r)\theta^2}{2f^2} \right) + \frac{r^2\theta^2 d\theta^2}{f(r)} - -2r\theta dr d\theta .$$

Combining these two results show that

$$dx^2 + dy^2 + dz^2 = dS^2 + r^2\theta^2 d\theta^2 \left(\frac{1}{f(r)} - 1 \right) + r\theta^2 d\chi^2 \frac{f'(r)}{2f^2} ,$$

where dS^2 is from (3.32). The two last terms on the right of this equation are of the order θ^2 , so we can use the approximations $f(r) = f(r_0)$, $d\chi = dz$, and $r^2\theta^2 d\theta^2 = (xdx + ydy)^2/r_0^2$. With these approximations, show that the spatial metric is given by (3.44).

3.12 With $f(r)$ given by (3.47), the metric (3.44) becomes:

$$\begin{aligned} dS^2 = & a^2 [dz^2(1 - bx^2 - by^2) \\ & + dy^2(1 + by^2) + dx^2(1 + bx^2) + 2bxydxdy] . \end{aligned} \quad (3.128)$$

This metric does not appear to describe a space that is isotropic about the point ($x = y = z = 0$) because the coordinates x and y enter the metric differently from the coordinate z . This cosmetic problem can be cured with a third-degree transformation:

$$\begin{aligned} x' &= x + b(xy^2 + xz^2 + x^3)/4 \\ y' &= y + b(yx^2 + yz^2 + y^3)/4 \\ z' &= z - b(x^2z + y^2z - z^3)/4 . \end{aligned} \quad (3.129)$$

With this transformation, show that the metric becomes (after dropping the primes)

$$\begin{aligned} dS^2 &= a^2 [dx^2 + dy^2 + dz^2][1 - b(x^2 + y^2 + z^2)/2] . \\ &= a^2 [dr^2 + r^2(d\theta^2 + \sin^2 \theta d\phi^2)][1 - r^2/2] . \end{aligned} \quad (3.130)$$

where $x = r \sin \theta \cos \phi$, $y = r \sin \theta \sin \phi$, and $z = r \cos \theta$. Show that with the transformation $r' = r(1 - r^2/4)$, the metric takes the form:

$$dS^2 = a^2 [(1 + br^2)dr^2 + r^2(d\theta^2 + \sin^2 \theta d\phi^2)] . \quad (3.131)$$

This metric is identical to the original metric (3.32) to order r^2 .

3.13 Figure 3.25 shows the trajectories of two observers in flat space-time. Observer B recedes from observer A with a velocity $v = \beta c$. A emits a photon at t_1 as

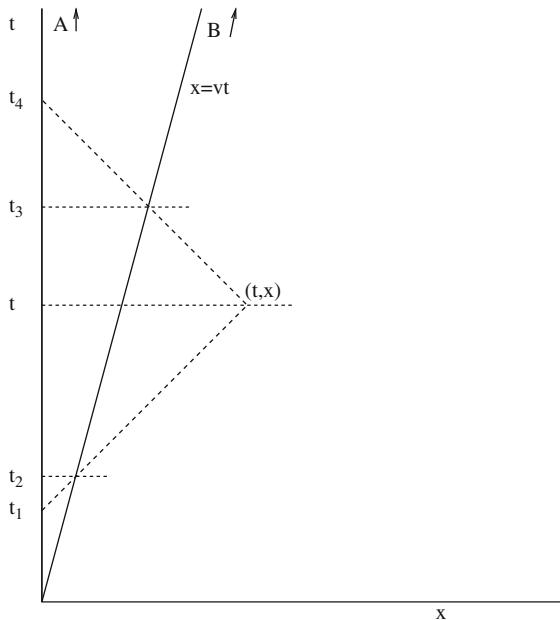


Fig. 3.25 The trajectories of two observers, observer A at the origin ($x = 0$) and observer B following the trajectory $x = vt$. A emits a photon at t_1 (as measured by A's clock). The photon is reflected at (t, x) returning to A at t_4 . A assigns coordinates to distant events by using its clock and radar, i.e. $t = (t_4 + t_1)/2$ and $x = (t_4 - t_1)/2$. B can do the same with its clock and radar

measured by its clock. The photon is reflected and returns at t_4 . A assigns the following coordinates to the reflection event:

$$t = (t_4 + t_1)/2 \quad x = (t_4 - t_1)/2 . \quad (3.132)$$

The photon passes by observer B at t_2 and then again at t_3 . (These two times are to be measured by A with its clock and radar.) The time measured by B's clock at these two events is

$$t'_2 = t_2 \sqrt{1 - \beta^2} \quad t'_3 = t_3 \sqrt{1 - \beta^2} . \quad (3.133)$$

B will assign the following coordinates to the reflection event:

$$t' = (t'_3 + t'_2)/2 \quad x = (t'_3 - t'_2)/2 . \quad (3.134)$$

(a) Show that (t', x') is related to (t, x) by a Lorentz transformation:

$$\begin{pmatrix} t' \\ x' \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma \\ -\beta\gamma & \gamma \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix} . \quad (3.135)$$

We see that only one observer is needed to set up a Lorentzian coordinate system. Of course, by exchanging photons observer A can synchronize his clock with other clocks at rest (with respect to A). If this is done, the time of (t, x) in A's frame can be measured locally.

- (b) The observer A carries a bar of length L oriented in the direction of B's movement. Use Fig. 3.26 to quickly show that B sees the bar to be Lorentz contracted by a factor $1/\gamma$.

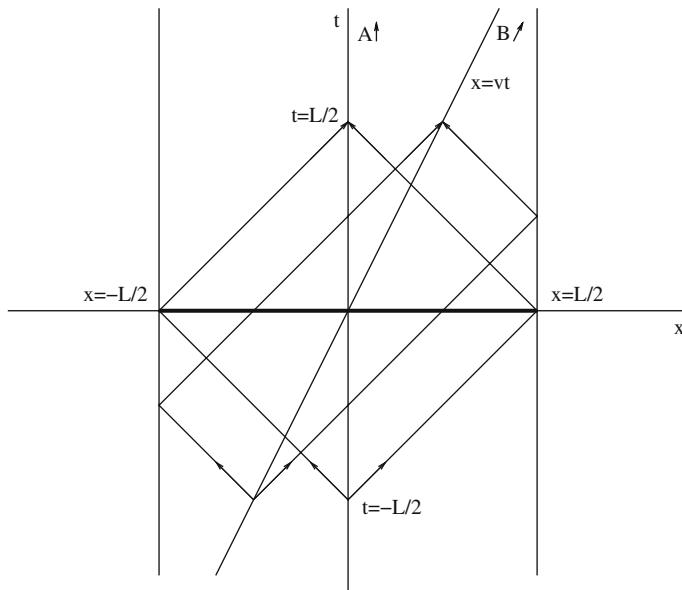


Fig. 3.26 As in Fig. 3.25 except that A carries a bar of length L in the direction of B's movement

- 3.14** Using the galaxy density (1.5), estimate the probability that a galaxy of redshift $z \sim 0.1$ has a background galaxy of redshift $0.4 < z < 0.5$ within its Einstein ring.

Chapter 4

The Field Equations

The evolution of the scale factor $a(t)$ is governed by gravitation via the Friedmann equation. In Chap. 1, we derived this equation using a rather dubious Newtonian argument. Here, we will re-derive the Friedmann equation within the framework of relativistic gravitation theory, i.e. general relativity. In this theory, the parameters of the metric are constrained by the Einstein field equation:

$$G^{\mu\nu} = -8\pi G T^{\mu\nu}. \quad (4.1)$$

The right-hand side is proportional to the energy-momentum tensor which acts as the source of gravitation. The left-hand side is the Einstein tensor $G^{\mu\nu}$ which is a function of the parameters of the metric. The 00 component of the field equation for the Robertson–Walker metric is called the Friedmann equation.

We will see that the general form of $G^{\mu\nu}$ is determined by the local conservation of energy:

$$\frac{\partial T^{\mu\nu}}{\partial x^\mu} = 0 \quad \Rightarrow \quad \frac{\partial G^{\mu\nu}}{\partial x^\mu} = 0. \quad (4.2)$$

This form of the conservation equation applies at the origin of a system of locally Lorentzian coordinates. The left equation simply says that the time derivative of the energy density (momentum density) is given by the divergence of the energy flux (momentum flux). The right equation follows from the left equation and the Einstein field equation. It will be all that we will need to fix the form of the Einstein tensor and to derive the Friedmann and Einstein field equations.

In this chapter, we will find it convenient to choose a coordinate system that is in free-fall. We will do this in Sect. 4.1, where we will transform the comoving coordinates of Chap. 3 into freely falling coordinates with the metric that is, up to second-order corrections, the metric of special relativity. The use of these coordinates will allow us, in Sect. 4.3, to introduce the energy-momentum tensor entirely within the framework of special relativity. After having found $T^{\mu\nu}$, it will be a simple matter to derive the Friedmann equation in Sect. 4.4 by imposing local energy conservation (4.2).

While the Friedmann equation is all that is necessary for homogeneous cosmology, for other applications, we will need the general form of the Einstein field equation. With this in mind, we will introduce the Riemann tensor in Sect. 4.7. It will allow us to understand the physical significance of the metric parameters since it is this tensor that governs the physically meaningful relative motion of test particles, i.e. geodesic deviation. Requiring that the Einstein tensor be the function of the Riemann tensor with zero 4-divergence will allow us, in Sect. 4.8, to find the field equation applicable in freely falling systems.

Finally, in Sect. 4.9, we will perform a general coordinate transformation that will give us the general form of the Einstein field equation.

4.1 Our Freely Falling Coordinates

In cosmology, one generally uses comoving coordinates (t, χ, θ, ϕ) or (t, r, θ, ϕ) with the Robertson–Walker metric:

$$\begin{aligned} dt^2 &= dt^2 - a(t)^2 \left[\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right], \\ g_{tt} &= 1 & g_{rr} &= \frac{a(t)^2}{1 - kr^2} & g_{\theta\theta} &= \frac{g_{\phi\phi}}{\sin^2 \theta} = a(t)^2 r^2 \end{aligned} \quad (4.3)$$

In spite of the utility of comoving coordinates, the dynamics of the universe, i.e. the time dependence of $a(t)$, is most easily understood if we choose a set of freely falling¹ coordinates $(\tilde{x}^\mu, \mu = 0, 1, 2, 3)$ with a metric that is locally Lorentzian near the origin $\tilde{x}^\mu = 0$:

$$g_{\mu\nu} = \eta_{\mu\nu} + \frac{1}{2}g_{\mu\nu,\alpha\beta}(0)\tilde{x}^\alpha\tilde{x}^\beta + \dots \quad (4.4)$$

Here $\eta_{\mu\nu}$ is the metric for flat space–time:

$$\eta_{00} = -\eta_{11} = -\eta_{22} = -\eta_{33} = 1 \quad \eta_{\mu\neq\nu} = 0, \quad (4.5)$$

and

$$g_{\mu\nu,\alpha\beta} \equiv \frac{\partial^2 g_{\mu\nu}}{\partial \tilde{x}^\alpha \partial \tilde{x}^\beta}. \quad (4.6)$$

The metric (4.4) is locally Lorentzian because the corrections to $\eta_{\mu\nu}$ are quadratic in the coordinates \tilde{x}^μ . The Robertson–Walker metric (4.3) is not locally Lorentzian because $\dot{a} \neq 0$.

¹ In a cosmological context, “freely falling” is perhaps a bad term since in a homogeneous universe there is no direction to fall.

The object of this section is to find a coordinate transformation $(t, r, \theta, \phi) \rightarrow \tilde{x}^\mu$ such that has the metric becomes

$$g_{00} = 1 - \alpha R^2 \quad g_{ii} = -(1 - \beta R^2/2) \quad i = 1, 2, 3, \quad (4.7)$$

with $R^2 = \tilde{x}^1 \tilde{x}^1 + \tilde{x}^2 \tilde{x}^2 + \tilde{x}^3 \tilde{x}^3$ and the coefficients α and β are

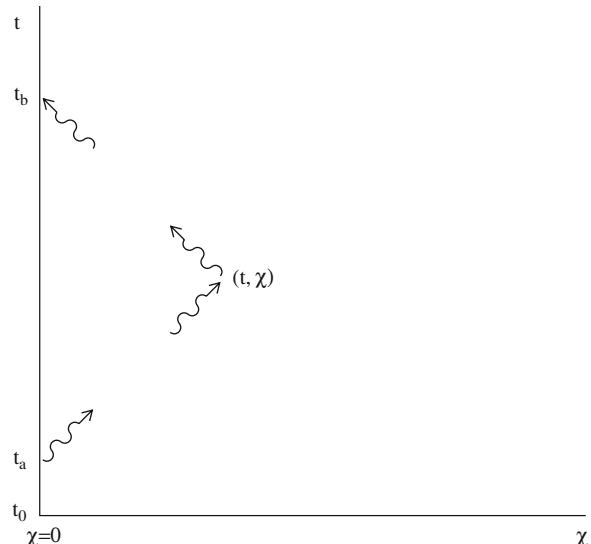
$$\begin{aligned} \alpha &= \left[\frac{\ddot{a}}{a} \right]_{t_0} \equiv -H_0^2 q_0 \\ \beta &= \left[\left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} \right]_{t_0} = H_0^2 + \frac{k}{a_0^2}. \end{aligned} \quad (4.8)$$

This metric is locally Lorentzian.

The transformation can be found by fiddling with the coordinates, but we choose to follow a more physical procedure. A general way to construct freely falling coordinates is to equip an observer in free-fall with a clock and a radar. The observer can then use his equipment in the usual way to measure times and distances. It turns out that this automatically gives coordinates with a metric that is locally Lorentzian. The observer that we will use is the one at the origin of the comoving coordinate system, i.e. us.

The situation is illustrated in Fig. 4.1. The observer at the origin ($\chi = 0$) attempts to measure the time and position of an event with comoving coordinates (t, χ) . At $t = t_a$ the observer emits a photon that is reflected at (t, χ) and returns to the origin at $t = t_b$. It is natural for the observer to give the coordinates (T, R) to the event with

Fig. 4.1 The construction of freely falling coordinates by an observer at the origin equipped with a clock and a radar. The observer emits a photon at $t = t_a$ that is reflected at the point with comoving coordinates (t, χ) . The photon returns to the origin at $t = t_b$. The observer can associate with the event (t, χ) a radial coordinate $R = (t_b - t_a)/2$ and the temporal coordinate $T = (t_b + t_a)/2 - t_0$. The metric of the system (T, R) is locally Lorentzian at the origin



$$T = \frac{t_b + t_a}{2} - t_0 \quad (4.9)$$

$$R = \frac{t_b - t_a}{2}. \quad (4.10)$$

The relation between the coordinates (T, R) and (t, r) is simple to find because the trajectory of the photon between t_a and t is governed by

$$\chi = \int_{t_a}^t \frac{dt'}{a(t')} = \int_{t_a}^t \frac{dt'}{a(t)[1 + H(t)(t' - t) + \dots]}, \quad (4.11)$$

and between t and t_b by

$$\chi = \int_t^{t_b} \frac{dt'}{a(t')} = \int_t^{t_b} \frac{dt'}{a(t)[1 + H(t)(t' - t) + \dots]}. \quad (4.12)$$

Taking the sum and the difference of these two equations we find (Exercise 4.1)

$$T = t - t_0 + a(t)\dot{a}(t)r^2/2 + O(r^3) \quad (4.13)$$

$$R = a(t)r + O(r^3). \quad (4.14)$$

To find the metric in the system (T, R) , we need the inverse transformations $t(T, R)$ and $r(T, R)$. Substituting (4.14) into (4.13) to get

$$t = t_0 + T - (1/2)R^2 \frac{\dot{a}}{a} \quad (4.15)$$

$$r = \frac{R}{a(t)}. \quad (4.16)$$

To eliminate the t dependence of $a(t)$ in these two equations we use the Taylor series for $a(t)$ (1.98):

$$a(t) = a_0 [1 + H_0(t - t_0) - (1/2)q_0 H_0^2(t - t_0)^2],$$

which implies

$$\frac{\dot{a}}{a} = H_0 [1 - (1 + q_0) H_0 (t - t_0)].$$

Substituting this into (4.15) and (4.16) we find

$$t = t_0 + T - (1/2)R^2 H_0 [1 - (1 + q_0) H_0 (t - t_0)]$$

$$r = \frac{R}{a_0} [1 - H_0(t - t_0) + (1 + q_0/2) H_0^2(t - t_0)^2].$$

Since we are working to order R^2 and T^2 , we can use $t - t_0 \sim T - (1/2)R^2H_0$ to find the final transformations $t(T, R)$ and $r(T, R)$:

$$t = t_0 + T - (1/2)R^2H_0[1 - (1 + q_0)H_0T] + \dots \quad (4.17)$$

$$r = \frac{R}{a_0}[1 - H_0(T - R^2H_0/2) + (1 + q_0/2)H_0^2T^2] + \dots \quad (4.18)$$

Substituting (4.17) and (4.18) into the Robertson–Walker metric, we obtain the metric in the system (T, R) to order R^2 :

$$\begin{aligned} dt^2 &= dT^2(1 - \alpha R^2) \\ &\quad - dR^2(1 + \beta R^2) - R^2(d\theta^2 + \sin^2\theta d\phi^2), \end{aligned} \quad (4.19)$$

where the coefficients α and β are given by (4.8). By applying the standard transformations $X = R \sin\theta \cos\phi$, $Y = R \sin\theta \sin\phi$, and $Z = R \cos\theta$, the metric takes a form that is locally Lorentzian:

$$\begin{aligned} d\tau^2 &= dT^2 - dX^2 - dY^2 - dZ^2 \\ &\quad - \alpha dT^2(X^2 + Y^2 + Z^2) - \beta(XdX + YdY + ZdZ)^2. \end{aligned} \quad (4.20)$$

While this metric is locally Lorentzian, it is convenient to make a further transformation:

$$\begin{aligned} \tilde{x}^1 &= X(1 - \beta R^2/4) \\ \tilde{x}^2 &= Y(1 - \beta R^2/4) \end{aligned} \quad (4.21)$$

$$\begin{aligned} \tilde{x}^3 &= Z(1 - \beta R^2/4) \\ \tilde{x}^0 &= T. \end{aligned} \quad (4.22)$$

This transformation yields a metric that is locally Lorentzian and “orthogonal” ($g_{\mu\nu} = 0$):

$$\begin{aligned} d\tau^2 &= d\tilde{x}^0 d\tilde{x}^0 [1 - \alpha R^2] \\ &\quad - [d\tilde{x}^1 d\tilde{x}^1 + d\tilde{x}^2 d\tilde{x}^2 + d\tilde{x}^3 d\tilde{x}^3] \left[1 - \frac{1}{2}\beta R^2\right], \end{aligned} \quad (4.23)$$

i.e. the metric (4.7).

The coordinates \tilde{x} define the freely falling system that we will use for most of this chapter. It is obviously not the only such system. Besides third-order transformations like (4.21), Lorentz transformations and spatial rotations (Exercise 3.2) applied to the coordinates \tilde{x} also yield coordinates with a metric that is locally Lorentzian. The \tilde{x} is, however, special because it is centered on a comoving observer (us). The universe must therefore be isotropic in the system \tilde{x} . Because of this, the

energy-momentum tensor and the Einstein field equation take simple forms, as we will see in the next sections.

4.2 A Universe with $\rho = 0$

Without knowing the Einstein equation, we can guess the Friedmann equation for a universe with $\rho = 0$. In this case, it should be possible to find a globally Lorentzian coordinates system with $\alpha = \beta = 0$:

$$\alpha = \frac{\ddot{a}}{a} = 0 \quad \beta = \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = 0. \quad (4.24)$$

There are two solutions to these equations. The first is $a(t) = a = \text{constant}$ with $k = 0$:

$$d\tau^2 = dt^2 - a^2 [dr^2 + r^2(d\theta^2 + \sin^2 \theta d\phi^2)]. \quad (4.25)$$

This solution gives directly the locally Lorentzian metric (4.7) with $R = ar$ and $T = t$.

The second solution is $a = t$ with $k = -1$, i.e.

$$d\tau^2 = dt^2 - t^2 \left[\frac{dr^2}{1+r^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right]. \quad (4.26)$$

In this case, the transformations (4.14) and (4.13) are equivalent to

$$R = rt, \quad T = t \sqrt{1+r^2} - t_0. \quad (4.27)$$

These transformations yield *exactly* a Lorentzian metric:

$$\begin{aligned} d\tau^2 &= dT^2 - dR^2 - R^2(d\theta^2 + \sin^2 \theta d\phi^2) \\ &= dT^2 - dX^2 - dY^2 - dZ^2. \end{aligned} \quad (4.28)$$

It is perhaps surprising that a universe with a flat space–time can be described by a Robertson–Walker metric with $k = -1$ and $a(t) = t$. To resolve this paradox, we need to remind ourselves of the difference between curved spaces and curved space–times. The curvature of space is seen by making *instantaneous* geometrical measurements, e.g. testing the Euclidean relation $Rd\theta = dS$ as in Fig. 3.7. Since the measurements take place over large distances, we need to choose a network of clocks and decide how to synchronize them. Any conclusions about flatness will necessarily depend on these conventions. As we have seen in the example of a universe with $\rho = 0$, the choice of comoving clocks gives a curved space while the use of laboratory clocks gives a flat space.

We will see in Sect. 4.7 that the curvature of space–time is seen by following the trajectories of nearby test particles as illustrated in Fig. 4.3. The measurement is done in a freely falling coordinate system in which the particles have the same initial velocities. If the separation between the particles is time independent, the space–time is flat.

Unlike measurements of spatial curvature which are done instantaneously over large distances, measurements of space–time curvature are done locally but require a non-vanishing amount of time. They only require one observer with one clock.

4.3 The Energy-Momentum Tensor

In this section we will find the “source” of gravitation, the energy-momentum tensor, $T^{\mu\nu}$. The Einstein field equation is a local equation and we can therefore concern ourselves only with a small region of space near the origin. As such, we can use the freely falling coordinates found in the previous section. These coordinates have a metric that is the metric of flat space–time up to quadratic corrections in the distance from the origin. As it turns out, we will need $T^{\mu\nu}$ only up to linear corrections so our discussion of $T^{\mu\nu}$ will be entirely within the framework of special relativity.

We are looking for the relativistic generalization of Poisson’s equation for the gravitational potential Φ :

$$\nabla^2 \Phi = 4\pi G\rho, \quad (4.29)$$

where the source of gravitation on the right is the mass (or energy) density ρ . To find the relativistic generalization of ρ , we will be guided by the generalization of the Poisson equation for the electrostatic potential Φ_q :

$$\nabla^2 \Phi_q = -4\pi \rho_q \quad \rightarrow \quad \left(\frac{\partial^2}{\partial t^2} - \nabla^2 \right) A^\mu = 4\pi j^\mu. \quad (4.30)$$

In the relativistic generalization, we replaced the potential by the 4-vector potential A^μ and ∇^2 by the invariant operator $(\partial^2/\partial t^2 - \nabla^2)$. The charge density was replaced by the current 4-vector j^μ and this is easy to justify. Consider a set of particles of charge q that are at rest. The charge density is $\rho_q = qn$ where n is the particle number density. If we suppose that ρ_q is the 0th component of a 4-vector, a Lorentz transformation gives

$$\begin{pmatrix} \gamma & \beta\gamma \\ \beta\gamma & \gamma \end{pmatrix} \begin{pmatrix} qn \\ 0 \end{pmatrix} = \begin{pmatrix} q\gamma n \\ q\beta\gamma n \end{pmatrix} = \begin{pmatrix} qn' \\ q\beta n' \end{pmatrix} = \begin{pmatrix} \rho' \\ j'_x \end{pmatrix}, \quad (4.31)$$

where the number density in the primed frame is $n' = \gamma n$ because of the Lorentz contraction. We see that in the primed frame, the 0th component is still the charge density, confirming that ρ_q is, indeed, the 0th component of a 4-vector.

We are tempted to try the same thing for gravitation. However, if we suppose that the energy density is the 0th component of a 4-vector, a Lorentz transformation gives us an object whose 0th component is no longer the energy density. This is simply because the Lorentz transformation modifies by a factor γ both the number density of particles *and* their energies.

The second possibility to come to mind is to suppose that the energy density is the 00 component of a tensor. In this case, a Lorentz transformation returns an object whose 00 component is still the energy density:

$$\begin{pmatrix} \gamma & \beta\gamma \\ \beta\gamma & \gamma \end{pmatrix} \begin{pmatrix} mn & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \gamma & \beta\gamma \\ \beta\gamma & \gamma \end{pmatrix} = \begin{pmatrix} (\gamma m)(\gamma n) & (\beta\gamma m)(\gamma n) \\ (\beta\gamma m)(\gamma n) & (\beta^2\gamma m)(\gamma n) \end{pmatrix} \\ = \begin{pmatrix} E' & p'_x \\ p'_x & (p'_x p'_x/E') \end{pmatrix} n'. \quad (4.32)$$

We can therefore take the energy density ρ to be the 00 component of a tensor, $T^{\mu\nu}$.

The other components of $T^{\mu\nu}$ are perhaps still a little obscure so we perform a rotation about the z axis:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} E & p_x & 0 & 0 \\ p_x & p_x p_x/E & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & \sin\theta & 0 \\ 0 & -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ = \begin{pmatrix} E & p_x & p_y & 0 \\ p_x & p_x p_x/E & p_x p_y/E & 0 \\ p_y & p_y p_x/E & p_y p_y/E & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (4.33)$$

At this point, the reader is certainly convinced that the general components of $T^{\mu\nu}$ for a mono-kinetic set of particles is

$$T^{\mu\nu} = n(\tilde{x}) \frac{p^\mu p^\nu}{E}. \quad (4.34)$$

The behavior of this object under Lorentz transformations and spatial rotations is entirely determined by the quantity $p^\mu p^\nu$ since n/E is invariant, since both n and E are increased by a factor γ over the rest-frame values.

Since the sum of two tensors is a tensor, for a poly-kinetic set of particles the energy-momentum is found by integrating (4.34) over momentum:

$$T^{\mu\nu} = \int \frac{d^3 p}{E} F(\mathbf{p}, \tilde{x}) p^\mu p^\nu = n(\tilde{x}) \left\langle \frac{p^\mu p^\nu}{E} \right\rangle, \quad (4.35)$$

where $F(\mathbf{p}, \tilde{x}) = dN/d^3 p d^3 x$ is the density of particles in phase space and $n(\tilde{x})$ is the density of particles in space. We see that the 00 component is the energy density, $T^{00} = \rho = n\langle E \rangle$, as expected.

We would like to leave open the possibility of an energy density that is not associated with the presence of particles or macroscopic fields. We will call this energy density the vacuum energy density ρ_Λ :

$$T^{00} = \rho_\Lambda + n(\tilde{x}) \langle E \rangle. \quad (4.36)$$

In the absence of particles or classical fields, all Lorentz frames should be equivalent and it is natural to suppose that ρ_Λ is a Lorentz invariant. This requirement determines the form of the vacuum energy-momentum tensor since, as shown in appendix A, the only Lorentz invariant tensor is $\eta^{\mu\nu} = \eta_{\mu\nu}$:

$$T_\Lambda^{\mu\nu} = \rho_\Lambda \eta^{\mu\nu}. \quad (4.37)$$

The final type of energy-momentum tensor interesting for cosmology is the tensor associated with macroscopic fields. We will consider this case in Sect. 4.6.

We will now calculate $T^{\mu\nu}$ in our freely falling coordinate system \tilde{x} . For the Einstein field equation at the origin, we need only $T^{\mu\nu}$ at the origin. At this point, the Hubble velocity is zero so the galaxies (or particles) in a small box will have a momentum distribution that is isotropic, i.e. a distribution dominated by peculiar velocities. It follows that at the origin $\langle p^i \rangle = \langle p^i p^{j \neq i} \rangle = 0$ so the off-diagonal elements of $T^{\mu\nu}$ vanish at the origin

$$T^{\mu \neq \nu}(\tilde{x}^1 = \tilde{x}^2 = \tilde{x}^3 = 0) = 0. \quad (4.38)$$

The 00 component is, as always

$$T^{00}(\tilde{x}^1 = \tilde{x}^2 = \tilde{x}^3 = 0) = \rho, \quad (4.39)$$

where ρ is the energy density. The space-space components at the origin are

$$T^{11}(0) = T^{22}(0) = T^{33}(0) = n(\tilde{x}) \langle |\mathbf{p}|^2 / 3E \rangle \equiv p, \quad (4.40)$$

where this equation defines the “pressure” p . It can be shown (Exercise 4.3) that this definition of the pressure agrees with the usual definition for an ideal gas at temperature T in both the non-relativistic ($T \ll m$) and relativistic ($T \gg m$) limits. However, it is important to remember that in our context, T^{ii} has the role of a source of gravity. In view of our intuitive association of “pressure” with a force per unit area, it would perhaps be better to call p the “density of $|\mathbf{p}|^2 / 3E$.”

We note that the definition (4.35) implies the following relations between energy density and pressure (“equations of state”):

$$|p| \ll \rho \quad \text{non-relativistic particles} \quad (4.41)$$

$$p = \rho/3 \quad \text{relativistic particles.} \quad (4.42)$$

Equation (4.37) gives a curious value of the vacuum pressure:

$$p = -\rho \quad \text{vacuum.} \quad (4.43)$$

The vacuum pressure is negative if the vacuum energy density is positive. From this point of view, the vacuum is like an elastic medium on which work must be performed to *increase* its volume (in contrast to a normal gas on which work must be performed to *decrease* its volume).

Before we derive the field equations, it is useful to study the local conservation of energy (4.2). For a system of particles with $T^{\mu\nu}$ given by (4.35), the 0 component of (4.2) is

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j}_E = 0, \quad (4.44)$$

where the “energy current” takes the expected form

$$\mathbf{j}_E = n \langle E \mathbf{v} \rangle \quad \mathbf{v} = \frac{\mathbf{p}}{E}. \quad (4.45)$$

Equation (4.44) is the energy analog of the equation for local charge conservation $\partial \rho_q / \partial t + \nabla \cdot \mathbf{j}_q = 0$ with $\mathbf{j}_q = q n \langle \mathbf{v} \rangle$. In order to apply it, we need the energy current at $(d\tilde{x}^1, d\tilde{x}^2, d\tilde{x}^3) \neq (0, 0, 0)$ so the derivatives can be evaluated. The situation is illustrated in Fig. 4.2. It is easy to evaluate the energy current for non-relativistic particles where the Hubble velocity dominates away from the origin. In this case, $\mathbf{v} = (\dot{a}/a)\mathbf{R}$ where $\mathbf{R} = (\tilde{x}^1, \tilde{x}^2, \tilde{x}^3)$ giving

$$\mathbf{j}_E = \frac{\dot{a}}{a} \rho \mathbf{R} \quad \Rightarrow \quad \nabla \cdot \mathbf{j}_E = 3 \frac{\dot{a}}{a} \rho. \quad (4.46)$$

The energy conservation equation (4.44) is then

$$\frac{\partial \rho}{\partial T}(\mathbf{R} = 0) = -3 \rho \frac{\dot{a}}{a} \quad \text{non-relativistic particles.} \quad (4.47)$$

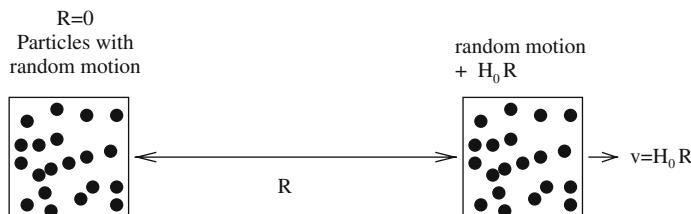


Fig. 4.2 Two groups of particles, one at the origin and a second at a distance R . The calculation of the energy current (4.45) requires knowledge of the velocity distribution as a function of position. The velocity of a particle is the relativistic sum of a random component and a Hubble velocity proportional to the distance from the origin

To include the effect of particle random motions, the easiest way to calculate $T^{\mu\nu}$ away from the origin is to take the $T^{\mu\nu}$ seen by a comoving observer and to Lorentz transform it with $\beta = RH_0 = R/d_H \ll 1$. To order R/d_H we find:

$$\begin{pmatrix} \gamma & \beta\gamma \\ \beta\gamma & \gamma \end{pmatrix} \begin{pmatrix} \rho & 0 \\ 0 & p \end{pmatrix} \begin{pmatrix} \gamma & \beta\gamma \\ \beta\gamma & \gamma \end{pmatrix} = \begin{pmatrix} \rho & \beta(\rho + p) \\ \beta(\rho + p) & p \end{pmatrix}, \quad (4.48)$$

i.e.

$$T^{00}(\tilde{x}) = \rho \quad (4.49)$$

$$T^{ij}(\tilde{x}) = p \delta_{i,j} \quad i, j = 1, 2, 3 \quad (4.50)$$

$$T^{i0}(\tilde{x}) = (\rho + p) H_0 \tilde{x}^i \quad i = 1, 2, 3. \quad (4.51)$$

The energy current is the non-relativistic current (4.46) with ρ replaced by $\rho + p$. It is important to emphasize that ρ and p are the energy density and pressure measured locally by a comoving observer, i.e. a comoving observer using its local Lorentz frame.

We note that $T^{i0}(\tilde{x})$ has the expected form for an energy current, i.e. the mass density times the Hubble velocity in the case of $p = 0$. We also note that the vacuum energy current vanishes as might be expected in view of the fact that the vacuum does not contain moving particles.

Taking the 4-divergence of $T^{\mu\nu}$ given by (4.49), (4.50), and (4.51) we find the equation for the local conservation of energy:

$$\frac{\partial \rho}{\partial t}(\mathbf{R} = 0) = -3(\rho + p) \frac{\dot{a}}{a}, \quad (4.52)$$

where we have used the fact that $\tilde{x}^0 = T = t$ at the origin. The value of ρ at $\mathbf{R} = 0$ is the density viewed by all comoving observers so we can replace the partial derivative by a full derivative. The solutions of this equation for matter, radiation, and vacuum energy are:

$$\rho_M(a) = \rho_M(a_0) \hat{a}^{-3} \quad |p| \ll \rho \quad \text{non-relativistic matter} \quad (4.53)$$

$$\rho_R(a) = \rho_R(a_0) \hat{a}^{-4} \quad p = \rho/3 \quad \text{relativistic matter} \quad (4.54)$$

$$\rho_\Lambda(a) = \rho_\Lambda(a_0) \quad p = -\rho \quad \text{vacuum.} \quad (4.55)$$

These solutions confirm the a dependences found in Chap. 1. Again, the fact that the vacuum energy density is a constant is due to its value being independent of the number of particles present. Because of their random movement, massive particles (galaxies) eventually leave the region around the origin. They are not replaced by other particles because of the Hubble velocity of distant particles. This leaves only the vacuum.

Equation (4.52) is equivalent to

$$\frac{d(\rho a^3)}{da^3} = -p. \quad (4.56)$$

This is, in turn, equivalent to the first law of thermodynamics for an adiabatic expansion:

$$d(\rho a^3) = -p da^3 \quad \Rightarrow \quad dE = -pdV, \quad (4.57)$$

where E is the energy contained in a fixed comoving volume, i.e. a volume bounded by fixed comoving coordinates. We will use this relation in Chap. 6 when we study the thermodynamics of the universe.

We end with a note on energy conservation in cosmology. Local energy conservation is expressed through (4.44) and (4.52). Global energy conservation comes about if we can apply Gauss's theorem to (4.44) by integrating over some finite volume V that contains all the energy:

$$\frac{dE}{dt} = \int dV \frac{\partial \rho}{\partial t} = \int dS \hat{n} \cdot \mathbf{j}_E, \quad (4.58)$$

where the first integral is over the volume and the second over the boundary of the volume. Total energy is only conserved if there is a surface where the energy current vanishes. There are two reasons why this is not the case in cosmology. The first is that if the universe is homogeneous there is no surface where the current vanishes. The second is that (4.44) only applies in Lorentz coordinates. Since it is only possible to find locally Lorentzian coordinates in the presence of gravity, we cannot apply (4.44) even if the energy is concentrated in a finite volume.

There are, of course, quantities that are approximately conserved in cosmology. During the radiation epoch ρa^4 is conserved and during the matter epoch ρa^3 is conserved. This second quantity has dimensions of energy and therefore leads to some confusion because it invites us to interpret it as the total energy of the universe, in spite of the fact that the volume of the universe may be infinite. The Friedmann equation implies that the quantity $\dot{a}^2 - 8\pi G\rho a^2/3$ is time independent. We interpreted this in Sect. 1.2.2 in terms of the conservation of total energy of a single nearby comoving particle.

4.4 The Friedmann Equation

The form of the Einstein tensor $G^{\mu\nu}$ will be determined by the conservation of energy (4.2) which implies that the 4-divergence of $G^{\mu\nu}$ vanishes. It is thus not surprising that we can find directly the Friedmann equation by imposing local energy conservation without bothering to find the general form of $G^{\mu\nu}$. To do this, we suppose that in the field equation for a homogeneous universe the components

of $T^{\mu\nu}$ are linear functions of the parameters of the metric (4.7), $\alpha = \ddot{a}/a$ and $\beta = (\dot{a}/a)^2 + k/a^2$:

$$GT^{00} = G\rho = A\alpha + B\beta \quad (4.59)$$

$$GT^{ii} = Gp = C\alpha + D\beta, \quad (4.60)$$

where A , B , C , and D are constants which, by dimensional analysis, must be pure numbers. Using (4.60) to eliminate α in (4.59) and then differentiating we find

$$\begin{aligned} \dot{\rho} &= (A/C)\dot{p} + \dot{\beta}(BC - AD)/GC \\ &= (A/C)\dot{p} + 2(\dot{a}/a)(\alpha - \beta)(BC - AD)/GC \\ &= \frac{A}{C}\dot{p} - \frac{2}{C}[(D + C)\rho - (B + A)p]\frac{\dot{a}}{a}. \end{aligned} \quad (4.61)$$

Comparing the last form with (4.52), we see that we must impose that the coefficient of \dot{p} vanish, $\Rightarrow A = 0$, if we want (4.52) to be satisfied for all possible relations between ρ and p . Matching the coefficients of ρ and p , we find $C = -2B/3$ and $D = -B/3$:

$$G\rho = B \left[\left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} \right], \quad (4.62)$$

and

$$Gp = -\frac{B}{3} \left[\frac{2\ddot{a}}{a} + \left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} \right]. \quad (4.63)$$

The Einstein field equation and its Newtonian limit (Exercise 4.6) will give us the numerical constant $B = 3/8\pi$. Equations (4.62) and (4.63) then become the field equations for a homogeneous universe:

$$\left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} = \frac{8\pi G\rho}{3} \quad (\text{Friedmann equation}) \quad (4.64)$$

$$\frac{-2\ddot{a}}{a} - \left(\frac{\dot{a}}{a} \right)^2 - \frac{k}{a^2} = 8\pi Gp. \quad (4.65)$$

We summarize the assumptions necessary for this rapid derivation of the Friedmann equation (4.64):

- The universe is homogeneous implying that we can find a coordinate system with the Robertson–Walker metric (4.3).
- The energy-momentum tensor is a linear function of the coefficients of the quadratic terms in the locally Lorentzian metric (4.7) derived from the Robertson–Walker metric, i.e. \ddot{a}/a and $(\dot{a}/a)^2 + k/a^2$.
- Energy is locally conserved (4.52).

The general derivation of the Einstein field equation will eliminate the first assumption, permitting us to treat general energy configurations.

We can combine the 00 and space components of the field equation (4.64) and (4.65) to find an equation for the deceleration of the universe:

$$\frac{\ddot{a}}{a} = \frac{-4\pi G}{3}(\rho + 3p). \quad (4.66)$$

This equation summarizes the gravitational physics determining the expansion. If the universe is dominated by non-relativistic matter, ($p \ll \rho$), the universe is decelerated by the gravitation $\ddot{a}/a = -4\pi G\rho/3$, in agreement with the Newtonian calculation of Chap. 1. In this case, the term $3p$ gives a small “relativistic” correction. For a universe dominated by relativistic matter, ($p = \rho/3$), the deceleration is twice the non-relativistic value. For a universe dominated by vacuum energy, ($p = -\rho$), the deceleration is negative, i.e. the expansion is accelerated.

4.5 The Cosmological Parameters

It is convenient to put the Friedmann equation in a form that contains present-day expansion rate and density parameters. At $t = t_0$ (4.64) takes the form

$$\frac{k}{a_0^2} = \frac{8\pi G\rho(t_0)}{3} - H_0^2 = H_0^2(\Omega_T - 1). \quad (4.67)$$

We see that the sign of k is determined by Ω_T :

$$\begin{array}{ccc} > 1 & & +1 \\ \Omega_T = 1 & & k = 0. \\ < 1 & & -1 \end{array} \quad (4.68)$$

Equation (4.67) gives us the value of a_0 :

$$a_0 = \frac{H_0^{-1}}{\sqrt{|\Omega_T - 1|}} = \frac{d_H}{\sqrt{|\Omega_T - 1|}}. \quad (4.69)$$

Unless Ω_T is very close to unity, a_0 is of order the Hubble distance $d_H = cH_0^{-1}$. We remind the reader that a_0 gives the distance scale where violations of Euclidean geometry become large, e.g. $\Delta\Theta \neq \Delta S/R$. The fact that a_0 diverges for $\Omega_T = 1$ is therefore not surprising since in this case the Euclidean relation is exact. As we will see, the divergence of a_0 will cause no practical problems in calculations of angular or luminosity distances because the quantity $a_0\chi$ will always be well-behaved.

The curvature parameter k can now be eliminated from the Friedmann equation by substituting (4.67) into (4.64):

$$\begin{aligned} \left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G\rho}{3} + H_0^2(1 - \Omega_T)\hat{a}^{-2} \\ &= H_0^2 \left[\frac{\rho(a)}{3H_0^2/8\pi G} + (1 - \Omega_T)\hat{a}^{-2} \right]. \end{aligned} \quad (4.70)$$

To find the solution of the Friedmann equation, we need the energy density as a function of the scale parameter. For non-relativistic matter, radiation, and vacuum energy, we have found previously the results (4.53), (4.54), and (4.55). For a mixture of these three types of energy the Friedmann equation is therefore

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0^2 [\Omega_M \hat{a}^{-3} + \Omega_R \hat{a}^{-4} + \Omega_\Lambda + (1 - \Omega_T) \hat{a}^{-2}], \quad (4.71)$$

where the Ω s all refer to the present-day densities:

$$\Omega_i = \frac{\rho_i(a_0)}{3H_0^2/8\pi G} \quad i = M, R, \Lambda, \quad (4.72)$$

and where

$$\Omega_T = \Omega_M + \Omega_R + \Omega_\Lambda. \quad (4.73)$$

The form (4.71) of the Friedmann equation is correct only if all energy is either relativistic, non-relativistic, or vacuum energy. There are two situations where this is not the case. The first comes about when the temperature is of order of the mass of one of the particle species present. Here, the particles are neither relativistic nor non-relativistic. The technique for treating this case will be presented in Chap. 6. The second case arises when there is an energy associated with a classical field, as in the inflationary epoch. This case will be treated in the next section.

The deceleration parameter is given by (4.66) at t_0 :

$$q_0 \equiv -H_0^{-2} \left[\frac{\ddot{a}}{a} \right]_{t_0} = \frac{\Omega_M}{2} - \Omega_\Lambda. \quad (4.74)$$

As we have already remarked, a positive vacuum energy accelerates the expansion.

Finally, for the present-day conditions (4.67) and (4.74), the metric for the freely falling system (4.19) and (4.8) takes the form

$$\begin{aligned} d\tau^2 &= dT^2 \left(1 + (\Omega_M/2 - \Omega_\Lambda) \left(\frac{R}{d_H} \right)^2 \right) \\ &\quad - dR^2 \left(1 + \Omega_T \left(\frac{R}{d_H} \right)^2 \right) - R^2 (d\theta^2 + \sin^2 \theta d\phi^2). \end{aligned} \quad (4.75)$$

We see that for distances $R \ll d_{\text{H}}$, the metric is well approximated by the Lorentz metric. This justifies our ignoring gravity for purely local effects as we did in Sect. 1.2 when we derived the temperature evolution of the CMB.

4.6 Scalar Fields

Over the last two decades, macroscopic scalar fields have become important objects in speculative cosmology. For the early universe, scalar fields are assumed to dominate the universe during the inflationary epoch. For the present-day universe, they are used in “quintessence” models to provide the dynamical equivalent of a cosmological constant. Scalar fields are preferred over, say, vector fields because they do not give a preferred spatial direction that might lead to a violation of Lorentz invariance.

As with particle and vacuum energy, we will treat scalar fields in a freely falling system within the framework of special relativity. We thus consider a hypothetical field $\phi(\tilde{x})$ that is invariant under Lorentz transformations. A relativistically invariant equation for ϕ is

$$\eta^{\mu\nu} \frac{\partial^2 \phi}{\partial \tilde{x}^\mu \partial \tilde{x}^\nu} + V'(\phi) = \frac{\partial^2 \phi}{\partial T^2} - \nabla^2 \phi + V'(\phi) = 0, \quad (4.76)$$

where $\eta^{\mu\nu} = \eta_{\mu\nu}$ and $V(\phi)$ is the “potential energy” of the field. For $V(\phi) = m^2 \phi^2 / 2$, (4.76) is the well-known Klein–Gordon equation. Lacking constraints from fundamental physics, the potential in inflationary and quintessential models can be chosen to solve some cosmological problem.

The energy-momentum tensor for a scalar field is

$$T^{\mu\nu} = \eta^{\mu\alpha} \eta^{\nu\beta} T_{\alpha\beta}$$

with

$$T_{\mu\nu} = \frac{\partial \phi}{\partial \tilde{x}^\mu} \frac{\partial \phi}{\partial \tilde{x}^\nu} - \frac{1}{2} \eta_{\mu\nu} \left[\left(\frac{\partial \phi}{\partial T} \right)^2 - \nabla \phi \cdot \nabla \phi - 2 V(\phi) \right]. \quad (4.77)$$

The justification for the use of this form for $T_{\mu\nu}$ is nothing more or less than the fact that its 4-divergence vanishes:

$$\frac{\partial T^{\mu\nu}}{\partial \tilde{x}^\mu} = 0, \quad (4.78)$$

as can be readily verified by using (4.76).

In inflationary and quintessential cosmology, we are often interested in potentials that are non-zero and relatively flat, $V(\phi) \neq 0$ and $V'(\phi) \sim 0$. In this case, a solution of (4.76) is:

$$\phi \sim \text{constant}. \quad (4.79)$$

Ignoring all derivatives in $T_{\mu\nu}$ we find

$$T_{00} = \rho \sim V(\phi) \quad (4.80)$$

$$T_{xx} = T_{yy} = T_{zz} = p \sim -V(\phi). \quad (4.81)$$

We see that in this case the energy-momentum tensor is approximately that for the vacuum. It is this characteristic that is exploited in inflationary and quintessential models.

It will be useful to find the form of (4.76) in comoving coordinates (t, r) . Consider a field $\phi(R, T) = \phi(r, t)$. We will be mostly interested in the case where ϕ is spatially homogeneous in comoving coordinates:

$$\frac{\partial \phi}{\partial r} = 0. \quad (4.82)$$

By using the transformations (4.17) and (4.18), it is straightforward to show that

$$\eta^{\mu\nu} \frac{\partial^2 \phi}{\partial \tilde{x}^\mu \partial \tilde{x}^\nu} = \frac{\partial^2 \phi}{\partial T^2} - \frac{1}{R^2} \frac{\partial}{\partial R} R^2 \frac{\partial \phi}{\partial R} = \frac{\partial^2 \phi}{\partial t^2} + 3 \frac{\dot{a}}{a} \frac{\partial \phi}{\partial t}. \quad (4.83)$$

The equation for a spatially homogeneous scalar field is therefore

$$\frac{\partial^2 \phi}{\partial t^2} + 3 \frac{\dot{a}}{a} \frac{\partial \phi}{\partial t} + V'(\phi) = 0. \quad (4.84)$$

We will need this equation because a scalar field does not generally have a simple time dependence for the energy density so there is not a form of the Friedmann equation analogous to (4.71). In situations where the universe is dominated by a spatially homogeneous scalar field, one generally uses (4.84) coupled with the Friedmann equation:

$$\left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi G V(\phi)}{3} + \frac{k}{a^2}. \quad (4.85)$$

Incidentally (4.83) tells us that a field that is time dependent but spatially homogeneous in comoving coordinates must be spatially inhomogeneous in freely falling coordinates. This is because the synchronization of clocks is different in the two systems so a time dependence in one induces a spatial dependence in the other.

A final word on the “reality” of scalar fields. As is the case for electric and magnetic fields, the existence of a cosmological scalar field can be deduced from the movement of test particles. In particular, we will see (4.101) that a homogeneous scalar field with $V > 0$ causes test particles to accelerate away from each other.

4.7 The Riemann Tensor

We now have all the results necessary to do homogeneous cosmology. The impatient reader therefore has our permission to jump ahead to Chap. 5. However, to derive the general field equations and to better understand the physics of gravitation, we should do a little more work.

So far, it appears that the gravitational field is characterized by the 256 components of the tensor $g_{\mu\nu,\alpha\beta}$ that appear in the metric (4.4) of a freely falling system. Not all of the components are independent and we obviously have an enormous freedom to change them by making arbitrary coordinate transformations. For our local metric in a homogeneous universe, we have only two quantities, the coefficients α and β of (4.7).

In order to understand the physical significance of the $g_{\mu\nu,\alpha\beta}$, we must see their effect on particle movement. In Chap. 3 we saw that free particles follow trajectories that maximize their proper time, i.e. space–time geodesics. The equation for geodesics was found to be

$$\frac{d}{d\tau} \left(g_{\mu\nu}(x) \frac{dx^\nu}{d\tau} \right) - \frac{1}{2} \frac{\partial g_{\lambda\nu}}{\partial x^\mu} \frac{dx^\lambda}{d\tau} \frac{dx^\nu}{d\tau} = 0 \quad \mu = 0, 1, 2, 3. \quad (4.86)$$

Applied to the metric (4.4), the geodesic equation is, to first order in \tilde{x} ,

$$\frac{d^2 \tilde{x}^\alpha}{d\tau^2} = -\eta^{\alpha\gamma} [g_{\mu\gamma,\nu\beta} - (1/2)g_{\mu\nu,\gamma\beta}] \tilde{x}^\beta \frac{d\tilde{x}^\mu}{d\tau} \frac{d\tilde{x}^\nu}{d\tau}. \quad (4.87)$$

It is interesting to consider the movement of a slow particle, i.e. a trajectory with $d\tilde{x}^0/d\tau \sim 1$, $d\tilde{x}^i/d\tau \sim 0$, $i = 1, 3$. In this case (4.87) becomes

$$\frac{d^2 \tilde{x}^i}{d\tau^2} = -(1/2) g_{00,i\mu} \tilde{x}^\mu \quad i = 1, 2, 3, \quad (4.88)$$

where we have supposed that the metric is orthogonal $g_{\mu\neq\nu} = 0$. For the metric (4.7) this becomes

$$\frac{d^2 \tilde{x}^i}{d\tau^2} = \alpha \tilde{x}^i = \left[\frac{\ddot{a}}{a} \right]_{t_0} \tilde{x}^i \quad i = 1, 2, 3. \quad (4.89)$$

While the interpretation of (4.89) is relatively simple, one must be cautious about interpreting the acceleration of a particle with respect to a coordinate grid that is a priori arbitrary. The only physically significant acceleration is acceleration of particles with respect to other particles. For example, in electrodynamics the acceleration of charged particles with respect to neutral particles signals the presence of an electromagnetic field. In gravitation, all particles at the same position have the same acceleration because of the principle of equivalence. The presence of gravitational effects can only be signaled by the relative acceleration of test particles separated

by small but non-vanishing distances. Such accelerations exist because of the tidal effect in inhomogeneous gravitational fields.

We therefore compare the movement of the two test particles ($m \rightarrow 0$) near $\tilde{x} = 0$ with initial ($\tilde{x}^0 = 0$) spatial coordinates (Fig. 4.3)

$$\tilde{x}_a^i(t = \tau_a = 0) = -\epsilon^i/2 \quad (\text{particle a}) \quad (4.90)$$

$$\tilde{x}_b^i(t = \tau_b = 0) = +\epsilon^i/2 \quad (\text{particle b}) \quad (4.91)$$

for $i = 1, 2, 3$. As indicated, we start their respective clocks measuring proper time at $t = 0$. We suppose that the two particles start with the same velocity:

$$\frac{d\tilde{x}_a^\mu}{d\tau_a}(t = 0) = \frac{d\tilde{x}_b^\mu}{d\tau_b}(t = 0) \equiv v^\mu. \quad (4.92)$$

An appropriate measure of the separation of the two particles is the space-like analog of proper time, the proper distance :

$$\delta^2(t) = -[\eta_{\mu\nu}(\tilde{x}_a^\mu - \tilde{x}_b^\mu)(\tilde{x}_a^\nu - \tilde{x}_b^\nu)]_{\tilde{x}_a^0 = \tilde{x}_b^0 = t}. \quad (4.93)$$

As illustrated in Fig. 4.3, this quantity can be measured by the observer at the origin with a radar and clock.

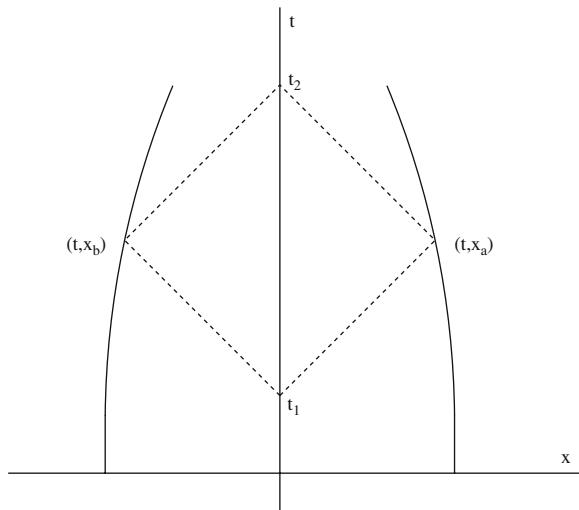


Fig. 4.3 The trajectories of two test particles $x_a(t)$ and $x_b(t)$. The particles start at $t = 0$ with the same velocities, in this case $dx/dt = 0$. The separation of the two particles is monitored by an observer at the origin equipped with a clock and a radar. The square of the space-like separation $\delta^2(t) = -\eta_{\mu\nu}(x_a^\mu - x_b^\mu)(x_a^\nu - x_b^\nu)$ is equal to the square of the round-trip photon flight time, $(t_2 - t_1)^2$. If there is a gravitational field, the test particles will have a relative acceleration and δ^2 will change with time

Because of the symmetry of the situation, we can use either the coordinate time t or the common proper time τ of the two particles to parameterize the trajectories. At $\tau = 0$ we have

$$\delta^2(\tau = 0) = -[\eta_{\mu\nu}(\tilde{x}_a^\mu - \tilde{x}_b^\mu)(\tilde{x}_a^\nu - \tilde{x}_b^\nu)]_{\tau=0} = -\eta_{\mu\nu}\epsilon^\mu\epsilon^\nu. \quad (4.94)$$

After a small proper time τ the coordinates are

$$\tilde{x}^\mu(\tau) = \tilde{x}^\mu(\tau = 0) + \tau \frac{d\tilde{x}^\mu}{d\tau}(0) + \frac{1}{2}\tau^2 \frac{d^2\tilde{x}^\mu}{d\tau^2}(0), \quad (4.95)$$

where $d^2\tilde{x}^\mu/d\tau^2$ is given by (4.87). Using the initial positions and velocities of particles a and b, we find that after a proper time τ , the invariant separation is

$$\delta^2(\tau) = \delta^2(0) - \tau^2 R_{\alpha\mu\beta\nu}(\tilde{x} = 0) \epsilon^\alpha v^\mu \epsilon^\beta v^\nu, \quad (4.96)$$

where $R_{\kappa\lambda\mu\nu}$ is the Riemann curvature tensor:

$$R_{\kappa\lambda\mu\nu}(\tilde{x} = 0) = \frac{1}{2} [g_{\kappa\mu,\lambda\nu} - g_{\lambda\mu,\kappa\nu} - g_{\kappa\nu,\lambda\mu} + g_{\lambda\nu,\kappa\mu}]_{\tilde{x}=0}. \quad (4.97)$$

In the absence of gravitation (flat space-time), $g_{\mu\nu} = \eta_{\mu\nu} \Rightarrow g_{\mu\nu,\alpha\beta} = 0$, and the Riemann tensor vanishes. It follows that in this case the separation of the two particles is time independent.

The non-vanishing relative acceleration of two nearby test particles is called “geodesic deviation.” As we have already emphasized, it is the true manifestation of a gravitational field. For instance, two test particles dropped in the Earth’s gravitational field (Fig. 3.4) will fall toward the Earth’s center. A freely falling observer will see the two particles accelerate toward each other. This convergence is governed by Earth’s Riemann tensor. In Exercise 4.6, we will use Newtonian dynamics to calculate the components $R_{\mu 0\nu 0}$.

We will now apply these results to the case of our cosmological freely falling system with metric (4.7). We consider two particles that are initially at rest ($v^0 = 1$, $v^i = 0$) as in Fig. 4.3. Particle a is initially at $(\tilde{x}^1 = \delta/2, \tilde{x}^2 = \tilde{x}^3 = 0)$ and particle b is initially at $(\tilde{x}^1 = -\delta/2, \tilde{x}^2 = \tilde{x}^3 = 0)$. Their initial separation is $\eta_{\mu\nu}\epsilon^\mu\epsilon^\nu = \delta^2$. Equation (4.96) becomes

$$\frac{d^2\delta^2}{d\tau^2} \sim 2\delta \frac{d^2\delta}{d\tau^2} = -2R_{1010}\delta^2, \quad (4.98)$$

where we have used $d\delta/d\tau = 0$ à $\tau = 0$.

The Riemann tensor is easily calculated for the metric (4.7):

$$R_{1010} = -\alpha = -\frac{\ddot{a}}{a}. \quad (4.99)$$

We will soon see that the Riemann tensor is determined by the energy-momentum tensor via the Einstein equation. In the case of a homogeneous universe we have already found (4.66):

$$\frac{\ddot{a}}{a} = -\frac{4\pi G(\rho + 3p)}{3}, \quad (4.100)$$

implying

$$\frac{d^2\delta}{d\tau^2} = -\frac{4\pi G}{3}(\rho + 3p)\delta. \quad (4.101)$$

Supposing that the pressure is negligible compared to the energy density, i.e. a universe dominated by non-relativistic matter, and taking the radial coordinate of a particle to be $R = \delta/2$ we find

$$\frac{d^2R}{d\tau^2} = -\frac{GM(R)}{R^2}, \quad (4.102)$$

where $M(R = \delta/2)$ is the mass inside an imaginary sphere of radius R centered on the origin. This acceleration is the same that we calculated in Chap. 1 using a Newtonian argument. This explains why the Newtonian argument gave the correct Friedmann equation.

4.8 The Einstein Tensor

The coefficients of the quadratic terms in the locally Lorentzian metric (4.4) or (4.7) are determined by the energy density and the pressure via the Einstein equation:

$$G^{\mu\nu}(\tilde{x} = 0) = -8\pi G T^{\mu\nu}(\tilde{x} = 0), \quad (4.103)$$

where $T^{\mu\nu}$ is the energy-momentum tensor (Sect. 4.3) and where $G^{\mu\nu}$ is the Einstein tensor. Since the physical effects of gravity are determined by the Riemann tensor, we can expect that the components of $G^{\mu\nu}$ will be some linear combination of the $R_{\alpha\beta\gamma\delta}$. The local conservation of energy requires that the 4-divergence of $G^{\mu\nu}$ vanish:

$$\left[\frac{\partial G^{\mu\nu}}{\partial \tilde{x}^\mu} \right]_{\tilde{x}=0} = 0. \quad (4.104)$$

It turns out (Exercise 4.4) that there is only one combination of the $R_{\mu\nu\alpha\beta}$ that has this property:

$$G^{\mu\nu}(\tilde{x} = 0) = \eta^{\mu\alpha}\eta^{\nu\beta}G_{\alpha\beta}(\tilde{x} = 0),$$

with

$$G_{\mu\nu}(\tilde{x} = 0) = \eta^{\lambda\kappa} \left(R_{\lambda\mu\kappa\nu}(0) - \frac{1}{2} \eta_{\mu\nu} \eta^{\alpha\beta} R_{\lambda\alpha\kappa\beta}(0) \right). \quad (4.105)$$

Apart from a numerical factor, we see that the conservation of energy determines the form of the Einstein equation (4.103). The factor of proportionality $8\pi G$ is determined by examining the Newtonian limit (Exercise 4.6).

The form of the Einstein tensor is relatively simple for an orthogonal coordinate system ($g_{\kappa\neq\lambda} = 0$):

$$\begin{aligned} G_{\mu\mu}(\tilde{x} = 0) &= (1/2) \sum_{\lambda \neq \mu} \eta^{\lambda\lambda} \left(\frac{\partial^2 g_{\lambda\lambda}}{\partial x^\mu \partial x^\mu} + \frac{\partial^2 g_{\mu\mu}}{\partial x^\lambda \partial x^\lambda} \right) \\ &\quad - (1/2) \eta_{\mu\mu} \sum_{\lambda} \sum_{\alpha \neq \lambda} \eta^{\lambda\lambda} \eta^{\alpha\alpha} \frac{\partial^2 g_{\lambda\lambda}}{\partial x^\alpha \partial x^\alpha}. \end{aligned} \quad (4.106)$$

It is thus simple to calculate the Einstein tensor for the metric (4.7) of our freely falling system: $G_{\mu\neq\nu}(0) = 0$ and

$$G_{00}(0) = -3\beta = -3 \left[\left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} \right]_{t_0} \quad (4.107)$$

$$G_{11}(0) = G_{22}(0) = G_{33}(0) = 2\alpha + \beta = \left[2 \frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} \right]_{t_0}. \quad (4.108)$$

Substituting these functions into (4.103), we find the Friedmann equation (4.64) and its partner (4.65).

4.9 The General Einstein Equation

In Sect. 4.8, we found the Einstein tensor (4.105) at the origin of a freely falling coordinate system that can be used in the Einstein field equation (4.103). Since the principle of equivalence assures us that it is always possible to find a freely falling coordinate system, (4.103) contains all the physics of gravitation. This is illustrated by the fact that we used these equations to find the Friedmann equation which applies everywhere in a homogeneous universe, not just at the origin.

However, in order to apply (4.103), we first needed to transform comoving coordinates into freely falling coordinates. While this was a relatively simple procedure, it is useful to have the generalization of (4.103) that applies to a general coordinate system.

To do this, we consider coordinates $(x^\mu, \mu = 0, 1, 2, 3)$ that are functions $x^\mu(\tilde{x}^\nu)$ of the freely falling coordinates \tilde{x}^ν . The metric $g_{\alpha\beta}$ for the coordinates x^μ is easily calculated:

$$\begin{aligned} d\tau^2 &= g_{\mu\nu}(x) dx^\mu dx^\nu = g_{\alpha\beta}(\tilde{x}) d\tilde{x}^\alpha d\tilde{x}^\beta \\ &= g_{\alpha\beta}(\tilde{x}) \frac{\partial \tilde{x}^\alpha}{\partial x^\mu} \frac{\partial \tilde{x}^\beta}{\partial x^\nu} dx^\mu dx^\nu, \end{aligned} \quad (4.109)$$

i.e.

$$g_{\mu\nu}(x) = \frac{\partial \tilde{x}^\alpha}{\partial x^\mu} \frac{\partial \tilde{x}^\beta}{\partial x^\nu} g_{\alpha\beta}(\tilde{x}), \quad (4.110)$$

or

$$g_{\alpha\beta}(\tilde{x}) = \frac{\partial x^\mu}{\partial \tilde{x}^\alpha} \frac{\partial x^\nu}{\partial \tilde{x}^\beta} g_{\mu\nu}(x). \quad (4.111)$$

It will be useful to define the inverse metric $g^{\alpha\mu}$:

$$g^{\alpha\mu} g_{\mu\beta} = \delta^\alpha_\beta, \quad (4.112)$$

where $\delta^\alpha_\beta = 1$ if $\alpha = \beta$ and $\delta^\alpha_\beta = 0$ if $\alpha \neq \beta$. The inverse metrics of the two systems are related by

$$g^{\mu\nu}(x) = \frac{\partial x^\mu}{\partial \tilde{x}^\alpha} \frac{\partial x^\nu}{\partial \tilde{x}^\beta} g^{\alpha\beta}(\tilde{x}), \quad (4.113)$$

or

$$g^{\alpha\beta}(\tilde{x}) = \frac{\partial \tilde{x}^\alpha}{\partial x^\mu} \frac{\partial \tilde{x}^\beta}{\partial x^\nu} g^{\mu\nu}(x). \quad (4.114)$$

This equation can be verified by multiplying it by (4.111).

Equation (4.103) contains the derivatives of the metric $g_{\mu\nu}(\tilde{x})$ of the system \tilde{x} . We would like to transform (4.103) so that it contains the derivatives of the metric $g_{\mu\nu}(x)$ of the system x . Finding inspiration in the form of Lorentz transformations, we *define* the tensors $G_{\mu\nu}$ and $T_{\mu\nu}$ in the system x by

$$G_{\mu\nu}(x) \equiv \frac{\partial \tilde{x}^\alpha}{\partial x^\mu} \frac{\partial \tilde{x}^\beta}{\partial x^\nu} G_{\alpha\beta}(\tilde{x} = 0) \quad x = x(\tilde{x} = 0) \quad (4.115)$$

$$T_{\mu\nu}(x) \equiv \frac{\partial \tilde{x}^\alpha}{\partial x^\nu} \frac{\partial \tilde{x}^\beta}{\partial x^\mu} T_{\alpha\beta}(\tilde{x} = 0) \quad x = x(\tilde{x} = 0). \quad (4.116)$$

Applying this transformation to the two sides of (4.103) we find *trivially* that

$$G_{\mu\nu}(x) = 8\pi G T_{\mu\nu}(x) \quad x = x(\tilde{x} = 0). \quad (4.117)$$

An explicit form for $G_{\mu\nu}(x)$ can be found by substituting (4.105) into (4.115) and then using (4.114):

$$G_{\mu\nu}(x) = g^{\lambda\kappa} \left(R_{\lambda\mu\kappa\nu}(x) - \frac{1}{2} g_{\mu\nu} g^{\alpha\beta} R_{\lambda\alpha\kappa\beta}(x) \right), \quad (4.118)$$

where

$$R_{\kappa\lambda\mu\nu}(x) \equiv \frac{\partial \tilde{x}^\alpha}{\partial x^\kappa} \frac{\partial \tilde{x}^\beta}{\partial x^\lambda} \frac{\partial \tilde{x}^\gamma}{\partial x^\mu} \frac{\partial \tilde{x}^\delta}{\partial x^\nu} R_{\alpha\beta\gamma\delta}(\tilde{x} = 0) \quad x = x(\tilde{x} = 0). \quad (4.119)$$

By substituting (4.97) into (4.119) and using (4.111) we find (after a rather long calculation)

$$\begin{aligned} R_{\kappa\lambda\mu\nu}(x) &= \frac{1}{2} (g_{\kappa\mu,\lambda\nu} - g_{\lambda\mu,\kappa\nu} - g_{\kappa\nu,\lambda\mu} + g_{\lambda\nu,\kappa\mu}) \\ &\quad + g_{\eta\sigma} (\Gamma_{\kappa\mu}^\eta \Gamma_{\lambda\nu}^\sigma - \Gamma_{\kappa\nu}^\eta \Gamma_{\lambda\mu}^\sigma), \end{aligned} \quad (4.120)$$

where the “affine connection” Γ is defined by

$$\Gamma_{\kappa\mu}^\eta \equiv (1/2) g^{\eta\nu} \left(\frac{\partial g_{\mu\nu}}{\partial x^\kappa} + \frac{\partial g_{\kappa\nu}}{\partial x^\mu} - \frac{\partial g_{\kappa\mu}}{\partial x^\nu} \right). \quad (4.121)$$

We note that to find (4.120) it is perhaps simpler to go backward by showing

$$R_{\alpha\beta\gamma\delta}(\tilde{x} = 0) = \frac{\partial x^\kappa}{\partial \tilde{x}^\alpha} \frac{\partial x^\lambda}{\partial \tilde{x}^\beta} \frac{\partial x^\mu}{\partial \tilde{x}^\gamma} \frac{\partial x^\nu}{\partial \tilde{x}^\delta} R_{\alpha\beta\gamma\delta}(x) \quad x = x(\tilde{x} = 0), \quad (4.122)$$

where $R_{\alpha\beta\gamma\delta}(\tilde{x} = 0)$ is given by (4.97) and where $R_{\alpha\beta\gamma\delta}(x)$ is given by (4.120).

Equation (4.120) is the general form of the Riemann tensor. The terms containing the affine connection were not present in the form (4.97) because the first derivatives of the metric vanish at the origin of a freely falling coordinate system.

Equations (4.118), (4.120), and (4.121) determine the left-hand side of the Einstein equation (4.117) without explicitly referring to the freely falling system.

To find the right-hand side (4.116) of (4.117), it is simple to define $T^{\mu\nu}$ by

$$T^{\mu\nu}(x) \equiv \frac{\partial x^\mu}{\partial \tilde{x}^\alpha} \frac{\partial x^\nu}{\partial \tilde{x}^\beta} T^{\alpha\beta}(\tilde{x} = 0) \quad x = x(\tilde{x} = 0). \quad (4.123)$$

It is simple to show that

$$T_{\mu\nu}(x) = g_{\mu\alpha}(x) g_{\nu\beta}(x) T^{\alpha\beta}(x). \quad (4.124)$$

Using $T_{00} = \rho$ and $T_{ii} = p$, we find

$$\begin{aligned} T^{\mu\nu}(x) &= \frac{\partial x^\mu}{\partial \tilde{x}^0} \frac{\partial x^\nu}{\partial \tilde{x}^0} \rho + \frac{\partial x^\mu}{\partial \tilde{x}^i} \frac{\partial x^\nu}{\partial \tilde{x}^i} p \\ &= U^\mu U^\nu \rho + (U^\mu U^\nu - g^{\mu\nu}) p, \end{aligned} \quad (4.125)$$

where

$$U^\mu = \frac{\partial x^\mu}{\partial \tilde{x}^0} \quad (4.126)$$

is the 4-velocity of x with respect to \tilde{x} . Since the mean particle momentum vanishes in the system \tilde{x} , U^μ is also the mean 4-velocity of the particles in the system x .

The Einstein equation (4.117) with (4.118), (4.120), and (4.121) on the left and (4.125), (4.126), and (4.124) on the right are the much-awaited generalization of (4.103).

To check this result, we will apply (4.117) to the Robertson–Walker metric:

$$g_{tt} = \frac{1}{g^{tt}} = 1 \quad (4.127)$$

$$g_{rr} = \frac{1}{g^{rr}} = \frac{-a^2}{1 - kr^2} \quad (4.128)$$

$$g_{\theta\theta} = \frac{1}{g^{\theta\theta}} = -a^2 r^2 \quad (4.129)$$

$$g_{\phi\phi} = \frac{1}{g^{\phi\phi}} = -a^2 r^2 \sin^2 \theta. \quad (4.130)$$

The Einstein tensor is

$$G_{\mu\nu} = 0 \quad \text{if } \mu \neq \nu \quad (4.131)$$

$$G_{00} = -3 \left[\left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} \right] \quad (4.132)$$

$$G_{ii} = \left[2 \frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} \right] g_{ii} \quad i = r, \theta, \phi. \quad (4.133)$$

In comoving coordinates, the mean particle velocity vanishes ($U^0 = 1$, $U^i = 0$). Equation (4.125) for the energy-momentum tensor along with (4.124) gives $T_{\mu\neq\nu} = 0$ and

$$T_{00} = \rho \quad (4.134)$$

$$T_{ii} = -p g_{ii} \quad i = r, \theta, \phi. \quad (4.135)$$

By combining these expressions for the Einstein tensor with the expressions for the energy-momentum tensor, we rediscover (4.64) and (4.65).

We remark with satisfaction that, because of a great number of cancellations, the final equations do not depend on the spatial coordinates (r, θ, ϕ) . This confirms the fact that the Robertson–Walker metric can apply to a homogeneous universe. If we had replaced the factor of $1/(1 - kr^2)$ in the metric by a general factor $f(r)$, the form of the Einstein equations would have been more complicated but would

have had $f(r) = 1/(1 - kr^2)$ as a solution. The arguments of Chap. 3 used to find the Robertson–Walker metric were therefore not necessary because the form of the metric would have been in any case forced on us by the Einstein equation.

We finish this chapter with a few comments on the “generality” of general relativity. Much is made of the fact that this theory treats arbitrary coordinate systems. The manipulations that we have just performed show that there is nothing particularly profound about this. We could just as well have started with the metric of special relativity, $\eta_{\mu\nu}$, and then performed arbitrary coordinate transformations to put the metric in as strange a form as we like. The important thing is that in this chapter we did not start with $\eta_{\mu\nu}$ but rather with a locally Lorentzian metric constrained by the Einstein equation for a freely falling system. The essential physics is present in the equivalence principle that assures us that we can find such freely falling coordinates, and in local energy conservation that constrains the metric of such coordinates.

Exercises

4.1 An interval of time dt in our freely falling system is, according to (4.13) given by

$$dT = dt + a\dot{a}rdr + (1/2)\dot{a}^2r^2dt + (1/2)a\ddot{a}r^2dt. \quad (4.136)$$

- (a) Show that if $\ddot{a} = 0$ (i.e., in the absence of gravitation) the relation between dT and dt is simply due to special relativistic time dilation between fixed and (co)moving clocks.
- (b) Verify (4.14) and (4.13).

4.2 Verify that $G^{\mu\nu}$ given by (4.105) and (4.97) satisfies

$$\frac{\partial G^{\mu\nu}}{\partial x^\mu} = 0. \quad (4.137)$$

4.3 Consider the energy-momentum tensor (4.35) for free particles.

- (a) Show that the pressure p defined by (4.40) agrees with the usual relations for an ideal gas of temperature T : $p = nT$ for a non-relativistic gas and $p = \rho/3$ for a relativistic gas.
- (b) Use Liouville’s equation to show that $T^{\mu\nu}$ given by (4.35) satisfies

$$\frac{\partial T^{\mu\nu}}{\partial x^\mu} = 0. \quad (4.138)$$

4.4 Show that the Riemann tensor (4.97) and the Einstein tensor (4.105) are Lorentz tensors. This shows that the Einstein equation (4.103) is Lorentz invariant.

4.5 Show that the Einstein equation is equivalent to

$$R_{\mu\nu} = -8\pi G (T_{\mu\nu} - (1/2)g_{\mu\nu}g^{\alpha\beta}T_{\alpha\beta}), \quad (4.139)$$

where $R_{\mu\nu}$ is the Ricci tensor

$$R_{\mu\nu} = g^{\alpha\beta} R_{\alpha\mu\beta\nu}. \quad (4.140)$$

4.6 We want to consider the Newtonian limit of the Einstein field equation in order to justify the factor of 8π .

- (a) Consider two test particles having only gravitational interactions that are placed inside a sphere of uniform density ρ and pressure $p \ll \rho$. Use Newtonian gravitation to calculate the relative acceleration of the two particles initially at rest and separated by a δ . Using (4.98), show that in a coordinate system freely falling with the two particles, the components R_{i0i0} of the Riemann tensor are

$$R_{i0i0} = \frac{4\pi G\rho}{3} \quad i = 1, 2, 3. \quad (4.141)$$

It follows that

$$R_{00} = -4\pi G\rho. \quad (4.142)$$

This is consistent with (4.139), thus justifying the factor of 8π in the Einstein equation.

- (b) Use Newtonian gravity to calculate the relative acceleration of the four particles in Fig. 3.4. Find the components R_{0i0i} of the Riemann tensor in the freely falling frame and show that

$$R_{00} = 0. \quad (4.143)$$

This is consistent with (4.139) in vacuum.

4.7 The “instantaneous velocity” of a galaxy at (t, χ) is

$$\frac{d(a\chi)}{dt} = (a\chi) \frac{\dot{a}}{a} = \frac{a\chi}{d_H}. \quad (4.144)$$

This velocity is greater than $c = 1$ for $a\chi$ greater than the Hubble distance. This is not in contradiction with special relativity which only applies to Lorentzian coordinate systems. In our freely falling system \tilde{x} , galactic velocities are always subluminal but, at any rate, the system only applies to $R \sim a\chi \ll d_H$.

Consider a universe with $\rho = 0$. This universe can have a globally Lorentzian system of coordinates defined by (4.27)

$$R = rt \quad , \quad T = t \sqrt{1 + r^2} - t_0. \quad (4.145)$$

Consider a test particle at fixed r . Show that in the system (T, R) velocities are always subluminal, $dR/dT < 1$.

4.8 The Einstein field equations for the metric (3.89) are easily found if $|\Phi| \ll 1$ because in this case we can ignore the terms proportional to the affine connection in (4.120) which are second order in Φ . Show that for vanishing pressure, the field equations yield the Poisson equation for the gravitational potential Φ .

Chapter 5

Friedmannology

The time evolution of the scale factor $a(t)$ is determined by two coupled equations. The first is the Friedmann equation:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho(a)}{3} + H_0^2(1 - \Omega_T)\hat{a}^{-2}, \quad (5.1)$$

where

$$\hat{a}(t) = \frac{a(t)}{a_0} \quad \text{and} \quad \Omega_T = \frac{\rho(a_0)}{3H_0^2/8\pi G}. \quad (5.2)$$

The second equation is the “equation of state” giving the pressure, p , as a function of the energy density ρ . For a mixture of non-relativistic matter ($p \sim 0$), relativistic matter ($p_{\text{rad}} = \rho_{\text{rad}}/3$), and dark energy, the pressure is

$$p = \frac{\rho_{\text{rad}}}{3} + w\rho_{\text{de}}, \quad (5.3)$$

where $w \equiv p_{\text{de}}/\rho_{\text{de}}$. If the dark energy is a time-independent vacuum energy we have simply $w = -1$. If it is due to a scalar field, w can be a function of the value of the field and, therefore, of $a(t)$.

The equation of state then determines, via the law of local energy conservation (4.52), the dependence of ρ on the scale factor and this dependence can be imposed on the Friedmann equation. In a Λ CDM universe, the Friedmann equation is of the form:

$$\frac{\dot{a}}{a} = H_0 \left(\Omega_R \hat{a}^{-4} + \Omega_M \hat{a}^{-3} + \Omega_\Lambda + (1 - \Omega_T) \hat{a}^{-2} \right)^{1/2}, \quad (5.4)$$

where Ω_R , Ω_M , and Ω_Λ are the present-day contributions of relativistic matter, non-relativistic matter, and vacuum energy:

$$\Omega_T = \Omega_R + \Omega_M + \Omega_\Lambda. \quad (5.5)$$

For the more general case of a dark energy with $w \neq -1$, the Friedmann equation is

$$\frac{\dot{a}}{a} = H_0 (\Omega_R \hat{a}^{-4} + \Omega_M \hat{a}^{-3} + \Omega_{de} \hat{a}^{-3(w+1)} + (1 - \Omega_T) \hat{a}^{-2})^{1/2}. \quad (5.6)$$

We emphasize the Ω_R term in (5.4) and (5.6) is “fragile” because particles that are not relativistic today were so in the past. It is thus necessary to be careful in calculations where this term is important, e.g. Exercise 5.2.

The Friedmann equation (5.4) allows us to calculate most of the interesting quantities for a homogeneous universe. For example, the time as a function of the scale factor is:

$$\begin{aligned} t - t_1 &= \int_{a_1}^{a(t)} \frac{da}{a(\dot{a}/a)} \\ &= H_0^{-1} \int_{a_1}^{a(t)} \frac{da}{a(\Omega_R \hat{a}^{-4} + \Omega_M \hat{a}^{-3} + \Omega_\Lambda + (1 - \Omega_T) \hat{a}^{-2})^{1/2}}. \end{aligned} \quad (5.7)$$

The radial coordinate χ of an object of redshift z is

$$\begin{aligned} \chi(z) &= \int_{a_0/(1+z)}^{a_0} \frac{da}{a^2(\dot{a}/a)} \\ &= \frac{H_0^{-1}}{a_0} \int_{(1+z)^{-1}}^1 \frac{d\hat{a}}{\hat{a}^2(\Omega_R \hat{a}^{-4} + \Omega_M \hat{a}^{-3} + \Omega_\Lambda + (1 - \Omega_T) \hat{a}^{-2})^{1/2}}. \\ &= \frac{H_0^{-1}}{a_0} \int_0^z \frac{dz'}{(\Omega_R \hat{a}^{-4} + \Omega_M \hat{a}^{-3} + \Omega_\Lambda + (1 - \Omega_T) \hat{a}^{-2})^{1/2}}. \end{aligned} \quad (5.8)$$

Knowing $\chi(z)$, we can calculate the luminosity and angular distances:

$$d_L(z) = a_0 r(z)(1+z) \quad (5.9)$$

$$d_A(z) = a_0 r(z)/(1+z), \quad (5.10)$$

where the radial coordinate r is

$$\begin{aligned} \sin \chi &= \chi - \chi^3/6 + \dots & \Omega_T > 1 \\ r &= \chi & \Omega_T = 1. \\ \sinh \chi &= \chi + \chi^3/6 + \dots & \Omega_T < 1 \end{aligned} \quad (5.11)$$

The integrals (5.7) and (5.8) are functions of the cosmological parameters (H_0 , Ω_R , Ω_M , Ω_Λ) and in this chapter we will show how this fact can be used to measure their values. In Sect. 5.1, we will calculate what is commonly called “the age of the universe” by setting the lower limit of integration in (5.7) to $a_1 = 0$, i.e. the moment of the “big bang.” The integral will be a function of (H_0 , Ω_M , Ω_Λ) and a constraint on these three parameters will be obtained by requiring that the calculated age be greater than the age of the oldest known objects in the universe.

In Sect. 5.2 we will calculate the integral (5.8) and use it to calculate the flux of type Ia Supernovae at high redshift. Comparison with the data will indicate that $\Omega_A > \Omega_M$ so that the universe is presently accelerating. In Sect. 5.3, we will calculate the “sound horizon at recombination.” This length is imprinted on the spatial distribution of galaxies and on the angular distribution of the CMB temperature. Comparison with the data will allow us to measure Ω_M and Ω_T .

In the next sections, we will discuss two problems of classical cosmology, the “horizon” problem and the “ Ω ” problem. We will then show how these two problems can be, in some sense, solved by inflationary models of the early universe.

We will end the chapter with a discussion of absorption and scattering of photons in intergalactic space. This will allow us to understand how measurements of the Compton scattering of CMB photons have shown that the universe was reionized near a redshift of $z \sim 10$.

5.1 The Age of the Universe

The elapsed time since the initial singularity ($a = 0$) can be calculated from (5.7). The duration of the radiation epoch is negligible (Exercise 5.2), so we can ignore the relativistic term in the integrand of (5.7). In this approximation, the calculated age is H_0^{-1} times a function of (Ω_M, Ω_A) :

$$t_0 = \int_0^{a_0} \frac{da}{\dot{a}} \sim H_0^{-1} \int_0^1 \frac{d\hat{a}}{\hat{a} (\Omega_M \hat{a}^{-3} + \Omega_A + (1 - \Omega_M - \Omega_A) \hat{a}^{-2})^{1/2}} . \quad (5.12)$$

Unless the integrand is badly behaved, the order of magnitude of the age is the Hubble time $t_H = H_0^{-1}$. The age can be easily calculated for two cases:

$$(\Omega_M, \Omega_A) = (0, 0) \Rightarrow t_0 = t_H , \quad (5.13)$$

$$(\Omega_M, \Omega_A) = (1, 0) \Rightarrow t_0 = (2/3)t_H . \quad (5.14)$$

For the currently favored values, $(\Omega_M, \Omega_A) = (0.27, 0.73)$ [1], we find numerically

$$t_0 = 0.99t_H = (1.372 \pm 0.025) \times 10^{10} \text{ yr} . \quad (5.15)$$

The numerically calculated age is shown in Fig. 5.1 as a function of Ω_M for $\Omega_A = 0$ and $\Omega_A = 1 - \Omega_M$. We see that the age is a decreasing function of Ω_M . This is quite simple because normal gravitation decelerates the expansion, $\ddot{a} < 0$. Since the present value of \dot{a}/a_0 is fixed ($= H_0$), any deceleration would imply a greater value of \dot{a}/a_0 in the past and therefore less time to go from $a = 0$ and $a = a_0$. Similarly, the age is an increasing function of Ω_A , because a positive vacuum energy accelerates the expansion implying a smaller value of \dot{a} in the past and therefore more time to go from $a = 0$ and $a = a_0$.

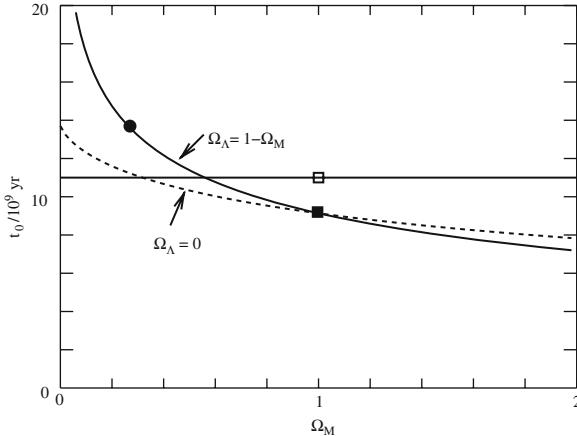


Fig. 5.1 The age of the universe, t_0 , calculated with $H_0 = 70 \text{ km sec}^{-1} \text{ Mpc}^{-1}$, as a function of Ω_M for $\Omega_\Lambda = 0$ and for $\Omega_\Lambda = 1 - \Omega_M$. The *filled circle* shows the value for the WMAP [1] parameters for a Λ CDM universe, $(\Omega_M, \Omega_\Lambda) = (0.27, 0.73)$. The *horizontal line* shows the lower limit on the ages of the oldest stars in the Milky Way. The *filled square* shows the age for $(\Omega_M, \Omega_\Lambda) = (1, 0)$. This possibility is excluded because the age is less than the age of these stars. The *open square* shows the age for $(\Omega_M, \Omega_\Lambda) = (1, 0)$ but assuming $H_0 = 60 \text{ km sec}^{-1} \text{ Mpc}^{-1}$

Note that for $\Omega_M = 1$, the calculated age is less than the age of the oldest stars on the Milky Way. Among the oldest [114] are those contained in certain globular (star) clusters in the halo of the Milky Way. The ages of these stars can be estimated from the observed color-magnitude diagram which evolves with time as in Fig. 5.2. Estimates of the ages of the oldest known cluster (M4) based on ages of main-sequence stars give [116]

$$t_{\text{ms}} = (12.6 \pm 1.1) \times 10^9 \text{ yr} \quad t_{\text{ms}} > 10.4 \times 10^9 \text{ yr (95\% CL)}.$$

Adding a reasonable amount of time between the big bang and the formation of globular clusters, one can conclude [116]

$$t_0 > 11.2 \times 10^9 \text{ yr (95\% CL).} \quad (5.16)$$

Note that the estimated age of M4 based on main sequence stars is consistent with the ages of the oldest white dwarfs in M4 [117]:

$$t_{\text{wd}} = (12.7 \pm 0.7) \times 10^9 \text{ yr.}$$

We note that this age can be compared with the ages of the oldest white dwarfs in the disk of the Milky Way, $(7.3 \pm 1.5) \times 10^9 \text{ yr}$. The difference is consistent with the theoretical expectation that the disk formed by radiative cooling after the formation of the spherical halo.

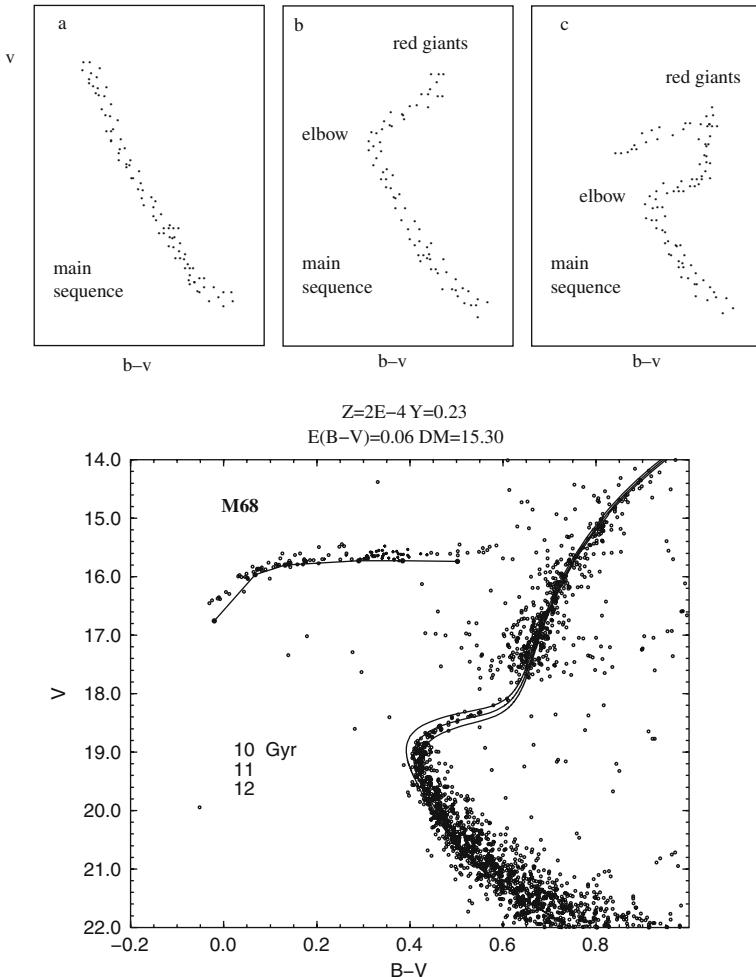


Fig. 5.2 The *top panel* shows the evolution of the color-magnitude diagram of a globular cluster. Globular clusters are star clusters (containing $\sim 10^5$ members) that can be found sprinkled about galactic disks and halos. When the cluster is formed, all stars are along the main sequence. The bright stars on the upper left of the m.s. have shorter hydrogen-burning lifetimes than the faint stars on the lower right ($L \propto M^3 \rightarrow \tau \propto M^{-2}$, Exercise 2.9.) As a result, as time passes, the bright stars leave the mean sequence and pass through various stages as red giants. The age of the globular cluster can therefore be deduced from the hydrogen-burning lifetime of the brightest star still on the main sequence. The *bottom panel* shows the color-magnitude diagram of the globular cluster M68. The diagram shows the main sequence along the diagonal from $(B-V = 0.8, V = 22)$ and ending abruptly at $(B-V = 0.4, V = 19)$. The three curves show the calculated position of the brightest main-sequence stars (the “main-sequence turnoff”) for M68 ages of 10, 11, and 12 Gyr. The calculated turnoff point agrees with the observations for an age of 11 Gyr [115]. Courtesy of A. Weiss

Requiring $t_0 > 1.1 \times 10^9$ yr requires $\Omega_M < 0.6$ for a Λ CDM model or $\Omega_M < 0.3$ for a $\Lambda = 0$ model. In the 1990s the limit assuming $\Lambda = 0$ caused considerable worry for advocates of inflation who expected $\Omega_T = 1$. At that time, the preferred way to avoid conflict with the inflationary prediction was to suppose $H_0 \sim 50 \text{ km s}^{-1} \text{Mpc}^{-1}$, corresponding to a longer Hubble time. The other way of avoiding the conflict was to invoke a vacuum energy to increase the age. In the end, it was discovered that this was Nature's choice.

This so-called “age” problem of the 1990s followed a much greater age problem in the 1930s. Originally, Hubble measured an expansion rate that was more than an order of magnitude larger than the presently accepted value. (The “Cepheids” he identified where, in reality, a more luminous type of star, causing him to underestimate galactic distances.) This gave a Hubble time of order $t_H \sim 10^9$ yr, significantly less than the known age of the Earth from radioactive dating. This conflict was a source of inspiration for “steady-state models.” Such models hypothesize the continuous creation of matter so as to keep a constant energy density. These models are dynamically equivalent to vacuum-dominated models and give an infinite age for the universe if the energy density is truly constant.

Better estimates of H_0 eliminated this motivation for the steady-state model. The fact that the oldest known stars have ages of order t_H is a strong argument against a steady-state universes since in that model one would expect to see much older stars. It is known that stars were formed $\sim 5 \times 10^8$ yr after the big bang at the epoch of reionization. It is thus encouraging that the estimated ages of the oldest Milky Way globular clusters is not too far from the estimated age of the universe.

Finally, we emphasize that the calculation of the age of the universe supposes that nothing strange happens for $a \rightarrow 0$. Since we do not know the physics at early time (high temperatures), we cannot justify this assumption. It is easy to think of ways of modifying $\rho(a)$ for $a \rightarrow 0$ so as to increase significantly the age. This can be seen in the general expression for $t(a)$:

$$t_0 = \int_0^{a_0} \frac{da}{a \left(8\pi G\rho(a)/3 + H_0^2(1 - \Omega_T)\hat{a}^{-2} \right)^{1/2}}. \quad (5.17)$$

For example, the integral diverges if $\Omega_T = 1$ and $\rho(a) \rightarrow \text{constant}$ for $a \rightarrow 0$. This can be the case in inflationary models. For this reason, the “age of the universe” is better called the elapsed time since the beginning of the epoch of known physics, for instance since $T \sim \text{MeV}$.

5.2 Type Ia Supernovae as Standard Candles

In Sect. 3.5, we showed that the radial coordinate χ_1 of an object of redshift z is given by (5.8). Visible galaxies were all created long after the end of the radiation epoch so we can safely neglect the Ω_R term in the integrand. The present distance to an object of redshift z is therefore

$$a_0 \chi_1(z) = d_H \int_0^z \frac{dz'}{\left(\Omega_M \hat{a}^{-3} + \Omega_\Lambda + (1 - \Omega_M - \Omega_\Lambda) \hat{a}^{-2} \right)^{1/2}}, \quad (5.18)$$

where $\hat{a} = 1/(1+z')$. For nearby objects (distances $\ll d_H \Rightarrow z \ll 1$) we have already seen (3.61) that this reduces to

$$a_0 \chi_1(z) \sim z d_H \left[1 - \frac{1+q_0}{2} z + \dots \right] \quad z \ll 1, \quad (5.19)$$

where the deceleration parameter q_0 is

$$q_0 = \frac{\Omega_M}{2} - \Omega_\Lambda. \quad (5.20)$$

The zeroth-order term in (5.19) $a_0 \chi_1(z) = H_0^{-1} c z$ is just an expression of the Hubble law, i.e. the velocity $c z$ is proportional to the distance $a_0 \chi$. The first-order correction to the Hubble law depends on q_0 , for basically the same reason that the age of the universe depends on the deceleration. A given redshift corresponds to a fixed expansion factor a_0/a_1 between emission and absorption. Deceleration implies fast expansion (in the past) diminishing the expansion time and, therefore, the photon flight time. A diminished flight time implies a diminished χ_1 (Fig. 3.13).

The distance $a_0 \chi$ to an object of known luminosity, a so-called “standard candle,” can be deduced from the measured flux through (3.63) or (3.69). For $z \ll 1$, the flux is independent of the Ω 's and simply inversely proportional to $(z/H_0)^2$. At $z = 0.5$, the Ω dependence is non-negligible and gives a $\sim 50\%$ difference between the fluxes of $(\Omega_M, \Omega_\Lambda) = (1, 0)$ and $(\Omega_M, \Omega_\Lambda) = (0.27, 0, 73)$.

The best standard candles that are observable at cosmological distances are type Ia supernovae (Sect. 2.1). At the time of maximum luminosity, they have luminosities that are a function of the supernova's color and time duration (Fig. 2.9) with an event-to-event dispersion of about 15% about the mean. This dispersion is less than the difference in flux of 50% between $(\Omega_M, \Omega_\Lambda) = (1, 0)$ and $(0.27, 0, 73)$, so in principle it is simple to distinguish the two cases by observing a few $z \sim 0.5$ supernovae. This was first done in 1998 by the HighZ team [119] and the Supernova Cosmology Project [118]. Much more precise data has since been obtained by the Supernova Legacy Survey (SNLS) [2]. Their data is shown in Fig. 5.3 which plots the flux-derived distance modulus:

$$\mu \equiv 5 \log(d_L/10 \text{ pc}) = 2.5 \log f - 2.5 \log[(L/4\pi(10 \text{ pc})^2)] \quad (5.21)$$

where the luminosity distance is $d_L = (1+z)a_0 \chi_1$ and the total flux and luminosity are f and L . (In fact, the flux integrated over pass-bands is used with (3.69).) If the flux is measured and the luminosity calculated from the observed supernova color and time-duration, the distance modulus can be calculated and plotted against the redshift. For $z \ll 1$ the distance modulus is

$$\mu \sim 5 \log z + \frac{5}{2 \ln 10} (1 - q_0) z + 5 \log d_{\text{H}}. \quad (5.22)$$

The data in Fig. 5.3 follow the $5 \log z$ term for $z < 0.1$ while the q_0 correction is apparent for $z > 0.1$. The data clearly show that supernovae at $z \sim 0.5$ are significantly fainter than one would expect for $(\Omega_M, \Omega_\Lambda) = (1, 0)$. The ellipses in Fig. 5.4 show the range of values of $(\Omega_M, \Omega_\Lambda)$ that are consistent with the data. The allowed region corresponds roughly to

$$\Omega_\Lambda - \Omega_M \sim 0.45 \pm 0.12 \quad \Omega_M < 0.5 \quad (5.23)$$

which requires $q_0 < 0$. (The ellipses are more closely aligned with $\Omega_M - \Omega_\Lambda$ than with $q_0 = \Omega_M/2 - \Omega_\Lambda$ because at $z \sim 0.5$ (5.19) is not a very good approximation.)

Note that it is not necessary to know L (Eq. 5.21) or d_{H} (Eq. 5.22) since these two quantities only shift the two curves in Fig. 5.3 up or down without affecting $(\Omega_M, \Omega_\Lambda)$ dependence. This can be seen by writing the difference, $\Delta\mu$, in distance moduli between high redshift (say $z = 0.5$) and low redshift (say $z = 0.05$):

$$\Delta\mu = 2.5 \log \left(\frac{f(z=0.05)}{f(z=0.5)} \right) = 5 + \frac{5}{2 \ln 10} (1 - q_0) z \dots .$$

This shows that it is the ratio between fluxes of high and low redshift that gives the information on the Ω 's. From a theoretical point of view, the only important things are that the luminosity be redshift-independent and that there be no redshift-dependent absorption. Observationally, the difficulty of this measurement comes from the necessity of accurately comparing high and low-redshift fluxes that differ in flux by a factor ~ 100 and that cover different wavelength bands (because of the redshift).

It must be emphasized that the explanation of the Hubble diagram based on $q_0 < 0$ must compete with two more mundane astrophysical interpretations:

- The mean supernova luminosity may have been smaller in the past.

The elemental abundances in the interstellar medium is known to evolve with time as heavy nuclei produced in stars are progressively ejected into interstellar space. Since it is this interstellar material that collapses to form stars, recently formed stars have an *initial* chemical composition that is different from that of stars formed long ago. While the chemical composition just before the explosion is primarily determined by post-main-sequence nuclear burning stages, the initial compositions have some influence that may cause differences in the luminosities of high- and low-redshift supernovae. It might be expected that such luminosity evolution would lead to the evolution of supernova spectra. No spectral differences between high- and low-redshift supernovae has been observed that would hint at a large effect [121]. Progress in the theoretical modeling of supernovae would be important in increasing our confidence that there is no evolution or in identifying and correcting for it.

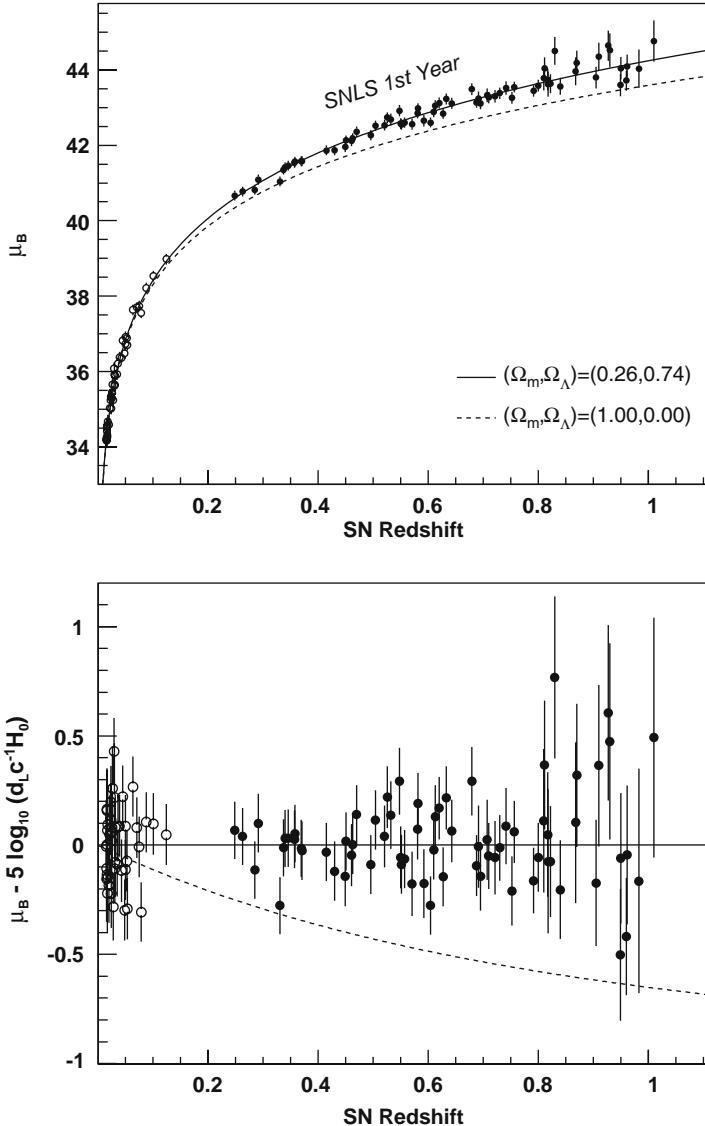


Fig. 5.3 Distance measurements using type Ia supernova from the Supernova Legacy Survey (SNLS) and a collection of low-redshift supernovae [2]. The *top panel* gives the distance modulus, $\mu_B = 5 \log(d_L/10 \text{ pc})$, as a function of the redshift z . At low redshift, the data follows the $1/z^2$ law while for $z > 0.1$ the data follows a curve that depends on the values of $(\Omega_M, \Omega_\Lambda)$. The luminosity distance of a supernova, $d_L = (1 + z)a_0\chi_1(z)$, is deduced from the measured flux. The *bottom panel* shows the difference between the data and the best fit model $(0.26, 0.74)$.

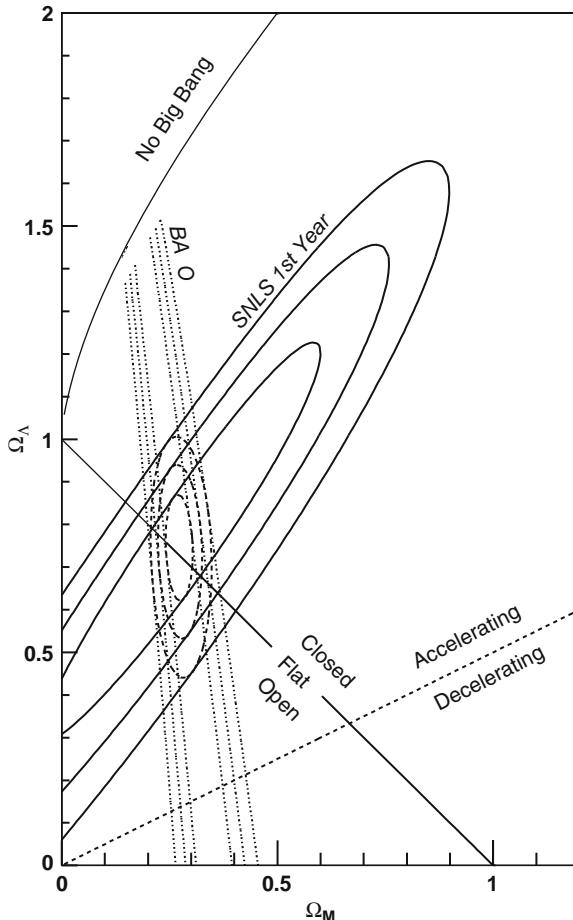


Fig. 5.4 The allowed ranges of Ω_M and Ω_Λ from the distance measurement by the Supernova Legacy Survey [2]. The ellipses marked “SNLS first year” show the one, two, and three standard deviation contours derived from the data shown in Fig. 5.3. Also shown are the constraints from the BAO measurement of SDSS [120]. Also shown is the line $\Omega_T = 1$ (Flat), which is required to a precision of $\sim 1\%$ by the BAO data combined with WMAP. The three constraints require $\Omega_M = 0.273 \pm 0.014$ and $\Omega_\Lambda = 0.726 \pm 0.015$ (Table 1.1)

The second astrophysical interpretation is

- The supernova light may be absorbed somewhere between the supernova and Earth.

The absorption must be roughly wavelength independent because the shape of the high-redshift spectra are observed to be nearly identical to those of nearby supernovae. This would require the existence of an absorbing dust with characteristics different from dust in the Milky Way (which leads to reddening of the spectra). The

dust would necessarily re-emit starlight in the infrared creating an isotropic infrared background. The present limit on the diffuse background is not strong enough to eliminate the possibility of the existence of the required amount of dust [122].

5.3 The Sound Horizon as a Standard Ruler

The two problems with the use of type Ia supernova as standard candles are hypothetical luminosity evolution and photon absorption by dust. These problems are avoided with measurements of baryon acoustic oscillations (BAO) that use the sound horizon at recombination as a standard ruler. Standard rulers are clearly insensitive to absorption, which changes the visibility of a ruler but not its angular size. Furthermore, we will see that the sound horizon is a ruler whose length can be accurately calculated from well-understood fundamental physics. Thus, there are no large uncertainties associated with the evolution of the ruler's size with redshift.

The sound horizon at recombination is the coordinate distance, χ_s , that a sound wave supported by the photon–baryon–electron plasma can travel between the big bang and recombination. We will calculate this distance below as a function of the Ω 's. For $(\Omega_M, \Omega_\Lambda) = (0.27, 0.73)$ the corresponding present distance is [1]

$$a_0 \chi_s = (153.3 \pm 2.0) \text{ Mpc} . \quad (5.24)$$

This distance is seen on the sky by the slight preference for pairs of galaxies at time t to be separated by a distance $a(t)\chi_s$. The preferred distance expands with the universe so the size of the ruler today is 153 Mpc.

This preference of galaxies to be separated by the sound horizon comes about because galaxies form preferentially in over-dense regions of space. The over-densities were seeded by small initial inhomogeneities in the early universe (quantum fluctuations of a scalar field in models of inflation). The time evolution of an over-density consisting of CDM, baryons, and photons is illustrated in Fig. 5.5. In the standard model of structure formation, initial density perturbations are “adiabatic” meaning that perturbations of all components (CDM, baryons, radiation) are initially proportional to each other. The heavy and pressureless CDM component stays near the origin but the baryon–photon pressure causes this component to propagate away from the initial perturbation as a sound wave. At recombination, the initial CDM perturbation is surrounded by a shell of excess baryons.

The radius of the shell at recombination is $a_{\text{rec}}\chi_s$ which we will now calculate. For a perfect fluid, the square of the sound velocity is determined by the relation between the pressure, p , and energy density ρ : $c_s^2 = dp/d\rho$, where the variations dp and $d\rho$ are taken for *adiabatic* expansion of a fluid element by a volume change dV . For an adiabatic expansion, $d(\rho V) = -pdV$ giving $d\rho = -(p + \rho)dV/V$. The pressure comes only from the photons, $p = \rho\gamma/3$ so $3dp = d\rho - d\rho_b$ where ρ_b is the baryon density. The pressureless baryons have $d\rho_b = \rho_b dV/V$. Putting all this together we have

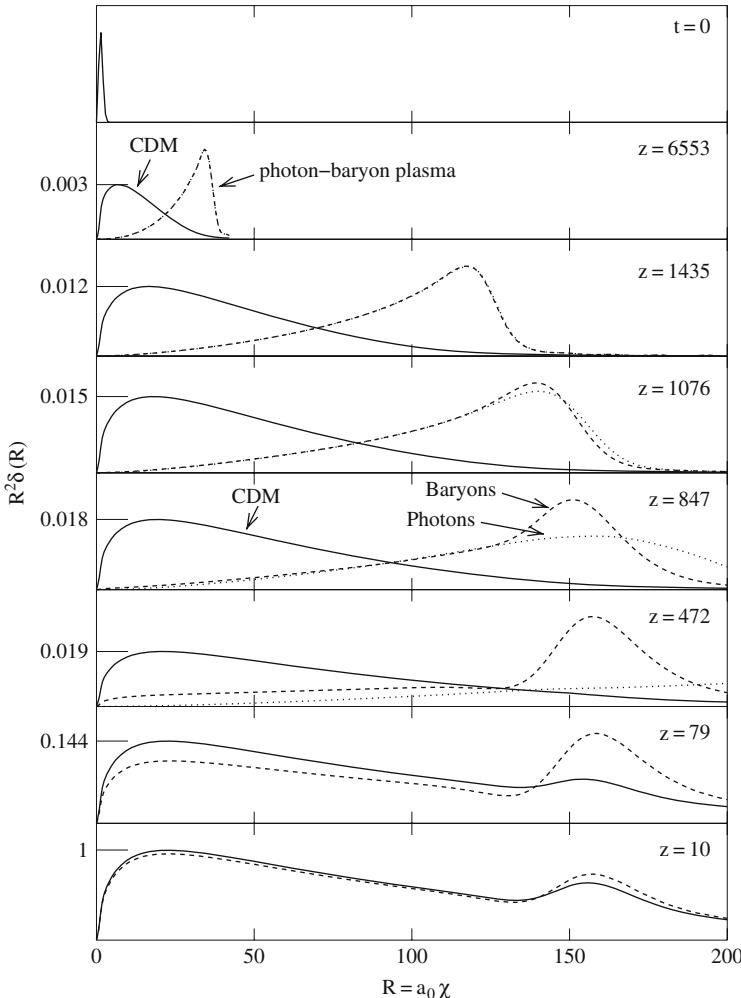


Fig. 5.5 The time development of an initial adiabatic over-density in a universe with CDM, neutrinos, baryons, and photons [152, 123]. At $t = 0$, the over-density of all components are superimposed but the pressure of the baryon–photon plasma causes it to propagate away from the origin at the speed of sound. Light neutrinos (not shown) free stream away with the speed of light. The baryons stop shortly after recombination when the baryons and photons decouple, allowing the photons to free stream away. CDM and baryons from the homogeneous reservoir will then be gravitationally attracted into the potential wells formed by the CDM at the origin and the shell of baryons. This infall of homogeneous matter will generate CDM–baryon over-densities where galaxies will preferentially form. This results in the galaxy–galaxy correlation function seen in Fig. 5.7.

$$c_s^2 = \frac{c^2}{3(1 + \epsilon(a))} \quad (5.25)$$

where

$$\epsilon(a) = \frac{3\rho_b}{4\rho\gamma} = \frac{3m_p\eta n\gamma}{4\rho\gamma} \sim 0.89 \frac{a}{a_{\text{rec}}} . \quad (5.26)$$

where $\eta = n_b/n\gamma = (6.18 \pm 0.2) \times 10^{-10}$. The speed of sound ranges from $c_s = c/\sqrt{3}$ for $a \ll a_{\text{rec}}$ to $c_s = 0.73c/\sqrt{3}$ at $a = a_{\text{rec}}$.

Wave fronts propagated with the same equation of motion as that for electromagnetic waves except that the speed of light is replaced with the speed of sound:

$$\frac{d\chi}{dt} = \frac{c_s}{a(t)} \quad (5.27)$$

This equation can be integrated over t or over $a(t)$ to calculate $\chi(t)$. The coordinate distance traveled between $t = 0$ and recombination is finite, corresponding to a present distance

$$a_0\chi_s = d_H \int_0^{\hat{a}_{\text{rec}}} \frac{c_s d\hat{a}}{\hat{a}^2 (\Omega_M \hat{a}^{-3} + \Omega_R \hat{a}^{-4})^{1/2}} . \quad (5.28)$$

where we neglect vacuum energy and curvature during the pre-recombination epoch. A good first approximation for this distance can be made by taking $c_s = 1/\sqrt{3}$ and neglecting the Ω_R term. This gives

$$a_0\chi_s \sim \frac{d_H}{\sqrt{3}\Omega_M} \int_0^{\hat{a}_{\text{rec}}} \frac{d\hat{a}}{\hat{a}^{1/2}} = \frac{2d_H}{\sqrt{3}\Omega_M(1 + z_{\text{rec}})} . \quad (5.29)$$

Numerically, this gives $a_0\chi_s = 300h_{70}^{-1}$ Mpc for $\Omega_M = 0.27$. Including the effects of Ω_R and $c_s < c/\sqrt{3}$ lowers this to

$$a_0\chi_s \sim 153h_{70}^{-1}\text{Mpc} \quad (5.30)$$

for $\Omega_M = 0.27$. Note also that the integration limit in (5.28) should be slightly higher than \hat{a}_{rec} (1/1020 instead of 1/1090; see Table 7.2). This is because the baryons continue to be dragged along by the numerous photons for a short time after recombination (defined as the mean liberation time for the photons).

This shell surrounding the CDM perturbation has two observable effects:

1. A slight preference for pairs of galaxies to be separated at time t by distances of $a(t)\chi_s$. Such pairs correspond to a galaxy that forms at the position of the initial perturbation and to another galaxy formed on the shell.

2. A preference for hot and cold regions on the CMB sky to be separated by $\sim a_{\text{rec}} \chi_s$.

The distance $a_0 \chi_s$ is much greater than the size of any bound objects in the universe so galaxy pairs separated today by this distance were separated by a smaller distance, $a(t) \chi_s$ in the past. For galaxies at redshift $\sim z$, the preferred distance is $a_0 \chi_s / (1 + z)$. This corresponds to a preferred angular separation (Eq. 3.70) of

$$\Delta\theta_s = \frac{a_0 \chi_s / (1 + z)}{d_A(z)} = \frac{\chi_s}{r_1(z)} \quad (5.31)$$

and a preferred redshift separation (Eq. 3.72) of

$$\Delta z_s = \frac{a_0 \chi_s}{H(z)^{-1}} = \frac{a_0 \chi_s}{4300 h_{70} \text{Mpc}} (\Omega_M(1+z)^3 + \Omega_\Lambda)^{1/2} \quad (5.32)$$

where $H(z)$ is the expansion rate at t_1 . Figure 5.6 shows the angular and redshift separations as a function of redshift for three combinations of Ω_M and Ω_Λ . As expected from (5.29), at low redshift, the separations depend mostly on Ω_M . At high redshift, the angular separation depends mostly on Ω_Λ .

The effect on the distribution of galaxies was observed by two large redshift surveys, the Sloan Digital Sky Survey (SDSS) [120] and the 2dF Galaxy Redshift Survey (2dFGRS) [4]. The correlation between pairs of SDSS galaxies is shown in Fig. 5.7. The position of the peak constrains Ω_M to be within a narrow range as shown in Fig. 5.4. For the Λ CDM model, SDSS reports

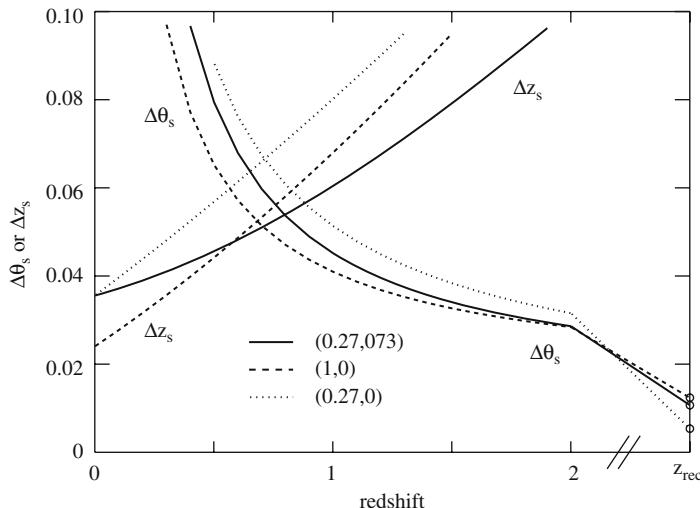


Fig. 5.6 The angular size (in radians) and redshift size of the sound horizon at recombination as a function of redshift for three combinations of $(\Omega_M, \Omega_\Lambda)$

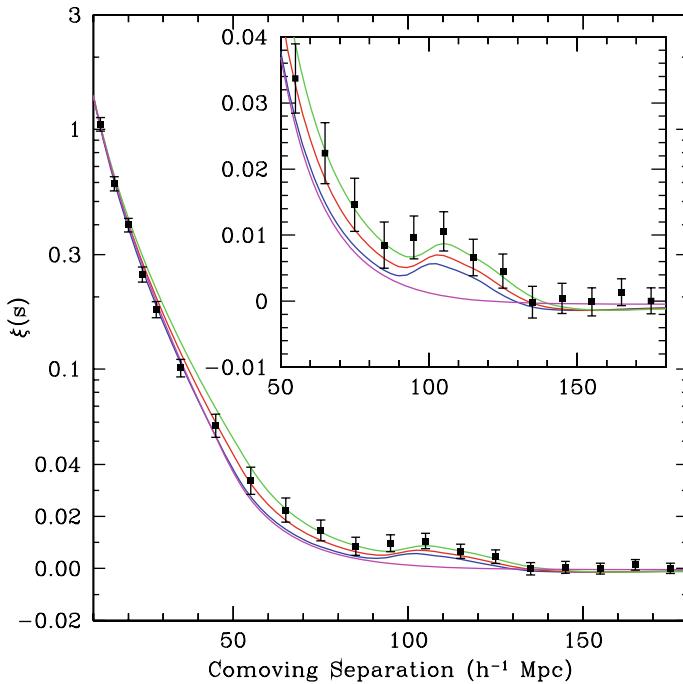


Fig. 5.7 The galaxy–galaxy correlation function measured by the SDSS [120]. The small peak at a separation of $105 h^{-1} \text{Mpc} = 150 h_{70}^{-1} \text{Mpc}$ corresponds to the sound horizon at recombination

$$\Omega_M = 0.273 \pm 0.025 + 0.137(1 - \Omega_T). \quad (5.33)$$

The slight dependence of Ω_M on $(1 - \Omega_T)$ comes about from the fact that the mean redshift of the galaxies used by SDSS was ~ 0.3 , sufficiently high that the angular and redshift size of the ruler depends on Ω_M and Ω_A .

The angular size of the sound horizon at recombination, $z \sim 1090$, can be derived from the spectrum of CMB anisotropies. To calculate the theoretical angle, we need the high-redshift angular distance. For $\Omega_M = \Omega_T = 1$ we easily calculate

$$\lim_{z \rightarrow \infty} a_0 \chi_1(z) = 2H_0^{-1} \quad \Omega_M = \Omega_T = 1. \quad (5.34)$$

In the case of $\Omega_A = 0$, the result, calculated in Exercise 5.10, is:

$$\lim_{z \rightarrow \infty} a_0 r_1 = \frac{2H_0^{-1}}{\Omega_M} \quad (\Omega_A = 0). \quad (5.35)$$

Using the approximation (5.29), this gives

$$\Delta\theta_s(z_{\text{rec}}) \sim \sqrt{\frac{\Omega_M}{3(1+z_{\text{rec}})}} \sim 0.02\sqrt{\Omega_M} \quad \Omega_A = 0, \quad c_s = \frac{1}{\sqrt{3}}.$$

The angular size is an increasing function of $\Omega_M = \Omega_T$.

For $\Omega_A \neq 0$, the angular size must be calculated numerically but it turns out that the correlation between $\Delta\theta_s(z_{\text{rec}})$ and Ω_T is maintained. Figure 5.8 shows the numerically calculated angular size as a function of (Ω_M, Ω_A) . The lines of constant angle follow roughly the lines of constant Ω_T , so a measurement of the angle will strongly constrain Ω_T .

The first measurement of $\Delta\theta_s(z_{\text{rec}})$ was made by the Boomerang [124] and Maxima [125] experiments in 2000. Much more precise data has since been obtained by the WMAP satellite. Their spectrum, Fig. 7.11, sets very precisely the angular scale (Table 7.2):

$$\Delta\theta_s(z_{\text{rec}}) = 0.01039 \pm 0.00003 \text{ rad.} \quad (5.36)$$

Combined with the SDSS result (5.33) implies very tight constraints on Ω_M , Ω_A , and Ω_T :

$$(\Omega_T, \Omega_M, \Omega_A) = (0.99 \pm 0.009, 0.273 \pm 0.027, 0.74 \pm 0.01). \quad (5.37)$$

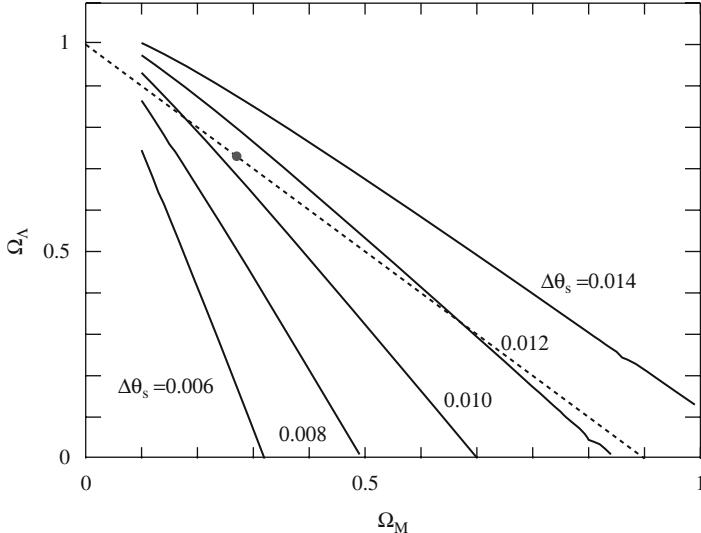


Fig. 5.8 The angular size (in radians) of the sound horizon, $\Delta\theta_s(z_{\text{rec}})$, at recombination as a function of Ω_M and Ω_A assuming $\Omega_\gamma = 5 \times 10^{-5}$ and three massless neutrino species. Note that the lines of constant angular size are nearly orthogonal to the lines of constant $\Omega_M - \Omega_A$, making CMB measurements of the sound horizon give constraints that are complementary to those of SNIa measurements (Fig. 5.4)

5.4 The Horizon Problem

While the sound horizon is the largest distance that a sound wave can travel since the big bang, the *particle horizon* (or simply “horizon”) is the largest distance that a particle can travel at the speed of light. It is found by replacing c_s with $c = 1$ in (5.28):

$$a_0 \chi_{\text{hor}}(a) = c/H_0 \int_0^{\hat{a}} \frac{d\hat{a}/\hat{a}^2}{(\Omega_M \hat{a}^{-3} + \Omega_R \hat{a}^{-4} + \Omega_\Lambda + (1 - \Omega_T) \hat{a}^{-2})^{1/2}} .$$

For $a \rightarrow 0$, the integrand is dominated by the radiation term and the integral is finite. Photons emitted at points $\chi > \chi_{\text{hor}}$ have not yet reached us. The present distance to the horizon, $a_0 \chi_{\text{hor}}(a_0)$, is of order c/H_0 :

$$\begin{aligned} (\Omega_M, \Omega_\Lambda) &= (1, 0) & \Rightarrow a_0 \chi_{\text{hor}}(a_0) &= 2c/H_0 \\ (\Omega_M, \Omega_\Lambda) &= (0.27, 0.73) & \Rightarrow a_0 \chi_{\text{hor}}(a_0) &\sim 3.3c/H_0 . \end{aligned}$$

Since there is nothing special about the present epoch, it is not surprising that the distance to the horizon at a given epoch is of the order of the Hubble distance at that epoch. From the Friedmann equation we know that $d_H \propto a^2$ during the radiation epoch and $d_H \propto a^{3/2}$ during the matter epoch. It follows that

$$\chi_{\text{hor}}(a) \propto a \quad a < a_{\text{eq}} \quad (5.38)$$

$$\chi_{\text{hor}}(a) \propto a^{1/2} \quad \text{matter epoch} . \quad (5.39)$$

The evolution of the horizon with time is illustrated in Fig. 5.9 which shows the trajectories $\chi(t)$ of photons that intersect the origin $\chi = 0$.

The existence of a horizon is perhaps surprising in view of the fact that co-moving observers are separated by distances that are proportional to $a(t)$ and therefore vanish at the big bang.¹ However, the age of the universe vanishes even faster, $t \propto a^2$ for $t < t_{\text{eq}}$. It follows that

$$\frac{\text{distances}}{\text{time available}} \propto a^{-1} , \quad (5.40)$$

so co-moving observers are completely isolated for $a \rightarrow 0$.

The existence of a horizon presents us with a paradox concerning the uniformity of the CMB. At t_{rec} the horizon was of the order of the Hubble distance at this epoch, $a_{\text{rec}} \chi_{\text{hor}}(a_{\text{rec}}) \sim d_H(a_{\text{rec}})$. As we saw in the previous section, this corresponds to an angle on the sky of about 1 deg. As illustrated in Fig. 5.10, points separated

¹ While this is true, it is important to remember that the volume of a homogeneous universe with $\Omega_T \leq 1$ is infinite for all values of $a > 0$.

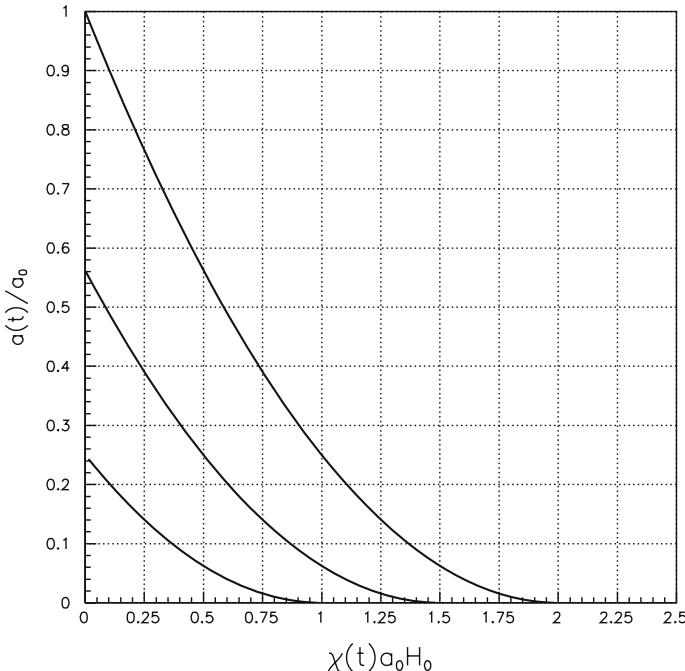


Fig. 5.9 Photon trajectories $\chi(t)$ that start at the big bang and intersect the origin $\chi = 0$ (supposing that the universe is transparent). The trajectories are calculated for a $\Omega_T = \Omega_M = 1$ universe $a(t) \propto t^{2/3}$ which implies that $\chi(t) = \chi(t=0) - 2a_0^{-1}H_0^{-1}\sqrt{a(t)/a_0}$. The time (vertical axis) is parameterized by $a(t)/a_0$. The trajectory $\chi(t)$ (horizontal axis) is multiplied by $a_0 H_0$. We see that a photon emitted at a co-moving point that is currently at a distance $\chi a_0 = 2H_0^{-1}$ reaches the origin today ($a(t)/a_0 = 1$). This means that our horizon is at a distance of $2H_0^{-1}$. As time increases, an observer at $\chi = 0$ sees photons arriving from co-moving points that are more and more distant

by angles greater than this had not communicated before t_{rec} . The uniformity of the CMB is therefore a mystery since no causal mechanism, e.g. heat diffusion, could have made the temperature uniform.

There are two responses to this mystery. The first is to simply say that the initial conditions were such that the temperature was uniform. This might be equivalent to saying that the initial temperature was fixed by the “laws of physics” or maybe that it has nothing to do with physics.

The second, more popular solution, is to modify the energy density $\rho(a)$ in the limit $a \rightarrow 0$ in such a way as to considerably increase χ_{hor} . A simple example would be to suppose $\rho(a) = \text{constant}$ for $a_2 < a < a_1 \ll a_{\text{eq}}$ as shown in Fig. 5.11. This is equivalent to supposing that for $a_2 < a < a_1$ the universe is dominated by a vacuum energy that is somehow transformed into radiation at a_1 . This is more or less what happens in the inflationary models that will be discussed in Sect. 5.6. Since the evolution of ρ is no longer described by a simple mixture of particles and vacuum, we must go back to the general expression for $\chi(z)$

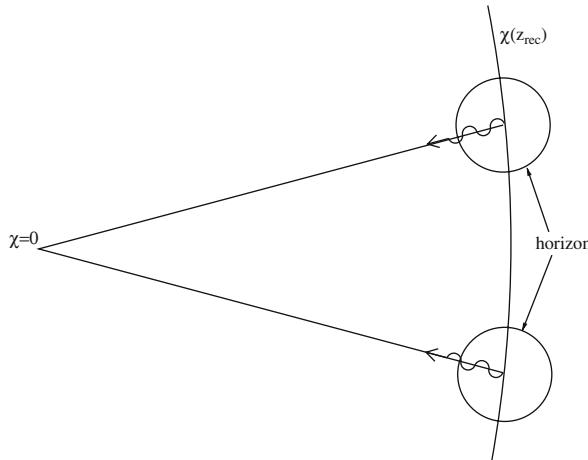


Fig. 5.10 Two points on the surface $\chi = \chi(z_{\text{rec}})$ with their horizons at t_{rec} . Pairs of points separated by more than ~ 1 deg on the sky were outside each other's horizon. The near equality of the temperatures of the two points could not have resulted from a causal process like heat diffusion

$$\chi_{\text{hor}} = \int_0^{a_0} \frac{da}{a\dot{a}} = \int_0^{a_0} \frac{da}{a^2 (8\pi G\rho/3 + H_0^2(1 - \Omega_T)\hat{a}^{-2})^{1/2}}. \quad (5.41)$$

For a constant energy density for $a_2 < a < a_1$, this expression diverges for $a_2 \rightarrow 0$. We can thus formally solve the horizon problem with this “trick.” The reason is that

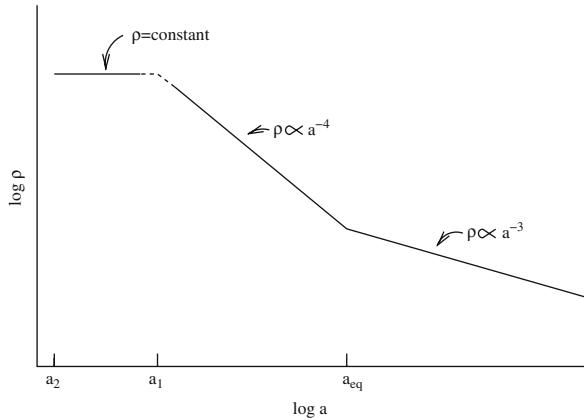


Fig. 5.11 The energy density $\rho(a)$ in a simple inflationary scenario. The radiation and matter epochs are preceded by an inflationary epoch $a_2 < a < a_1$ when the universe is dominated by an effective vacuum energy. At the end of inflation, this energy must be partly transformed into radiation, providing the initial conditions for classical cosmology. This process will be described in Sect. 5.6

during the epoch of domination by a constant energy density, the universe expands exponentially with a constant Hubble time H_1^{-1} for $a_2 < a < a_1$:

$$a(t) = a_2 \exp H_1(t - t_2) \quad H_1^2 \sim \frac{8\pi G\rho(a_1)}{3}. \quad (5.42)$$

The universe spends a Hubble time H_1^{-1} with $a \sim a_2$. If the universe had been radiation-dominated for $a_2 < a < a_1$ it would have spent considerably less time with $a \sim a_2$: $H_2^{-1} \sim H_1^{-1} a_2^2/a_1^2$. The increased time spent when the universe was “small” allowed communication over larger co-moving distances. This is illustrated in Fig. 5.12.

Inflationary models will be discussed in more detail in Sect. 5.6. Without knowing the details, we can easily estimate the duration of inflation needed to solve the horizon problem. During inflation, the horizon is

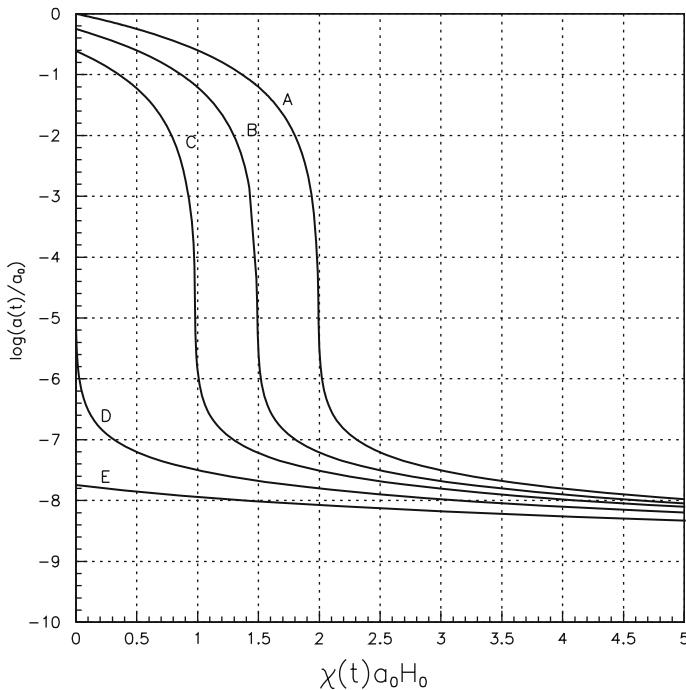


Fig. 5.12 Photon trajectories $\chi(t)$ as in Fig. 5.9 except that the universe is vacuum-dominated ($\rho = \text{constant}$) during an inflationary epoch $a(t)/a_0 < 10^{-5}$. (In more realistic inflationary models, the inflationary epoch ends after the GUT epoch $a(t)/a_0 \sim 10^{-28}$, $T \sim 10^{15}$ GeV). During the matter epoch, the three trajectories A, B, and C follow the three trajectories of Fig. 5.9. At the end of inflation, trajectory A passes through the classical horizon at a present distance of $2H_0^{-1}$. During the inflationary epoch, the curves continue to the right, considerably increasing the horizon. This is due to the increased time the universe spends at small $a(t)$. We note (trajectory D) that the point $\chi = 0$ was, at the end of inflation, already able to communicate with the entire classical horizon if inflation started before $a(t)/a_0 = 10^{-8}$

$$\chi_{\text{hor}}(t) \sim \int_{t_2}^t \frac{dt}{a(t)} \sim \frac{1}{H_1 a_2} [1 - e^{-H_1(t-t_2)}] \sim \frac{1}{H_1 a_2}, \quad (5.43)$$

where we have neglected the period before inflation. We see that the horizon is essentially fixed at the beginning of inflation when the distances between co-moving points were smallest. To solve the horizon problem, we want the horizon during inflation to include the present Hubble volume:

$$a_0 \chi_{\text{hor}}(t \sim t_2) > H_0^{-1}. \quad (5.44)$$

Using (5.43) this becomes

$$\frac{a_0}{a_1} \frac{a_1}{a_2} H_1^{-1} > H_0^{-1}. \quad (5.45)$$

But, using the Friedmann equation and neglecting curvature we have

$$H_1^2 \sim \frac{\rho_1}{\rho_0} H_0^2 \sim \frac{\rho_R(t_0) (a_0^4/a_1^4)}{\rho_M(t_0)} H_0^2. \quad (5.46)$$

(We neglect effects due to the changing number of degrees of freedom as discussed in Chap. 6.) Taking $\rho_R(t_0)/\rho_M(t_0) \sim 10^{-4}$ and substituting into (5.45) we find that in order to solve the horizon problem we must have

$$\frac{a_1}{a_2} > 10^{-2} \frac{a_0}{a_1}, \quad (5.47)$$

i.e. that the expansion during inflation must be at least 1% of the expansion after inflation.

The success of the theory of primordial nucleosynthesis tells us that inflation could not have occurred after $T \sim \text{MeV}$. Often it is assumed that inflation has something to do with interactions in Grand Unified Theories of particle physics. This suggests that inflation occurred somewhat below the GUT scale, $T_1 \sim 10^{15} \text{ GeV}$, implying $a_0/a_1 \sim 10^{28}$. (Inflation is needed *below* the GUT scale in order to dilute certain unobserved particles, e.g. magnetic monopoles, that are created at the GUT scale.) While there are models where inflation occurs at scales as low as $\sim \text{TeV}$, we use the GUT scale for numerical estimates and we see that to solve the horizon problem we need

$$\frac{a_1}{a_2} > 10^{26} \sim e^{60} \quad \text{if } T_1 \sim 10^{15} \text{ GeV}. \quad (5.48)$$

The ~ 60 Hubble times during inflation lasts only $\sim 60H_1^{-1} \sim 10^{-34} \text{ s}$. During this time the present Hubble volume expands from a size of $\sim 10^{-28} \text{ m}$ to $\sim 10^{-2} \text{ m}$ before expanding in the radiation and matter epochs to its present size of $\sim 10^{26} \text{ m}$.

To have a reasonably uniform present-day Hubble volume, we require that the original, causally connected volume be reasonably uniform. The arbitrary *non-causal* initial condition is thus replaced with an arbitrary but *causal* condition that may be explainable with some yet-to-be established physics. It is thus not true that inflation results necessarily in a homogeneous universe, but only that the observable homogeneous universe could have developed causally. In the case of inflationary models where the energy is due to a scalar field, it is necessary that the field be well-approximated by a homogeneous classical field, requiring that the corresponding quantum field be in its ground state. It is a challenge for models of pre-inflationary physics, e.g. superstring landscape theories, to understand how this can come about and to determine if it is likely that observers like ourselves find themselves in a region that underwent inflation.

Inflationary models increase considerably our real horizon so we can ask what is the role of the “classical horizon” calculated ignoring inflation. The answer to this question can be seen in Fig. 5.12 where we see that while we can receive signals from points beyond the classical horizon, they were sent at a time when there were no objects like stars or galaxies or even CMB photons. The classical horizon therefore remains the practical horizon for doing astronomy.

It is interesting to see what will happen to the classical horizon in the future. It is simple to show (Exercise 5.9) that the classical horizon in a $\Omega_T = \Omega_M = 1$ universe will continue to increase with time. Referring to Fig. 5.9, we see that the photon trajectories will sweep out greater and greater areas as time increases. If we wait long enough, we will see everything.

On the other hand, we will see in Exercise 5.9 that in a universe with $\Omega_\Lambda > 0$, χ_{hor} approaches a finite limit for $t \rightarrow \infty$. The exponential expansion in such a universe does not leave enough time for photons to reach us from rapidly receding objects. One says that a universe with $\Omega_\Lambda > 0$ has an “event horizon” beyond which we will never see. This is similar to the situation with black holes where we cannot see within the Schwarzschild radius.

5.5 The Ω Problem

The ratio between the density $\rho(a)$ and the critical density as a function of the scale factor is

$$\Omega_T(a) = \frac{\rho(a)}{3H^2(a)/8\pi G} , \quad (5.49)$$

where $H(a) = \dot{a}/a$ is given by the Friedmann equation. If $\rho(a)$ is a mixture of relativistic, non-relativistic, and vacuum energy, this becomes

$$\Omega_T(a) = \frac{\Omega_R \hat{a}^{-4} + \Omega_M \hat{a}^{-3} + \Omega_\Lambda}{\Omega_R \hat{a}^{-4} + \Omega_M \hat{a}^{-3} + \Omega_\Lambda + (1 - \Omega_T) \hat{a}^{-2}} . \quad (5.50)$$

For $a \ll a_0$ we find that Ω_T approaches unity:

$$\Omega_T(a) = 1 - \frac{1 - \Omega_T}{\Omega_R + \Omega_M \hat{a}} \hat{a}^2 \quad \hat{a} \ll 1. \quad (5.51)$$

During the radiation epoch, $\Omega_T(a) - 1$ is thus proportional to the square of the scale factor. Extrapolating to the GUT epoch, $\hat{a}_{\text{GUT}} \sim 10^{-28}$ we find

$$\begin{aligned} 1 - \Omega_T(a_{\text{GUT}}) &\sim \frac{1 - \Omega_T(a_0)}{\Omega_R(a_0)} \hat{a}_{\text{GUT}}^2 \\ &\sim 10^{-52}[1 - \Omega_T(a_0)]. \end{aligned} \quad (5.52)$$

Since it is probably safe to say that $|1 - \Omega_T(a_0)| < 0.05$, we can conclude that at the GUT epoch, it was tuned to unity to a precision of more than 53 decimal places.

The peculiarity of this situation is better understood by rewriting (5.52) starting at the GUT epoch. During the radiation epoch we have

$$[1 - \Omega_T(a)] \sim [1 - \Omega_T(a_{\text{GUT}})] \left(\frac{a}{a_{\text{GUT}}} \right)^2. \quad (5.53)$$

We see that Ω_T starts to differ from unity after an expansion of $a/a_{\text{GUT}} \sim 1/\sqrt{|1 - \Omega_T(a_{\text{GUT}})|}$. The fact that Ω_T is still near unity is thus possible only because Ω_T was so closely tuned to unity at the GUT epoch.

As it turns out, an inflationary epoch that lasted long enough to solve the horizon problem will also push Ω_T sufficiently close to unity to satisfy (5.52). During inflation, Ω evolves according to

$$\Omega_T(a) = \frac{\Omega_T(a_2)}{\Omega_T(a_2) + (1 - \Omega_T(a_2))(a/a_2)^{-2}}, \quad (5.54)$$

where a_2 is the beginning of inflation. We see that during inflation $\Omega_T \rightarrow 1$ for $a \gg a_2$.

At the end of inflation we have

$$1 - \Omega_T(a_1) \sim \frac{1 - \Omega_T(a_2)}{\Omega_T(a_2)} \frac{a_2^2}{a_1^2}. \quad (5.55)$$

In order to solve the horizon problem, a_2/a_1 must respect the conditions (5.47) or (5.48) implying

$$|1 - \Omega_T(a_1)| < \frac{|1 - \Omega_T(a_2)|}{\Omega_T(a_2)} 10^{-52}. \quad (5.56)$$

The factor 10^{-52} is just what we needed. Combining this equation with (5.52) and setting $a_1 = a_{\text{GUT}}$, we find that today

$$|1 - \Omega_T(a_0)| = \frac{|1 - \Omega_T(a_2)|}{\Omega_T(a_2)} \frac{a_2^2/a_1^2}{10^{-52}} < \frac{|1 - \Omega_T(a_2)|}{\Omega_T(a_2)} . \quad (5.57)$$

We see that Ω_T should be near unity today unless $\Omega_T(a_2)$ was very small. While this is still a non-trivial requirement, the fact that Ω_T is near unity today is no longer to be considered the result of incredibly fine tuning at the GUT epoch but rather as a “reasonable” requirement on the conditions at the beginning of inflation and on the length of inflation.

5.6 Inflation

We have seen that the horizon problem and the Ω problem can be solved if, during a sufficiently long epoch ($a_2 < a < a_1$), the universe had a roughly constant energy density, as in Fig. 5.11. The constant energy cannot be due to a simple cosmological constant because some of the energy must be transformed into radiation at t_1 . Inflationary scenarios [126] provide a mechanism (or at least a differential equation) that does this.

The original inflationary scenario [127] proposed by Guth supposed that the universe was trapped in a metastable vacuum from which it escaped by quantum or thermal tunneling. This scenario was well-motivated by certain Grand Unified Theories. Unfortunately, the tunneling process, equivalent to a first-order phase transition, generated inhomogeneities during the nucleation process that are considerably larger than those observed.

The second generation of inflationary models use scalar fields to provide an effective vacuum energy. We saw in Sect. 4.6 that a scalar field ϕ characterized by a potential $V(\phi)$ could mimic a vacuum energy if the time and space derivatives are sufficiently small, in which case $\rho \sim V(\phi)$ and $p \sim -V(\phi)$. The field is governed by (4.84)

$$\ddot{\phi} + 3 \frac{\dot{a}}{a} \dot{\phi} + V'(\phi) = 0 . \quad (5.58)$$

An example of a potential that can give an energy density like that of Fig. 5.11 is shown in Fig. 5.13. Potentials of this sort are used in “new inflation” models proposed by Linde and by Steinhardt and Albrecht [128]. The potential is characterized by a flat region for $\phi_2 < \phi < \phi_1$ and then a sharp falloff to a minimum at ϕ_0 .

Inflation is assumed to start at a_2 with the field at ϕ_2 far from the minimum. According to (5.58), the fields “rolls” toward the minimum reaching a terminal velocity of

$$\dot{\phi} = \frac{-V'(\phi)}{3\dot{a}/a} \quad a_2 < a < a_1 . \quad (5.59)$$

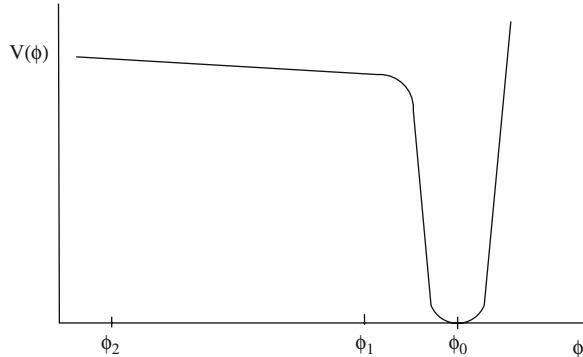


Fig. 5.13 An inflationary potential that yields an energy density similar to that shown in Fig. 5.11. The field is assumed to have an initial value $\phi = \phi_2$ at a_2 . Governed by (5.58), the field “rolls” toward ϕ_1 with a velocity that approaches the terminal velocity $\dot{\phi} = -V'/(3\dot{a}/a)$. After a_1 , the field performs damped oscillations about ϕ_0 before settling down to its minimum. During the oscillations part of the energy can be transformed into particles via appropriate couplings

If this velocity is reached it is said that the “slow-roll” condition is satisfied. Before reaching a_1 , the potential is relatively flat so we have near perfect inflation:

$$\rho(a) \sim V(\phi) \sim \text{constant} \quad a_2 < a < a_1 . \quad (5.60)$$

After falling into the potential well, the field performs damped oscillations about the minimum until it settles down to the minimum. The oscillations are damped because of the $\dot{\phi}$ term in (5.58). Radiation of particles can be accomplished by coupling ϕ to matter fields. This so-called “reheating” of the universe is described phenomenologically by adding a term $\Gamma \dot{\phi}$ to the l.h.s. of (5.58). Here, Γ is the inverse of the damping time from particle radiation. This will lead to a certain fraction of the field energy being converted to radiation energy.

We should emphasize that while the potential in Fig. 5.13 gives an energy density similar to that of Fig. 5.11, it is possible to solve the horizon and \mathcal{Q} problems with any relatively slowly varying potential. One that is often discussed because of its simplicity is

$$V(\phi) = \gamma \phi^4 , \quad (5.61)$$

where γ is a dimensionless constant.² The Friedmann equation in natural units is then

$$H^2 = \frac{8\pi\gamma\phi^4}{3m_{\text{pl}}^2} . \quad (5.62)$$

² ϕ has dimensions of energy, as can be seen in the general form of the energy momentum tensor for a scalar field (4.77).

To solve the horizon and \mathcal{Q} problems we need inflation to last $N \sim 60$ Hubble times. This sets a constraint on the initial value of ϕ_2 :

$$\int_{t_2}^{t_1} H dt \sim \int_{\phi_2}^0 H(\phi) d\phi / \dot{\phi} > 60 . \quad (5.63)$$

Using the slow-roll condition (5.59) we find that the field must be initially displaced from the minimum by a value comparable to the Planck mass:

$$\phi_2 > \sqrt{N} m_{\text{pl}} . \quad (5.64)$$

While the reader may or may not be impressed with this condition, it does provide an important constraint for people who worry about how inflation started.

A constraint on the coupling constant γ can be obtained by considering the quantum fluctuations of the field ϕ . These fluctuations are believed to be the seeds of the density fluctuations that result in the structures of the present-day universe. We will see in Chap. 7 that the present-day density fluctuations on the scale of H_0^{-1} are observed to be of order $\Delta\rho/\rho \sim 2 \times 10^{-5}$. Furthermore, these fluctuations will be shown to be related to the slope of the potential during inflation:

$$\left[\frac{\Delta\rho}{\rho} \right]_{H_0^{-1}} \sim \frac{H^3}{V'(\phi)} . \quad (5.65)$$

For the $\gamma\phi^4$ potential and $\Delta\rho/\rho \sim 10^{-5}$ this gives

$$\gamma \sim \frac{(\Delta\rho/\rho)^2}{N^3} \sim 10^{-15} . \quad (5.66)$$

This means that the ϕ is very weakly self-coupled, a conclusion which is generic for inflationary potentials that have quantum fluctuations of the size necessary to generate the observed density fluctuations.

While the condition (5.66) can certainly be chosen classically, it is difficult to achieve such a weak coupling in a quantum theory since quantum corrections to the potential generally result in couplings comparable to the other couplings, e.g. $\alpha \sim 10^{-2}$. This is a basic problem with the inflationary scenario. While a classical field and potential can be found to solve the horizon and \mathcal{Q} problems and give seeds for the observed inhomogeneities, there is at present no well-motivated quantum theory that satisfies all requirements. As is often the case in cosmology, use can be made here of anthropic selection because, as we will see in Sect. 7.3, fluctuations that have amplitudes that are much larger or smaller than those observed are inefficient in creating galaxies that are hospitable for life as we know it. It is thus possible that the conditions that led to inflation in our Hubble volume are not typical for the universe as a whole, but are typical for observers.

5.7 Intergalactic Scattering and Absorption

As a final application of the Friedmann equation, we will show in this section how to calculate the probability for absorption or scattering of photons in intergalactic space. At the epoch of recombination, the universe became nearly transparent because of the very low scattering cross-section for photons on neutral atoms. Measurements of CMB polarization indicate that photons started scattering again on free electrons that appeared at $\hat{a} \sim 0.1$. It is generally believed that this “reionization” of the universe was due to first quasars and stars which produced enough ultraviolet photons to ionize most atoms.

Consider a photon emitted at t_1 with energy E_1 . We want to calculate the probability for absorption or scattering before it can be detected by us at t_0 . We consider a general reaction that scatters or absorbs photons on a certain type of particle (free electrons or atoms) with cross-section $\sigma(E)$. The probability dP that a photon is scattered or absorbed between t and $t + dt$ is

$$dP = n(t) \sigma(E(t)) c dt , \quad (5.67)$$

where $n(t)$ is the density of scattering or absorbing particles. To simplify matters, we suppose that the number of particles is time independent:

$$n(t) = n_0 \left(\frac{a_0}{a(t)} \right)^3 . \quad (5.68)$$

Furthermore, the energy of the photon in the frame of the absorbing matter is $E(t) = E_1 a_1 / a(t)$ so we find

$$dP = n_0 \left(\frac{a_0}{a(t)} \right)^3 \sigma(E = E_1 a_1 / a(t)) c dt . \quad (5.69)$$

The total optical depth τ can be calculated by integrating dP between t_1 and t_0 . As usual, things are simpler if we make the variable change $t \rightarrow a(t)$. Using the Friedmann equation we find:

$$\tau = d_H n_0 \int_{\hat{a}_1}^1 \frac{\sigma(E = E_1 a_1 / a) d\hat{a}}{\hat{a}^4 [\Omega_M \hat{a}^{-3} + \Omega_\Lambda + (1 - \Omega_T) \hat{a}^{-2}]^{1/2}} . \quad (5.70)$$

For low-energy photons, the cross-section on free electrons is the Thomson cross-section, σ_T , so

$$\tau = d_H n_e \sigma_T \int_{\hat{a}_1}^1 \frac{d\hat{a}}{\hat{a}^4 [\Omega_M \hat{a}^{-3} + \Omega_\Lambda + (1 - \Omega_T) \hat{a}^{-2}]^{1/2}} , \quad (5.71)$$

where $d_{\text{H}} n_e \sigma_{\text{T}} = 2 \times 10^{-3}$ for $n_e \sim 0.87 n_b$. For objects with $z \ll 1$, the integral give a factor z so the probability for scattering is very small. For CMB photons, the scattering starts at reionization when $\hat{a} \ll 1$. The integral is then dominated by the interval near \hat{a}_{reion} giving

$$\tau \sim \frac{2 d_{\text{H}} n_e \sigma_{\text{T}}}{3\sqrt{\Omega_M}} \hat{a}_{\text{reion}}^{-(3/2)} = 0.09 \left(\frac{1 + z_{\text{reion}}}{12} \right)^{3/2} \quad (5.72)$$

It is possible to measure the optical depth for CMB photons from the polarization induced by Compton scattering after reionization. WMAP gives $\tau = 0.087 \pm 0.017$ from which we can deduce the redshift of reionization $z_{\text{reion}} = 11.0 \pm 0.4$.

While most of the baryonic matter became ionized at this epoch, some atomic hydrogen remains, probably in a steady state where recombination is balanced by photo-ionization. The most important reaction on atomic hydrogen is the resonant excitation of the ground state to an excited state. For example, photons with energies equal to $E_{1-2} = (3/4)0.5\alpha^2 m_e c^2 = 10.21 \text{ eV}$ can excite the 2p level, the so-called Lyman- α transition. The cross-section is

$$\sigma(E) = f \frac{\sigma_{\text{T}}}{\alpha^3} E_{1-2} \delta(E - E_{1-2}) \quad f = 0.416, \quad (5.73)$$

where σ_{T} is the Thomson cross-section and $\alpha \sim 1/137$. The expansion of the universe guarantees that a photon emitted with energy $E_1 > E_{1-2}$ will eventually be redshifted to the resonant energy.

Substituting the cross-section (5.73) into (5.70) we find:

$$\tau_{\text{Ly}\alpha} = \frac{\beta f}{\alpha^3} (d_{\text{H}} n_e \sigma_{\text{T}}) \times \left[\frac{1}{\hat{a}^3 [\Omega_M \hat{a}^{-3} + \Omega_A + (1 - \Omega_T) \hat{a}^{-2}]^{1/2}} \right]_{a=a_1 E_1 / E_{1-2}}, \quad (5.74)$$

where $n_e \sim 0.87 n_b$ is the present-day electron density and β is the fraction of electrons that are bound in the ground state of intergalactic hydrogen. Applying this formula to the photons with E_1 just slightly larger than E_{1-2} we find

$$\tau_{\text{Ly}\alpha} = \frac{(9.4 \times 10^2) \beta (z+1)^3}{[\Omega_M (z+1)^3 + \Omega_A + (1 - \Omega_T)(z+1)^2]^{1/2}}, \quad (5.75)$$

where z is the redshift of the source. For $z > 1$, the large numerical factor guarantees that the absorption will be complete unless $\beta \ll 1$.

Figure 6.8 shows the spectrum of a high-redshift quasar. The prominent peak at 580 nm is due to Lyman- α emission by the quasar. Photons blueward of this peak can be absorbed by Lyman- α absorption after being redshifted. Supposing that the emission just blueward of Lyman- α is not much different from the emission just

redward of Lyman- α , we see that the mean absorption of the high-energy photons is of order 0.5. This tells us that the fraction of hydrogen that was not ionized is of order $\beta \sim 10^{-4}$

Exercises

5.1 The integrals (5.7) and (5.8) are trivial for the following three cosmological models:

- | | |
|--|---|
| 1. $\Omega_M = 1$ $\Omega_A = 0$, $\Omega_R \sim 0$
2. $\Omega_M = \Omega_A \sim 0$ $\Omega_R \sim 0$
3. $\Omega_A = 1$ $\Omega_M \sim \Omega_R \sim 0$ | critical, matter – dominated
empty
critical, vacuum – dominated |
|--|---|

For these three models:

- (a) Calculate $a(t)$ and q_0 .
- (b) Calculate, neglecting Ω_R , the age of the universe. Which universe would have its age strongly modified by including radiation?
- (c) Calculate $d_L(z)$ and $d_A(z)$ (neglecting Ω_R) and draw their curves as a function of $\log z$ between $z = 10^{-3}$ and $z = 10^3$.
- (d) Note that for the first model, $d_A(z) \rightarrow 0$ for $z \rightarrow \infty$ (and for $z \rightarrow 0$). Find the redshift for which d_A is maximized and the angular size of an object minimized.

5.2 If all neutrinos have masses less than ~ 1 eV, the Friedmann equation at t_{eq} takes the form appropriate for three massless neutrino species with $n_V = (3/11)n_\gamma$ (per species):

$$\frac{\dot{a}}{a} = H_0 \left(1.668 \Omega_V \hat{a}^{-4} + \Omega_M \hat{a}^{-3} + \Omega_A + (1 - \Omega_T) \hat{a}^{-2} \right)^{1/2}. \quad (5.76)$$

During what period of time should this be a good approximation for the Friedmann equation? Calculate t_{eq} and t_{rec} as a function of Ω_M , Ω_R , and Ω_A considering only “realistic” values of these parameters. Justify any approximations made.

5.3 Calculate the age of an almost critical universe with $\Omega_A = 0$ and $\Omega_T = \Omega_M = 1 + \epsilon$, $\epsilon \ll 1$. Verify that for $\epsilon < 0$ the age is between that of a critical matter-dominated universe and that for an empty universe.

5.4 Calculate the age of a critical universe $\Omega_T = \Omega_M + \Omega_A = 1$ in the limit $\Omega_A \ll \Omega_M$. Verify that for $\Omega_A > 0$, the age is larger than the age of a universe with $\Omega_A = 0$.

5.5 The number of galaxies dN with radial coordinates between χ_1 and $\chi_1 + d\chi$, in the solid angle $d\Omega$, and at time t_1 is

$$dN = n_1 a_1 d\chi a_1^2 r_1^2 d\Omega , \quad (5.77)$$

where n_1 is the density of galaxies at t_1 . If we suppose that the number of galaxies is conserved, $n_1 a_1^3 = n_0 a_0^3$ this expression becomes

$$dN = n_0 a_0 d\chi a_0^2 r_1^2 d\Omega . \quad (5.78)$$

Use (5.19) to show that

$$\frac{dN}{dz d\Omega} = n_0 H_0^{-3} z^2 [1 - 2(q_0 + 1)z + \dots] . \quad (5.79)$$

This relation shows how the redshift distribution of galaxies can be used to determine q_0 . In practice, this relation is extremely difficult to use because as we emphasized in Sect. 2.2, the number of galaxies diverges logarithmically for low luminosities. It is therefore necessary to select galaxies above a certain intrinsic luminosity. Unfortunately, the intrinsic luminosities of galaxies are known to evolve with time and, therefore, with redshift. It is thus difficult to separate luminosity evolution effects from the q_0 effect.

5.6 Show that the relativistic generalization of the Sunyaev-Zel'dovich (SZ) relation (2.33) is

$$a_0 \chi_1(z) \sim \frac{P^2 c \alpha \Delta \theta \sqrt{m_e T}}{4\pi \phi_X \sigma_T (1+z)^3} \frac{\langle n_e^2 \rangle}{\langle n_e \rangle^2} . \quad (5.80)$$

The use of (5.19) then allows one to determine q_0 once the quantities on the right are measured.

5.7 Generalize the discussion of the expanding photosphere method in Sect. 2.6.1 to show that

$$a_0 \chi_1(z) = \frac{vt}{(1+z)^2} \left(\frac{\sigma T^4}{\phi} \right)^{1/2} , \quad (5.81)$$

where v is the photosphere expansion velocity, t is the observed time since the explosion, ϕ is the observed energy flux, and T is the surface temperature. The use of (5.19) then allows one to determine q_0 .

5.8 Consider a closed universe with $\Omega_T = \Omega_M > 1$.

- (a) Calculate the present horizon $\chi_{\text{hor}} = \chi_1(z = \infty)$.
- (b) What fraction of the universe is presently within the horizon ($\chi < \chi_{\text{hor}}$)? Find a simple expression for the case $(\Omega_T - 1) \ll 1$.
- (c) Show that at the moment of maximum expansion horizon includes the entire universe: $\chi_{\text{hor}}(a_{\text{max}}) = \pi$.
- (d) Show that the age of the universe at maximal expansion is

$$t(a_{\text{max}}) = H_0^{-1} \frac{\pi \Omega_T}{2(\Omega_T - 1)^{3/2}} \quad \Omega_M = \Omega_T > 1 . \quad (5.82)$$

- (e) Show that the age of a radiation-only universe at maximal expansion is

$$t(a_{\text{max}}) = H_0^{-1} \frac{1}{\Omega_T - 1} \quad \Omega_R = \Omega_T > 1 . \quad (5.83)$$

5.9 Consider an open universe ($\Omega_T \leq 1$). Show that if $\Omega_\Lambda = 0$ all of the universe will eventually come within our horizon. Show that if $\Omega_\Lambda > 0$ the horizon will approach a finite limit. This shows that a universe with $\Omega_\Lambda > 0$ has an “event horizon” beyond which we will never see.

5.10 Consider an open universe ($\Omega_T < 1$) with $\Omega_\Lambda = 0$. Show that

$$\lim_{z \rightarrow \infty} r(z) = \sinh(\chi_{\text{hor}}) = \frac{2\sqrt{1 - \Omega_T}}{\Omega_T} \quad \Omega_\Lambda = 0 . \quad (5.84)$$

Show that the angular distance for large redshift is

$$\lim_{z \rightarrow \infty} d_A(z) = \frac{2H_0^{-1}}{z\Omega_T} \quad \Omega_\Lambda = 0 . \quad (5.85)$$

5.11 Calculate $\Delta\theta_s$ for $\Omega_T = 1$ in the limit $\Omega_\Lambda \ll \Omega_M$. Explain why $\Delta\theta_s$ in Fig. 5.8 is nearly independent of the relative values of Ω_M and Ω_Λ if $\Omega_T = 1$.

5.12 In the 18th century, Olbers and Ch  seaux noted that if the universe is homogeneous in space *and* in time, the integrated starlight must be infinite. To reconstruct their reasoning, consider a static universe composed of fixed objects of number density n_0 and time-independent luminosity per object L .

- (a) Consider the objects inside a solid angle $d\Omega$ as viewed from the origin. What is the total photon flux at the origin from these objects? This is the “Olbers’ paradox.”
- (b) Explain why absorption by intervening matter will not resolve the paradox.
- (c) Suppose that the universe has existed only for a time t_0 . What is the total flux? This resolves the paradox in an ad hoc manner.
- Consider now the standard (20th century) cosmological model. To simplify things, we suppose that the number of galaxies is constant in time:

$$n(t) = n_0 \left(\frac{a_0}{a} \right)^3 .$$

We additionally suppose that their luminosity L is time independent.

- (d) What is the number of galaxies in the solid angle $d\Omega$ with radial coordinates between χ_1 and $\chi_1 + d\chi_1$? What is the flux from these objects?
- (e) What is the total flux from objects in the solid angle $d\Omega$. Express the flux as an integral over the emission time t_1 . Express the flux as an integral over the scale factor $a(t_1)$. Evaluate the integral in the cases $\Omega_M = \Omega_T = 1$ and $\Omega_M = \Omega_T = 0$. Compare the resulting total flux with the flux in part c.
- (f) In an inflationary cosmology with only photons, the CMB photons can be considered to have been produced at the end of the inflationary epoch by the transformation of vacuum energy into radiation. To calculate the flux of these photons, we can replace the factor of $n(t)L$ by $\rho_A \delta(t - t_1)$. Calculate the resulting flux and comment on the result.
- (g) Calculate the photon flux that we will receive in the far future. Compare the results for $\Omega_M = \Omega_T = 1$, $\Omega_M = \Omega_T = 0$ and the currently preferred model $\Omega_M = 0.3$, $\Omega_A = 0.7$.

5.13 (Continuation of Exercise 3.10) An explorer leaves our galaxy at t_0 with a velocity $v \ll c$ after having used all her fuel. Using the equation of motion found in Sect. 3.7, calculate the redshift of the most distant galaxy that can be visited in the cases ($\Omega_T = \Omega_M = 1$), ($\Omega_T = \Omega_M = 0$), and ($\Omega_T = \Omega_A = 1$).

5.14 We can suppose that at some future time t , technology will have advanced to the point where we can leave the Galaxy with a velocity $v \sim c$. Calculate the redshift of the most distant galaxy that can be visited in the cases ($\Omega_T(t) = \Omega_M(t) = 1$), ($\Omega_T(t) = \Omega_M(t) = 0$), and ($\Omega_T(t) = \Omega_A(t) = 1$). Interpret the result for a vacuum-dominated universe.

5.15 Consider a galaxy at radial coordinate χ . The galaxy is observed by us at t_0 and then again later at $t_0 + \Delta t_0$ corresponding to photons emitted t_1 and $t_1 + \Delta t_1$. Calculate the difference between the redshift measured t_0 and the redshift measured at $t_0 + \Delta t_0$ (to first order in $H_0 \Delta t_0 \ll 1$).

Chapter 6

The Thermal History of the Universe

These days, not much goes on in intergalactic space. There are no nuclear reactions occurring and photons and neutrinos very rarely scatter on matter. More quantitatively, the number of reactions per particle per unit time, Γ , is much less than one reaction per Hubble time. For example the rate of Compton scattering per photon is

$$\Gamma_{\gamma e \rightarrow \gamma e} = n_e \sigma_T c \sim 1.4 \times 10^{-3} H_0, \quad (6.1)$$

which means that only one photon out of 700 will scatter in the next Hubble time. As long as the universe continues to expand, reactions will become rarer and rarer as the density decreases. In fact, we will see that the typical photon will *never* scatter again.

Things were quite different in the early universe. Just before recombination, $\hat{a}_{\text{rec}} \sim 10^{-3}$, the density of electrons was 10^9 times the present density whereas the expansion rate was only $\sqrt{\Omega_M} 1000^{3/2} \sim 2 \times 10^4$ times the present rate:

$$\Gamma_{\gamma e \rightarrow \gamma e}(t_{\text{rec}}) \sim n_e(t_{\text{rec}}) \sigma_T c \sim 80 H(t_{\text{rec}}). \quad (6.2)$$

At this epoch, a typical photon suffered 80 collisions per Hubble time.

We will see that the thermal spectrum of photons resulted from the high reaction rate in the early universe. Elastic scattering, e.g.



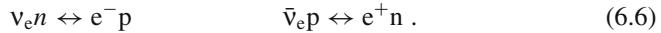
caused energy exchanges between particles and generated “kinetic” equilibrium, i.e. a thermal momentum distribution. Inelastic collisions, e.g.



changed the number of particles and generated “chemical” equilibrium where the particle densities have thermal values. Important examples for this chapter will be the elementary reactions



In the early universe, these reactions generated thermal (blackbody) densities of electron–positron pairs and neutrinos. Also important are the neutron–proton transitions:



These reactions established chemical equilibrium between protons and neutrons, thus determining the number of neutrons available for nucleosynthesis.

We will see that the minimal requirement for the establishment of thermal equilibrium is that the reaction rate per particle be greater than the expansion rate:

$$\Gamma \gg \frac{\dot{a}}{a} \quad \Rightarrow \quad \text{thermal equilibrium.} \quad (6.7)$$

The expansion rate is the relevant parameter because its inverse, t_H , gives the characteristic time for temperature and density changes due to the universal expansion. Collisions can therefore perform the necessary readjustments of momentum and chemical distributions only if each particle reacts at least once per Hubble time.

Because of the expansion, the collision epoch was bound to end when the reaction rate became less than the expansion rate, $\Gamma \ll \dot{a}/a$. What happens to the thermal distributions once the collisions cease depends on the type of equilibrium. For purely kinetic equilibrium, i.e. the momentum spectrum of particles, the thermal character of the spectrum may be maintained by the expansion. This is the case for the CMB photons.

On the other hand, chemical thermal equilibrium is maintained only by reactions and the equilibrium is lost once the reactions cease. This is the case for nuclei. At the present low temperature, chemical equilibrium would imply that nucleons would tend to be in their most bound states near ^{56}Fe . This is not the case because the nuclear reactions necessary to reach this state ceased when the temperature was much higher. Most nucleons were thus “stranded” in hydrogen and helium. We say that the nuclear reactions “froze” at a temperature T_f that turns out to be $T_f \sim 30\text{ keV}$. The “freeze-out” left the universe with a “relic” density of hydrogen and helium nuclei that is far from the equilibrium density.

The nuclear freeze-out left the universe with a reserve of free energy, i.e. energy that can now be degraded by entropy producing (exothermic) nuclear fusion reactions. In particular, hydrogen can be converted to helium and helium to heavier elements once matter is gravitationally confined in stars. Fusion reactions in stellar interiors transform mass into kinetic energy of the reaction products which is then degraded to thermal energy including a multitude of thermal photons. It is this increase in the number of photons which is primarily responsible for the entropy increase.¹

¹ Entropy is always at least approximately proportional to the number of particles. See Table 6.1.

After the photons escape from a star, entropy production can continue if the photons are intercepted by a cold planetary surface. On Earth, solar photons ($T \sim 6000$ K) are multiplied into ~ 20 thermal photons ($T \sim 300$ K). The accompanying entropy increase more than compensates for the entropy decrease associated with the organization of life induced by photosynthesis. Without the thermal gradient between the Sun and Earth, photosynthesis would not be possible because of the second law of thermodynamics. We see that the loss of thermal nuclear equilibrium in the early universe provides the free energy necessary for life on Earth. Without this energy source, life would depend on the photons from gravitational collapse (Exercise 2.9), a source that is more rapidly exhausted.

The explanation for the current thermal disequilibrium is one of the greatest triumphs of modern cosmology. Nineteenth century physicists were puzzled by the disequilibrium because they knew that all isolated systems tend toward thermal equilibrium. They also worried about the future “heat death” of the universe when equilibrium will be reached, terminating all intelligent activity. Modern cosmology appears to have inverted the sequence of events since the state of chemical thermal equilibrium occurred in the past rather than the future.

On the other hand, cosmology has introduced a new thermal equilibrium problem by including the effects of gravity. Photons are generally radiated when gravitationally bound objects are formed. An efficient way to increase the entropy of the universe is therefore through the binding of intergalactic or interstellar matter into stars. This is in fact what drives the increase in entropy today. A question that then arises is why did the universe start in a nearly homogeneous (low-entropy) state. The dynamical answer is inflation, but this only pushes the question back one step: how did inflation start when it corresponds to such a low entropy state. Attempts to explain this involve pre-inflation physics that is obviously very speculative.

Most of this chapter will be concerned with the transition from collisional epochs to non-collisional epochs and the resulting relic densities of different species. Precise calculations generally require complicated computer codes but we will generally limit ourselves to qualitative arguments so as to understand the basic physics.

The first two sections introduce the necessary tools. In Sect. 6.1 we will present the particle phase space distributions that describe the thermal state of the universe. In Sect. 6.2 we will show how the phase space distributions are governed by the Boltzmann equation coupled to the Friedmann equation.

The following sections will apply these concepts to electrons and positrons, neutrinos, nuclei, and WIMPs. (Atomic recombination will be treated in Exercise 6.5.) We will see that all these species were originally in thermal equilibrium implying that, for temperatures much greater than their masses, their number densities were comparable to that of photons. As the reaction rate dropped, each species dropped out of thermal equilibrium leaving a relic density of particles present today. We will see that the standard cosmology allows us to calculate the relic densities. In particular, the neutrino–photon ratio will be calculated “exactly,” the helium–hydrogen ratio will be calculated as a function of the baryon–photon ratio, and the WIMP–photon ratio will be calculated as a function of its (unknown) annihilation cross-section.

For temperatures much greater than $\sim \text{GeV}$, baryons and antibaryons in the form of quarks and antiquarks were present in numbers comparable to that of photons. The small relic density of baryons is due to the slight excess of baryons over antibaryons at this epoch. Possible sources of this asymmetry will be discussed in Sect. 6.7.

In Sect. 6.8 we will discuss the extent to which the expansion is reversible and identify processes that allow us to identify the thermodynamical arrow of time. Finally, in Sect. 6.9 we will present a very general discussion of what the thermal future might be like.

6.1 Equilibrium Distributions

In this chapter we will study the phase-space distribution for each particle species $i = \gamma, e^-, \dots$:

$$F_i(\mathbf{r}, \mathbf{p}, t) = \frac{dN_i}{d^3r d^3p}. \quad (6.8)$$

For some purposes it is convenient to renormalize the distribution to give the dimensionless density:

$$f_i(\mathbf{r}, \mathbf{p}, t) = \frac{F_i}{g_i/(2\pi\hbar)^3}, \quad (6.9)$$

where g_i is the number of spin states of the species. The normalization of f_i is such that $f_i = 1$ corresponds to the maximum phase-space density allowed by the Pauli principle for a fermion, $F_i = g_i/(2\pi\hbar)^3$. As usual, we will no longer write explicitly the factors of \hbar (nor of c and k).

We will only be concerned with local effects so \mathbf{r} and t can be taken to be those of our freely falling coordinate system. The assumption of homogeneity assures us that the conclusions we will reach for the conditions at the origin will apply to the conditions seen by all comoving observers. For the rest of this chapter we will generally suppress $\mathbf{r} = 0$ in the arguments of the distribution functions.

Apart from the constraint of the Pauli principle for fermions, the distributions f_i are a priori arbitrary. Of special importance, however, are the thermal equilibrium distributions:

$$f_i(\mathbf{p}_i) = \frac{1}{\exp[(E_p - \mu_i)/T] \pm 1}, \quad (6.10)$$

where T and μ_i are the temperature and chemical potential and where the sign $+$ ($-$) corresponds to fermions (bosons). In cosmological applications, the chemical potential is generally determined by the number of particles (if this number is fixed) or by the difference in the number of particles and antiparticles (see Exercise 6.1).

The number density n_i , the energy density ρ_i , and the partial pressure p_i are determined by f_i

$$n_i = \frac{g_i}{(2\pi)^3} \int d^3 p f_i(\mathbf{p}), \quad (6.11)$$

$$\rho_i = \frac{g_i}{(2\pi)^3} \int d^3 p E_p f_i(\mathbf{p}), \quad (6.12)$$

$$p_i = \frac{g_i}{(2\pi)^3} \int d^3 p \frac{|\mathbf{p}|^2}{3E} f_i(\mathbf{p}). \quad (6.13)$$

The definitions of ρ and p are consistent with those given in Chap. 4 in the context of the energy-momentum tensor.

In thermal equilibrium, f_i is given by (6.10), so the three quantities are functions of T and μ . In the relativistic and non-relativistic limits, it is easy to perform the necessary integrations (Exercise 6.1) with all but the numerical factors determined by dimensional analysis. The results are listed in Table 6.1 for the two limits that are most interesting for cosmology. The first is the relativistic limit ($T \gg m$) with ($T \gg \mu$). We consider only the case $\mu \sim 0$ because it corresponds to a number of particles nearly equal to the number of antiparticles as was the case, e.g. for electrons and positrons when $T \gg m_e$. In Table 6.1 we note the presence of the usual factors of T^4 and T^3 for the energy and number densities of blackbodies. The factors of (7/8) and (3/4) are due to the Pauli principle for fermions (sign + in (6.10)). The factors are less than unity because the Pauli principle requires $f_i \leq 1$.

The second case in Table 6.1 is the case of a non-relativistic dilute gas: ($m - \mu \gg T$) which implies $f \ll 1$. In this case, the Boltzmann factor $\exp(-m/T)$ drastically lowers the number of particles. If $\mu = 0$, there are no particles or antiparticles present if the temperature vanishes, $n(T = 0, \mu = 0) = 0$.

In thermal equilibrium, the entropy density s is given by (Exercise 6.2):

$$s(T, \mu) = \frac{\rho + p - \mu n}{T}. \quad (6.14)$$

As seen in Table 6.1, the entropy density is proportional to the number of particles. Because their numbers are not suppressed by a Boltzmann factor, relativistic species generally dominate the entropy density. Today it is dominated by photons and

Table 6.1 The equilibrium thermodynamical functions (6.11), (6.12), (6.13), and (6.14) in two limiting cases. The first is the relativistic limit ($m \ll T$) with an equal number of particles and antiparticles, i.e. $\mu \ll T$. The second is the dilute, non-relativistic limit ($m \gg T$) and $m - \mu \gg T$, implying $f \ll 1$

Function	$m \ll T, \mu \ll T$		$m - \mu \gg T$
$n(T, \mu)$	$g(1.2/\pi^2)T^3$	($\times 3/4$ fermions)	$g(mT/2\pi)^{3/2} e^{(\mu-m)/T}$
$\rho(T, \mu)$	$g(\pi^2/30)T^4$	($\times 7/8$ fermions)	$(m + 3T/2)n$
$p(T, \mu)$	$\rho/3$		Tn
$s(T, \mu)$	$g(2\pi^2/45)T^3$	($\times 7/8$ fermions)	$(m + 5T/2 - \mu)n/T$

neutrinos. At higher temperatures, it was dominated by particles in thermal equilibrium with $m < T$. Supposing a unique temperature T and vanishing chemical potentials, the entropy density is given by

$$s(T) = g_s(T) \frac{2\pi^2}{45} T^3, \quad (6.15)$$

where the effective number of spin states is given by the following sum over all relativistic particles:

$$g_s(T) \sim \sum_{\text{bosons}} g_i \theta(T - m_i) + (7/8) \sum_{\text{fermions}} g_i \theta(T - m_i), \quad (6.16)$$

where θ is the Heaviside step function. The function $g_s(T)$ is shown in Fig. 6.1 for the known particles listed in Table 6.2.

We will see that the approximation of a unique temperature is not good for $T < 1 \text{ MeV}$ when the neutrinos have a lower temperature than the photons $T_\gamma = (4/11)^{1/3} T_\gamma$. For low temperature, the effective number of spin states for photons and three neutrino species and their antineutrinos is therefore

$$g_s(T_\gamma \ll m_e) = 2 + 6 \times (7/8)(T_\gamma/T_\gamma)^3 = 3.91. \quad (6.17)$$

To calculate g_s it is important to remember that, for neutrinos and antineutrinos, only one of the two possible helicity states interacts at full strength and can therefore be expected to be in thermal equilibrium. The possible presence of the non-interacting helicity states is discussed in Exercise 6.8.

During the radiation epoch, the energy density is dominated by relativistic species. For $T > 1 \text{ MeV}$, there is a unique temperature and the energy density is

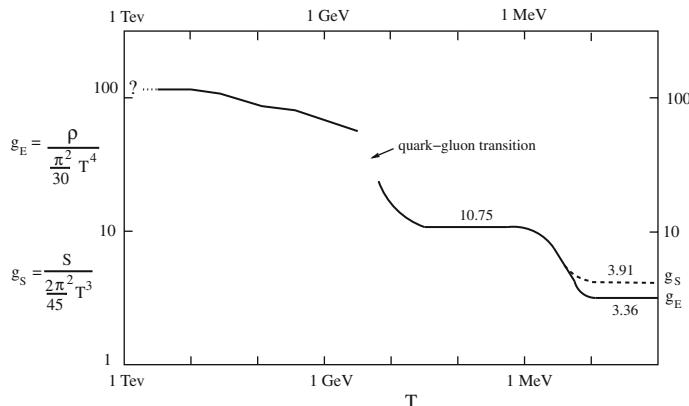


Fig. 6.1 The effective number of spin states, g_E and g_s , in equilibrium as a function of temperature. Only the known particles shown in Table 6.2 have been included. For $T < 1 \text{ MeV}$, $T_\gamma \neq T_\gamma$ implying $g_s \neq g_E$

Table 6.2 The known particle species expected to be in thermal equilibrium in the early universe [18]. The mass gives the minimum temperature for which the species is relativistic, except for the light quarks and gluons which are expected to exist as free particles only above the quark-gluon phase transition temperature ~ 400 MeV. Below this temperature quarks are bound in hadrons, mostly pions with small numbers of kaons, nucleons, and other excited hadrons. The effective number of spin states, g , is the product of 3 factors. The first factor is the number of subspecies including antiparticles, e.g. 1 photon and $6 = 3\nu + 3\bar{\nu}$. The second factor is the number of interacting spin states per subspecies, 1 for neutrinos, 2 for other spin 1/2 fermions, 2 for massless spin 1 bosons and 3 for massive spin 1 bosons. Finally, the factor (7/8) is due to the Pauli principle for fermions

Particle	Mass	Charge	Spin	g
Photon γ	0	0	1	$1 \times 2 = 2$
Neutrinos ν_e, ν_μ, ν_τ	< 10 eV	0	1/2	$6 \times 1 \times (7/8) = 5.25$
Charged leptons				
e	0.511 MeV	1	1/2	$2 \times 2 \times (7/8) = 3.5$
μ	105.66 MeV	1	1/2	$2 \times 2 \times (7/8) = 3.5$
τ	1777.05 MeV	1	1/2	$2 \times 2 \times (7/8) = 3.5$
Quarks (3 colors)				
d	~ 0	-1/3	1/2	$6 \times 2 \times (7/8) = 10.5$
u	~ 0	2/3	1/2	$6 \times 2 \times (7/8) = 10.5$
s	~ 170 MeV	-1/3	1/2	$6 \times 2 \times (7/8) = 10.5$
c	~ 1.3 GeV	2/3	1/2	$6 \times 2 \times (7/8) = 10.5$
b	~ 4.3 GeV	-1/3	1/2	$6 \times 2 \times (7/8) = 10.5$
t	~ 170 GeV	2/3	1/2	$6 \times 2 \times (7/8) = 10.5$
Gluons g	0	0	1	$8 \times 2 = 16$
Weak bosons				
W	80.41 GeV	1	1	$2 \times 3 = 6$
Z	91.187 GeV	0	1	$1 \times 3 = 3$

$$\rho(T) = g_E(T) \frac{\pi^2}{30} T^4 \quad g_E(T) = g_s(T) \quad (T > m_e). \quad (6.18)$$

For $T < m_e$ the temperature difference between photons and neutrinos modifies $g_E(T)$ in a way such that $g_E(T) \neq g_s(T)$:

$$g_E(T \gamma \ll m_e) = 2 + 6 \times (7/8)(T_V/T_\gamma)^4 = 3.36. \quad (6.19)$$

The importance of the entropy density $s(T, \mu)$ is that if the universe is in thermal equilibrium, the expansion is, to a good approximation, adiabatic. To see how this arises, we note that in Sect. 4.3 it was shown that local energy conservation implies

$$d(\rho a^3) = -p da^3. \quad (6.20)$$

For a comoving volume (the region $\chi < \chi_1 \ll 1$), this equation is equivalent to $dE = -pdV$ where V and E are the volume of and the energy in the sphere. The entropy S in the comoving volume varies according to the fundamental relation

$$dS = \frac{dE + pdV - \sum_i \mu_i dN_i}{T}, \quad (6.21)$$

where N_i is the number of particles of species i in the volume. The relation is equivalent to

$$d(sa^3) = \frac{d(\rho a^3) + p da^3 - \sum_i \mu_i d(n_i a^3)}{T}. \quad (6.22)$$

The first two terms cancel giving

$$\frac{ds a^3}{da^3} = - \sum_i \frac{\mu_i}{T} \frac{dn_i a^3}{da^3}. \quad (6.23)$$

Supposing that the chemical potentials are negligible, i.e. negligible particle-antiparticle asymmetries, it follows that

$$s(T) a^3 = g_s(T) T^3 a^3 = \text{constant}. \quad (6.24)$$

For the temperature ranges where $g_s(T)$ is constant, (6.24) implies

$$T \propto \frac{1}{a}, \quad (6.25)$$

giving the same dependence of T on $a(t)$ as in the absence of collisions.

6.2 The Boltzmann Equation

In the absence of collisions, particles freely propagate in phase space according to the Liouville equation. In Sect. 1.2.4 we showed that in a homogeneous universe, the Liouville equation for a massless species i is:

$$\frac{\partial F_i}{\partial t} = \frac{\dot{a}}{a} \mathbf{p} \cdot \nabla_p F_i \quad \Rightarrow \quad a \frac{\partial F_i}{\partial a} = \mathbf{p} \cdot \nabla_p F_i. \quad (6.26)$$

In fact, it can be shown that this equation also applies to massive particles (Exercise 6.3). The general solution is

$$F(\mathbf{p}, a) = F(\mathbf{p}a/a_0, a_0), \quad (6.27)$$

which means that momentum distributions are redshifted during the expansion by a factor a_0/a .

The evolution of the number density n_i in the absence of collisions is found by integrating (6.26) over momentum:

$$\frac{dn_i(t)}{dt} = -3 \frac{\dot{a}}{a} n_i \quad \Rightarrow \quad \frac{dn_i a^3}{dt} = 0, \quad (6.28)$$

with the solution

$$n_i = n_i(a_0) \left(\frac{a_0}{a} \right)^3. \quad (6.29)$$

As expected, the number density of particles falls as $1/a^3$.

In the presence of collisions, particles move in phase space not only continuously by free propagation but also discontinuously by discrete scatters. We want to find the generalizations of (6.26) and (6.28) that take into account elastic and inelastic reactions. To simplify things, we consider only reactions of the type



where i , j , k , and l refer to particle species. We will start with the generalization of (6.28) by writing symbolically

$$\frac{dn_i}{dt} = -3 \frac{\dot{a}}{a} n_i - c(ij \rightarrow kl) + c(kl \rightarrow ij), \quad (6.31)$$

where the two terms we have added take into account the loss and gain of i particles by reactions. We expect that these terms are proportional to the densities of the initial state particles:

$$c(kl \rightarrow ij) \equiv n_k n_l \gamma_{kl \rightarrow ij} \quad c(ij \rightarrow kl) \equiv n_i n_j \gamma_{ij \rightarrow kl}. \quad (6.32)$$

The constants of proportionality, γ , have dimensions of volume/time and are usually interpreted as “mean cross-section times relative velocity”:

$$\gamma_{kl \rightarrow ij} = \langle \sigma_{kl \rightarrow ij} v_{kl} \rangle \quad \gamma_{ij \rightarrow kl} = \langle \sigma_{ij \rightarrow kl} v_{ij} \rangle. \quad (6.33)$$

This is an intuitively appealing interpretation since we would expect the reaction rate to be proportional to the cross-section and to the relative velocities. (If the particles do not move they cannot collide.) The interpretation is especially clear for head-on collisions if we ignore identical particle effects (Pauli principle and stimulated emission). It should be emphasized, however, that quantum mechanical calculations give directly γ . The cross-sections are found by dividing γ by the normalization volume and the relative velocity.

It is useful to define the reaction rate per particle i

$$\Gamma_{ij \rightarrow kl} \equiv n_j \langle \sigma_{ij \rightarrow kl} v \rangle. \quad (6.34)$$

We note that $\Gamma_{ij \rightarrow kl} \neq \Gamma_{ji \rightarrow kl}$. This is simply because the reaction rate per particle i is proportional to the number of particles j and vice versa.

Equation (6.31) is now

$$\frac{dn_i}{dt} = -3 \frac{\dot{a}}{a} n_i + n_k \Gamma_{kl \rightarrow ij} - n_i \Gamma_{ij \rightarrow kl} \quad (6.35)$$

$$= -3 \frac{\dot{a}}{a} n_i + n_k n_l \langle \sigma_{kl \rightarrow ij} v \rangle - n_i n_j \langle \sigma_{ij \rightarrow kl} v \rangle. \quad (6.36)$$

The three terms in this equation describe the three effects that change n_i : the expansion of the universe, creation of i particles, and destruction of i particles. For example, if we ignore the first two terms in (6.35), the solution is $n_i \propto \exp(-t/\tau_i)$ where $\tau_i = 1/\Gamma(ij \rightarrow kl)$ is the mean lifetime of i particles by $(ij \rightarrow kl)$.

In the introduction to this chapter, we argued that equilibrium is obtained if the reaction rates are much greater than the expansion rates. The presence of these rates in (6.35) gives us a preliminary indication of how this will be obtained. If the expansion rate is much greater than the reaction rates, the first term dominates. In the opposite condition, the collision terms will drive the system to thermal equilibrium, as we will discover soon.

The expansion term can be combined with the scattering terms in (6.35) by calculating $d(na^3)/dt$ and then $d(na^3)/da^3$:

$$\frac{dn_i a^3}{da^3} = 1/3 \left(\frac{\dot{a}}{a} \right)^{-1} (n_k \Gamma_{kl \rightarrow ij} - n_i \Gamma_{ij \rightarrow kl}) \quad (6.37)$$

$$= 1/3 \left(\frac{\dot{a}}{a} \right)^{-1} (n_k n_l \langle \sigma_{kl \rightarrow ij} v \rangle - n_i n_j \langle \sigma_{ij \rightarrow kl} v \rangle). \quad (6.38)$$

Equation (6.38) is the required generalization of (6.28). We will use it in the next sections to study the evolution of the densities of electron–positron pairs, neutrinos, nuclei, and WIMPs.

Before leaving this section, it is useful to find the generalization of the Liouville equation (6.26). As with the generalization of (6.28), we write formally

$$\frac{\partial F_i(\mathbf{p}_i, t)}{\partial t} = \frac{\dot{a}}{a} \mathbf{p}_i \cdot \nabla_{\mathbf{p}} F_i(\mathbf{p}_i, t) - C(ij \rightarrow kl) + C(kl \rightarrow ij). \quad (6.39)$$

The collision term $C(ij \rightarrow kl)$ is found by summing over all ways to lose a particle i of momentum \mathbf{p}_i :

$$C(ij \rightarrow kl) = F_i(\mathbf{p}_i) \int d^3 p_j F_j(\mathbf{p}_j) \int g_k d^3 p_k g_l d^3 p_l \frac{d\sigma_{ij}}{d^3 p_k d^3 p_l} v_{ij} \times (1 \mp f_k(\mathbf{p}_k))(1 \mp f_l(\mathbf{p}_l)). \quad (6.40)$$

The integrals over the momenta of particle types j , k , and l and the factors g_j , g_k , and g_l , perform the required sum. The factors $F_i(\mathbf{p}_i)$ and $F_j(\mathbf{p}_j)$ reflect the fact that the rate for $ij \rightarrow kl$ is proportional to the number of particles i and j initially present. The factors of $(1 \mp f(\mathbf{p}))$ enforce the Pauli principle (sign – for fermions) or allow for stimulated emission (sign + for bosons).

The collision term $C(kl \rightarrow ij)$ is clearly

$$\begin{aligned} C(kl \rightarrow ij) = & \int F_k(\mathbf{p}_k) d^3 p_k F_l(\mathbf{p}_l) d^3 p_l \frac{d\sigma(kl \rightarrow ij)}{d^3 p_i d^3 p_j} v_{kl} g_j d^3 p_j \\ & \times (1 \mp f_i(\mathbf{p}_i))(1 \mp f_j(\mathbf{p}_j)). \end{aligned} \quad (6.41)$$

The Boltzmann equation is simplified if we use the fact that, to a good approximation, particle interactions respect time reversal invariance. We expect, therefore, that a movie of an individual scatter looks as physical as the movie run in reverse. This implies that the cross-section times velocity for the forward direction must be equal to the cross-section times velocity for the backward direction:

$$v_{ij} \frac{d\sigma(ij \rightarrow kl)}{d^3 p_k d^3 p_l} = v_{jk} \frac{d\sigma(kl \rightarrow ij)}{d^3 p_i d^3 p_j}. \quad (6.42)$$

In this approximation, we can combine the two collision terms to find

$$\begin{aligned} \frac{\partial F_i(\mathbf{p}_i, t)}{\partial t} = & \frac{\dot{a}}{a} \mathbf{p}_i \cdot \nabla_p F_i(\mathbf{p}_i, t) \\ & + \frac{1}{(2\pi)^3} \int g_j d^3 p_j g_k d^3 p_k g_l d^3 p_l v_{ij} \frac{d\sigma(ij \rightarrow kl)}{d^3 p_k d^3 p_l} \\ & \times [f_k f_l (1 \mp f_i)(1 \mp f_j) - f_i f_j (1 \mp f_k)(1 \mp f_l)]. \end{aligned} \quad (6.43)$$

In the general case, there is one Boltzmann equation for each species i , each equation containing a collision term for each type of reaction $ijk\dots \rightarrow \lambda\mu\nu\dots$. The Boltzmann equations coupled with the Friedmann equation for a series of integral-differential equations that determine the evolution of the universe.

Given the complexity of the Boltzmann equation, it is not surprising that there is no exact solution and that the equations must be integrated numerically to find approximate solutions. This is straightforward because all are first-order differential equations.

There are two limits with interesting approximate solutions that help us understand the numerical solutions. The first limit is that of no collisions ($\Gamma \ll H$) in which case the solution of the Boltzmann equation is given by (6.27) and (6.29).

The second limit is that of many collisions during a Hubble time, in which case we expect that we can, to a first approximation, ignore the \dot{a}/a term in (6.43). In this case, a time-independent solution ($\partial f / \partial t = 0$) is one where, for each momentum combination \mathbf{p}_i , \mathbf{p}_j , \mathbf{p}_k , \mathbf{p}_l conserving momentum and energy, the distributions satisfy

$$f_k f_l (1 \mp f_i)(1 \mp f_j) = f_i f_j (1 \mp f_k)(1 \mp f_l),$$

i.e.

$$\frac{f_k}{1 \mp f_k} \frac{f_l}{1 \mp f_l} = \frac{f_i}{1 \mp f_i} \frac{f_j}{1 \mp f_j}. \quad (6.44)$$

This condition is satisfied if for each species we have:

$$\frac{f_i}{1 \mp f_i} = \exp \left[(E_p - \mu_i)/T \right], \quad (6.45)$$

with

$$\mu_i + \mu_j = \mu_k + \mu_l. \quad (6.46)$$

Equation (6.45) gives the thermal equilibrium Fermi–Dirac or the Bose–Einstein distribution

$$f_i(p_i) = \frac{1}{\exp \left[(E_p - \mu_i)/T \right] \pm 1}, \quad (6.47)$$

where the sign $+$ ($-$) corresponds to fermions (bosons). T is obviously the temperature and the μ_i are the chemical potentials for each species. The condition (6.46) can be generalized in an obvious way for reactions involving more than two particles.

We note that the collision term in the Boltzmann equation (6.43) pushes the distributions toward their equilibrium value. This is because if the f_i are larger (smaller) than the equilibrium value the collision term is negative (positive).

The effect of the chemical constraint (6.46) depends on the type of reaction. Elastic scattering obviously puts no constraint on the chemical potentials. Inelastic reactions give non-trivial constraints. For example, the bremsstrahlung reaction $e^- p \leftrightarrow e^- p \gamma$ requires that the photon chemical potential vanish:

$$\mu_\gamma = 0. \quad (6.48)$$

This means that the equilibrium photon distribution is the Planck distribution. In the same way, the chemical potential of any species whose production is not constrained by the conservation of a quantum number must vanish in thermal equilibrium.

The reaction $e^+ e^- \rightarrow \gamma \gamma$ requires that in thermal equilibrium

$$\mu_{e^-} = -\mu_{e^+}. \quad (6.49)$$

This is also true for any particle x in equilibrium with its antiparticle \bar{x} :

$$\mu_x = -\mu_{\bar{x}}. \quad (6.50)$$

The Bose–Einstein and Fermi–Dirac distributions are not exact solutions of the Boltzmann equation (6.43) because they do not take into account the term proportional to a/a . At the very least, this term makes the temperature time dependent. If all species are approximately in thermal equilibrium, this effect can be included by simply using entropy conservation (6.24) for an adiabatic expansion to calculate the time dependence of the temperature.

Whether or not the collision terms in the Boltzmann equation are sufficiently strong to drive a system to thermal equilibrium depends on the type of equilibrium. We will see in the next section that for chemical equilibrium, it is generally sufficient to have a rate of species-changing reactions that is greater than the expansion rate. On the other hand, the approach to kinetic equilibrium depends on the efficiency of elastic scattering to exchange energy. As an example, consider photons produced by electron–positron annihilation when the temperature is far below the electron mass. These photons have energies that are much greater than the thermal photons. To thermalize, the photons must lose energy by Compton scattering. The time necessary to do this is the subject of Exercise 6.6.

For most applications in early-universe cosmology, the rate of elastic scattering is sufficient to impose kinetic equilibrium. The most interesting applications involve chemical equilibrium between species. To treat these problems, it is necessary to numerically integrate the Boltzmann equation (6.38) along with the Friedmann equation for the early universe

$$\frac{\dot{a}}{a} = \left(\frac{8\pi G\rho}{3} \right)^{1/2}, \quad (6.51)$$

and the energy conservation equation

$$\frac{d\rho a^3}{da^3} = -p, \quad (6.52)$$

where p is the pressure. In kinetic thermal equilibrium, the right-hand-sides of these equations are all function of T and of the μ_i . They can be integrated numerically starting with an initial condition for ρ , p and the μ_i s or equivalently for ρ , p , and the n_i s. At each step in the integration, ρ and the n_i are changed according to their derivatives which permits one to calculate the new values of T , p , and the μ_i . It is clear that in this procedure we are obliged to suppose that the momentum distributions are thermal.

Fortunately, qualitative arguments often make it easy to estimate the results obtained by numerical integration. We will see how this is done in the following sections.

6.3 Electrons and Positrons

As a first application of the Boltzmann equation, we will treat the case of electrons and positrons. These particles are created and destroyed principally by the reaction $e^+e^- \rightarrow \gamma\gamma$. The two Boltzmann equations (6.38) are

$$\frac{dn_e - a^3}{da^3} = \frac{dn_{e^+} a^3}{da^3} = 1/3 \left(\frac{\dot{a}}{a} \right)^{-1} \times (n_\gamma n_\gamma \langle \sigma \gamma \gamma \rightarrow e^+ e^- v \rangle - n_{e^-} n_{e^+} \langle \sigma_{e^+ e^-} \rightarrow \gamma \gamma v \rangle). \quad (6.53)$$

We can eliminate the cross-section for $\sigma_{\gamma \gamma \rightarrow e^+ e^-}$ in (6.53) by using the fact that the thermal equilibrium distributions are such that the rate per unit volume of $e^+ e^- \rightarrow \gamma \gamma$ is equal to the rate for $\gamma \gamma \rightarrow e^+ e^-$:

$$n_\gamma(T)^2 \langle \sigma \gamma \gamma \rightarrow e^+ e^- v \rangle_T = n_{e^-}(T) n_{e^+}(T) \langle \sigma_{e^+ e^-} \rightarrow \gamma \gamma v \rangle_T, \quad (6.54)$$

where the T s mean that densities and mean cross-sections are to be evaluated under conditions of thermal equilibrium at the temperature T . Equation (6.54) is called the “principle of detailed balance.” Supposing that the momentum distributions of the particles are such that we can approximate the averaged cross-sections by their averages in thermal equilibrium, $\langle \sigma v \rangle \sim \langle \sigma v \rangle_T$, (6.53) becomes

$$\frac{d(n_e - a^3)}{da^3} = \frac{d(n_{e^+} a^3)}{da^3} = 1/3 \left(\frac{\dot{a}}{a} \right)^{-1} \times (n_{e^-}(T) n_{e^+}(T) - n_{e^-} n_{e^+}) \langle \sigma_{e^- e^+} \rightarrow \gamma \gamma v \rangle. \quad (6.55)$$

The case of a symmetric universe, $n_{e^-} = n_{e^+} = n_e$ is especially simple:

$$\frac{d(n_e a^3)}{da^3} = 1/3 \frac{n_e(T)^2 - n_e^2}{n_e} \frac{\Gamma_{e^+ e^-} \rightarrow \gamma \gamma}{\dot{a}/a}. \quad (6.56)$$

This equation shows clearly the physics of the Boltzmann equation. If the expansion rate is much greater than the annihilation rate, the left side vanishes and the density decreases as a^{-3} . If the reaction rate is not negligible, then if n_e is greater (smaller) than the equilibrium value $n_e(T)$, the derivative of $n_e a^3$ is negative (positive). It follows that the value of n_e is pushed toward the equilibrium value as long as the proportionality factor $\Gamma / (\dot{a}/a)$ is not too small. There are therefore two limits and two simple solutions. The first limit $\Gamma \ll \dot{a}/a$ corresponds to few reactions per Hubble time in which case we have $d(na^3)/dt \sim 0$ and

$$n \propto a^{-3} \quad (\Gamma \ll \dot{a}/a). \quad (6.57)$$

The second limit is $\Gamma \gg \dot{a}/a$ corresponding to many reactions per Hubble time. In this case the abundances are pushed to their equilibrium values:

$$n = n(T) \quad (\Gamma \gg \dot{a}/a). \quad (6.58)$$

To determine if the electrons and positrons are in chemical equilibrium with the photons, we only need to calculate the thermal equilibrium value of the annihilation rate and compare it with the expansion rate given by the Friedmann equation:

$$\frac{\dot{a}}{a} = \left[\frac{8\pi G\rho}{3} \right]^{1/2} = \left[\frac{8\pi G}{3} g_E(T) \frac{\pi^2}{30} T^4 \right]^{1/2} \sim \sqrt{3g_E(T)} \frac{T^2}{m_{pl}}. \quad (6.59)$$

(Since we are only interested in qualitative results, we are tempted to ignore the numerical factors and we will sometimes do so in this chapter.)

The expansion and annihilation rates are shown in Fig. 6.2. For the annihilation rate, there are two simple limits, $T \gg m_e$ and $T \ll m_e$. For high temperatures we are interested in high center-of-mass energies, $E_{cm} \gg m_e$, where the annihilation cross-section is [129]

$$\sigma_{e^+e^- \rightarrow \gamma\gamma} = \frac{2\pi\alpha^2}{E_{cm}^2} [2 \ln(E_{cm}/m_e) - 1] \quad E_{cm} \gg m_e. \quad (6.60)$$

For low energy, $v \ll c$, the cross-section is proportional to $1/v$ as is often the case for exothermic reactions:

$$\sigma_{e^+e^- \rightarrow \gamma\gamma} v = \frac{\pi\alpha^2}{m_e^2} \quad v \ll c. \quad (6.61)$$

For $T \gg m_e$, the energy factors in the cross-section translate into temperature factors in the mean cross-section and (6.60) gives $\langle \sigma v \rangle \propto T^{-2}$ (ignoring the logarithmic factor). The densities of electrons and positrons are proportional to T^3 and the annihilation rate is therefore proportional to the temperature:

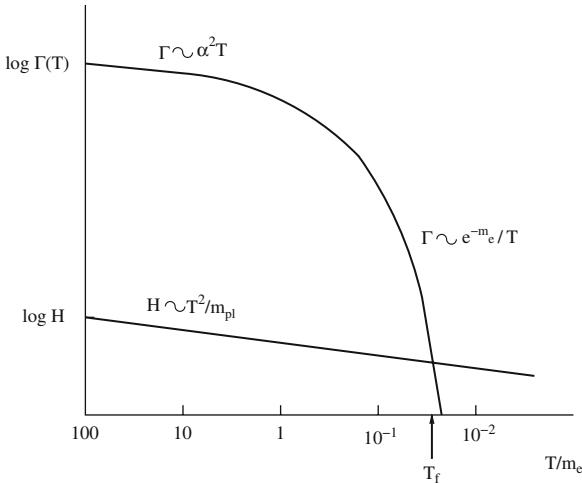


Fig. 6.2 The annihilation rate $\Gamma(e^+e^- \rightarrow \gamma\gamma)$ and the expansion rate $H = \dot{a}/a$ as a function of temperature under conditions of thermal equilibrium and with $n_{e^-} = n_{e^+} = n_e$. For $T > T_f$, $\Gamma > \dot{a}/a$ and n_e will take its equilibrium value. For $T < T_f$, $\Gamma < \dot{a}/a$ and the reactions are “frozen.” After the freeze-out, the number of electrons and positrons is constant so n_e decreases as $1/a^3$

$$\Gamma_{e^+e^- \rightarrow \gamma\gamma} \sim \alpha^2 T \quad T \gg m_e. \quad (6.62)$$

Comparing (6.59) and (6.62), we see that $\Gamma > \dot{a}/a$ for $T < \alpha^2 m_{pl} \sim 10^{14}$ GeV. For example, at $T \sim m_e$, $\Gamma \sim (\alpha^2 m_{pl}/m_e) \dot{a}/a \sim 10^{18} \dot{a}/a$, i.e. 10^{18} reactions per Hubble time. We can conclude that the electrons and positrons were in chemical equilibrium with the photons for $m_e < T < 10^{14}$ GeV.

The equilibrium is inevitably lost for $T \ll m_e$ because in equilibrium the number of electrons and positrons drops with temperature exponentially because of the Boltzmann factor $e^{-m/T}$ (Table 6.1). For the case $n_{e^-} = n_{e^+} = n_e$, the equilibrium annihilation rate is :

$$\begin{aligned} \Gamma_{e^+e^- \rightarrow \gamma\gamma} &= n_e(T) \frac{\pi \alpha^2}{m_e^2} \\ &\sim \alpha^2 \left(\frac{T^3}{8\pi m_e} \right)^{1/2} e^{-m_e/T} \quad (T \ll m_e, \quad n_{e^-} = n_{e^+}). \end{aligned} \quad (6.63)$$

The end of chemical equilibrium comes about roughly when the annihilation rate (6.63) is equal to the expansion rate (6.59). The temperature at which this occurs is called the “freeze-out” temperature T_f because after this temperature is reached the number of electrons and positrons per comoving volume is frozen. Equating the two rates, we find:

$$\sqrt{3g_E(T_f)} \frac{T_f^2}{m_{pl}} = n_e(T) \frac{\pi \alpha^2}{m_e^2} = \alpha^2 \left(\frac{T_f^3}{8\pi m_e} \right)^{1/2} e^{-m_e/T_f}. \quad (6.64)$$

The numerical solution is

$$T_f \sim m_e/40. \quad (6.65)$$

For $T < T_f$, the annihilation rate is smaller than the expansion rate and the Boltzmann equation is simply $d(na^3)/da^3 \sim 0$ implying

$$n_e(a) \sim n_e(T_f) \left(\frac{a_f}{a} \right)^3 \quad a > a_f. \quad (6.66)$$

For $a > a_f$ the number of photons per volume is also proportional to a^{-3} so the ratio between the number of electrons or positrons and the number of photons is constant in time. We can estimate the ratio using (6.65):

$$\begin{aligned} \frac{n_e(a_0)}{n\gamma(a_0)} &\sim \frac{n_e(a_f)}{n\gamma(a_f)} \sim \frac{n_e(a_f)}{T_f^3} \sim \sqrt{g_E(T_f)} \frac{m_e^2}{\pi \alpha^2 m_{pl} T_f} \\ &\sim 2 \times 10^{-16}. \end{aligned} \quad (6.67)$$

This number is about six orders of magnitude smaller than the observed ratio for electrons $n_{e^-} \sim \eta n_\gamma \sim 5 \times 10^{-10} n_\gamma$. This is, of course, due to our assumption of equal numbers of electrons and positrons.

Figure 6.3 shows the ratio between the number of electrons and the number of photons as a function of temperature calculated by numerically integrating the Boltzmann equation, the Friedmann equation (6.51), and the energy conservation equation (6.52). We see that the estimation (6.67) is relatively accurate. We emphasize, however, that we did not pay much attention to numerical factors and these can be accurately estimated only by numerically integrating the Boltzmann equation.

The evolution of n_{e^-} and n_{e^+} in a universe like ours with an excess of electrons is shown in Fig. 6.4. The relic density of electrons is simply determined by charge conservation:

$$a_0^3 n_{e^-}(t_0) = a^3(n_{e^-} - n_{e^+}). \quad (6.68)$$

We can relate this to the photon density by using the fact that as long as thermal equilibrium is maintained, the entropy of the electron–positron–photon system is conserved:

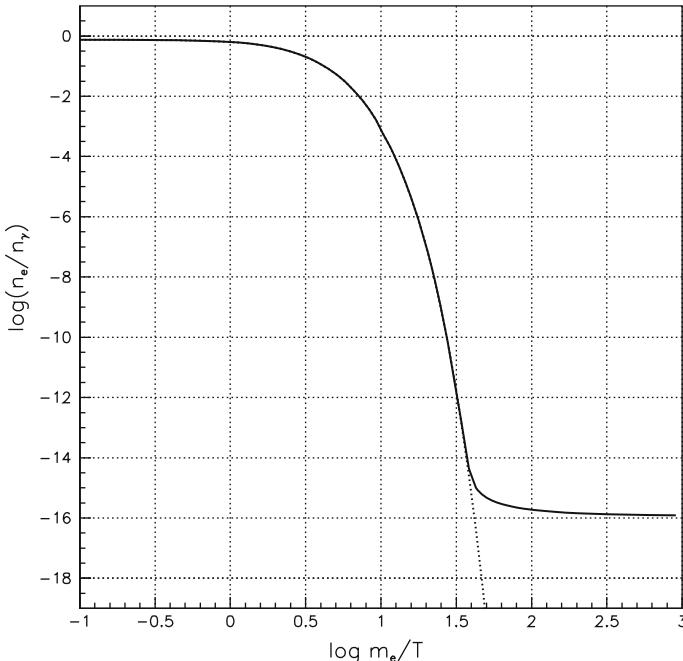


Fig. 6.3 $\log(n_e/n_\gamma)$ versus temperature for the case $n_e = n_{e^-} = n_{e^+}$ found by numerically integrating the Boltzmann equation. The dotted line shows $\log(n_e/n_\gamma)$ in the case of thermal equilibrium

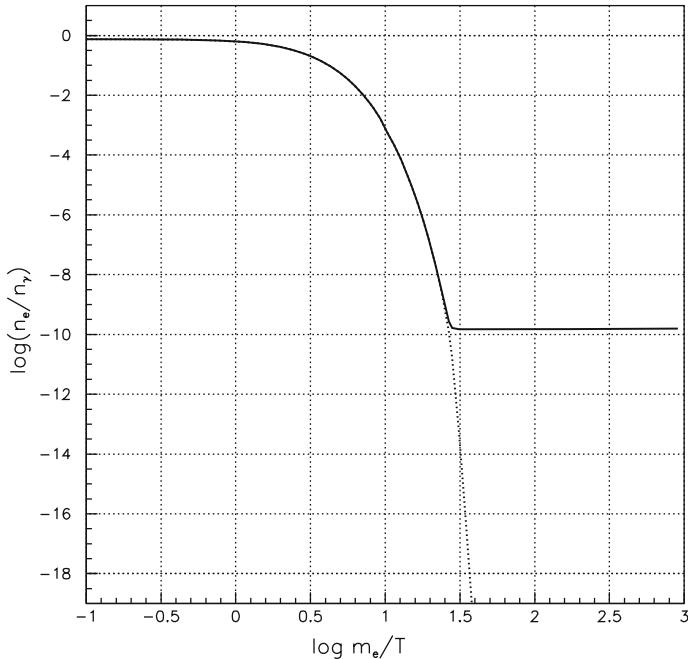


Fig. 6.4 $\log(n_{e^-}/n_\gamma)$ versus temperature for the case $(n_{e^-} - n_{e^+})/n_\gamma = 3 \times 10^{-10}$ found by numerically integrating the Boltzmann equation. The dotted line shows $\log(n_{e^+}/n_\gamma)$

$$s\gamma_{e^+e^-}a^3 = (2\pi^2/45)g\gamma_{e^+e^-}(T_\gamma)T_\gamma^3a^3 = \text{constant}, \quad (6.69)$$

where $g\gamma_{e^+e^-}(T_\gamma)$ is the effective number of spin states for the e^+e^- system at the temperature T_γ . Since equilibrium is maintained until there are very few remaining electrons, the entropy of the electron–positron–photon system for $T \gg m_e$ is nearly equal to the entropy of the photon system for $T \ll m_e$.

$$[n\gamma a^3]_{a_0} = [(1 + 2(7/8))n\gamma a^3]_{T \gg m_e}, \quad (6.70)$$

where we used the fact that $s\gamma \propto T^3 \propto n\gamma$. Dividing (6.68) by (6.70) we find

$$\left[\frac{n_{e^-}}{n\gamma} \right]_{a_0} = (11/4) \left[\frac{n_{e^-} - n_{e^+}}{n\gamma} \right]_{T > m_e}. \quad (6.71)$$

As for the positrons, they continue to annihilate on the fixed number of electrons leaving a positron relic density that is exponentially negligible (Exercise 6.7).

6.4 Neutrinos

The three neutrino species can be produced and destroyed at $T \sim \text{MeV}$ by the reaction

$$\nu \bar{\nu} \leftrightarrow e^+ e^- . \quad (6.72)$$

Since this reaction is due to the weak interactions (see Appendix C), the cross-section for all species is of order

$$\sigma \sim G_F^2 E_\nu^2 \quad m_e \ll E_\nu \ll m_W, \quad (6.73)$$

$$\sigma \sim \alpha^2 E_\nu^{-2} \quad E_\nu \gg m_W = 80.4 \text{ GeV}. \quad (6.74)$$

The annihilation rate is therefore

$$\Gamma_{\nu\bar{\nu} \rightarrow e^+e^-} = n_\nu \langle \sigma v \rangle \sim G_F^2 T^5 m_e \ll T \ll m_W. \quad (6.75)$$

Numerically, we find that the annihilation rate is greater than the expansion rate $H \sim T^2/m_{\text{pl}}$ for $T > 1 \text{ MeV}$ so we can conclude that for temperatures greater than 1 MeV, neutrinos were in chemical equilibrium. After the annihilation reactions freeze-out at $T_f \sim 1 \text{ MeV}$, the neutrinos decouple and their Boltzmann equation becomes

$$\frac{dn_\nu a^3}{da^3} \sim 0 \quad T < 1 \text{ MeV}. \quad (6.76)$$

We are now in a position to understand why the temperature for massless neutrinos is currently lower than the photon temperature. For $T > m_e$, photons, neutrinos, and electron–positrons were in thermal equilibrium with a unique temperature, $T_e = T_\gamma = T_\nu$. Apart from factors due to the Pauli principle and the small electron–positron asymmetry,² the numbers of electrons, positrons, neutrinos, and photons were equal. When T_e and T_γ dropped below m_e the e^+e^- pairs were transformed into photons (by $e^+e^- \rightarrow \gamma\gamma$) but not into the decoupled neutrinos. After these annihilations, the number of photons was therefore greater than the number of neutrinos. Since the distributions are thermal with $n \propto T^3$, it follows that $T_\gamma > T_\nu$ for $T < m_e$.

The ratio between the post-annihilation photon and neutrino temperatures can be calculated by using the fact that the electron–positron–photon system remains in thermal equilibrium until there are very few remaining electrons and positrons which implies that the post- and pre-annihilation entropies of this system are equal. As we found in the previous section (6.70), this implies

$$\frac{11}{4} \left[T_\gamma^3 a^3 \right]_{\text{before}} = \left[T_\gamma^3 a^3 \right]_{\text{after}}, \quad (6.77)$$

² We assume that any neutrino-antineutrino asymmetry is small.

where “before” and “after” refer to times well before and well after electron–positron annihilation. But after $T_f \sim 1$ MeV the temperature of the neutrinos simply falls as $1/a$ implying

$$[T_\nu a]_{\text{after}} = [T_\nu a]_{\text{before}} = [T_\gamma a]_{\text{before}}, \quad (6.78)$$

where in the last step we have used the fact that $T_\gamma = T_\nu$ before electron–positron annihilation. Substituting this into (6.77) we find that after electron–positron annihilation, the photon and neutrino temperatures are related by

$$T_\nu = (4/11)^{1/3} T_\gamma. \quad (6.79)$$

This corresponds to a neutrino number density per species of

$$n_\nu = (3/11)n_\gamma. \quad (6.80)$$

The temperature ratio (6.79) is maintained as long as the neutrinos remain relativistic since in this case both photon and neutrino temperatures fall as a^{-1} . If the neutrinos have masses that are sufficiently small, $m_\nu \ll T_\gamma(t_0)$, they are still relativistic and have a present temperature of

$$T_\nu(a_0) = (4/11)^{1/3} T_\gamma(a_0) \sim 2 \text{ K} \quad (\text{if } m_\nu \ll 10^{-4} \text{ eV}). \quad (6.81)$$

On the other hand, the density ratio (6.80) is maintained whether or not the neutrinos remain relativistic, so today we expect

$$n_\nu(a_0) = 1.12 \times 10^8 \text{ m}^{-3}. \quad (6.82)$$

A neutrino species with $m \gg 10^{-4}$ eV would now be non-relativistic and their spectrum would not be thermal because, in the absence of collisions, the form of a thermal spectrum is not conserved during the epoch when $T \sim m$. If their mass is $m > 10$ eV they would be sufficiently slow today to be trapped in the galactic potential wells and could constitute the galactic dark matter.

6.5 Primordial Nucleosynthesis

At sufficiently high temperatures and densities, nuclear reactions can take place and it is important for cosmologists to understand the mix of elements that was produced in the early universe. The calculations were originally performed by Gamow and collaborators in the 1940s with the hope that the relative abundances of all elements could be explained. We now know that nuclear reactions froze at $T \sim 30$ keV leaving most nuclei in the form of hydrogen and helium. Nucleosynthesis started up again once stars were formed providing “gravitational confinement” for astronomical “fusion reactors.”

In this section, we will present a very brief introduction to the theory and confirming observations [130]. The essential theoretical result will be predictions for the relative abundances of the light elements ($A \leq 7$, Table 6.3). Observationally, it is very difficult to determine the primordial abundances because of “pollution” by stellar nucleosynthesis. The best observational estimates are given in Table 6.4. Apart from ^1H and ^4He , small quantities of ^2H , ^3He , and ^7Li were produced in the early universe.

We will see that the abundances are predicted as a function of the baryon to photon ratio η (Fig. 6.7). The calculated abundances relative to ^1H for ^2H , ^3He , and ^4He agree with the best observations assuming the value for η taken from the CMB anisotropy measurements [1]:

$$\eta = \frac{n_b}{n_\gamma} \sim (6.18 \pm 0.2) \times 10^{-10} \quad (6.83)$$

i.e.

$$\Omega_b h_{70}^2 = 0.0456 \pm 0.0015. \quad (6.84)$$

Before the CMB measurement of η , primordial nucleosynthesis was the best way to measure the baryon content of the universe. The implied low value of Ω_b was

Table 6.3 The important nuclei [131] for nucleosynthesis with their binding energies per nucleon, B/A , their observed primordial abundances, their half-lives, and decay modes. (The half-life of ^7Be by electron capture is given for atomic beryllium.) We note the high binding energy of ^4He in comparison with the other light nuclei, which implies that this species will be the primary product of primordial nucleosynthesis. The absence of stable nuclei at $A = 5$ or $A = 8$ prevents the production of heavy elements by two-body reactions between ^1H and ^4He . Primordial nucleosynthesis therefore stops at $A = 7$. The production of heavy elements occurs in stars where the triple- α reaction $3^4\text{He} \rightarrow ^{12}\text{C}$ takes place. We note that B/A is a slowly varying function for $A > 12$ with a broad maximum at ^{56}Fe , the ultimate product of stellar nucleosynthesis. Elements with $A > 56$ can be produced by “explosive” nucleosynthesis in supernovae

Nucleus	B/A (MeV)	n_x/n_{H} Primordial (observed)	Half-life	Decay mode
p	0	1	$> 10^{32}$ yr	
n	0	0	10.24 min	$n \rightarrow p e^- \bar{\nu}_e$
^2H	1.11	$(2.68 \pm 0.27) \times 10^{-5}$		
^3H	2.83	0	12.3 yr	$^3\text{H} \rightarrow ^3\text{He} e^- \bar{\nu}_e$
^3He	2.57	$(1.1 \pm 0.2) \times 10^{-5}$		
^4He	7.07	0.064 ± 0.002		
^5Li	5.27	0	3×10^{-22} s	$^5\text{Li} \rightarrow p ^4\text{He}$
^6Li	5.33	$< 10^{-10}$		
^7Li	5.61	$(4.3 \pm 0.1) \times 10^{-10}$		
^7Be	5.37	0	53.3 d	$e^- ^7\text{Be} \rightarrow v_e ^7\text{Li}$
^8Be	7.06	0	6.7×10^{-17} s	$^8\text{Be} \rightarrow ^4\text{He} ^4\text{He}$
^{12}C	7.6	0		
^{16}O	8.0	0		
^{56}Fe	8.7	0		
^{208}Pb	7.7	0		

Table 6.4 Theoretical and observed primordial abundances relative to ^1H [130]. The theoretical abundances are for $\eta = 6.18 \times 10^{-10}$

Nuclei	Theory	Observation
$^2\text{H} : 10^5 n_2/n_1$	$2.83(1 \pm 0.03)$	2.68 ± 0.27
$^3\text{He} : 10^5 n_3/n_1$	$1.08(1 \pm 0.03)$	1.1 ± 0.2
$^4\text{He} : \rho_4/\rho_1$	0.2486 ± 0.0005	0.240 ± 0.006
$^7\text{Li} : 10^{10} n_7/n_1$	$4.6(1 \pm 0.1)$	1.25 ± 0.3

crucial in demonstrating the existence of non-baryonic dark matter. With the WMAP measurements, nucleosynthesis is no longer the most precise way to measure η but remains an essential demonstration that the physics of the early universe is well understood. An important caveat is that the estimated primordial abundance of ^7Li differs significantly from the calculated abundance. Cosmologists generally suppose that this discrepancy is due to poorly understood astrophysical effects that have corrupted the measurements.

Calculation of the primordial abundances are performed by numerically integrating the appropriate Boltzmann equations for each nuclear species. While this obviously requires a complicated computer code, we can understand things quantitatively because at high temperature most of the nucleons are free and remain so to a surprisingly low temperature, $T \sim 70\text{ keV}$. As illustrated in Fig. 6.5, the results can be understood by considering three epochs:

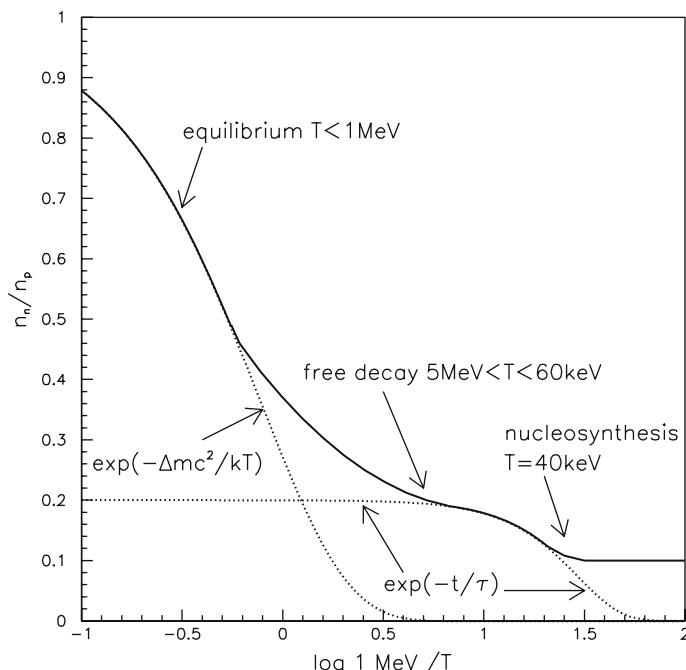
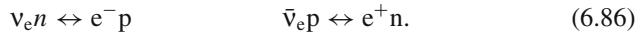


Fig. 6.5 The neutron–proton ratio as a function of time for temperatures $10\text{ MeV} > T > 10\text{ keV}$

1. $T > 800 \text{ keV}$. Neutrons and protons are free and in chemical equilibrium implying

$$\frac{n_n}{n_p} \sim \exp(-\Delta m/T) \quad \Delta m = m_n - m_p = 1.29 \text{ MeV}. \quad (6.85)$$

The chemical equilibrium is possible because of reactions transforming neutrons into protons and vice versa:



The cross-sections for these weak reactions are of the same order of magnitude as that for $v\bar{v} \leftrightarrow e^+e^-$ considered in the previous section. The reaction rate per baryon is therefore of order:

$$\Gamma_{n \leftrightarrow p} \sim G_F^2 T^5 \quad T \gg m_e, \Delta m. \quad (6.87)$$

The equilibrium is lost when the reaction rates fall below the expansion rate $\dot{a}/a \sim \sqrt{g(T)T^4/m_{pl}^2}$. The freeze-out temperature is near the freeze-out temperature of the reaction $v\bar{v} \leftrightarrow e^+e^-$:

$$T_f^3 \sim \sqrt{g(1 \text{ MeV})} G_F^{-2} m_{pl}^{-1} \Rightarrow T_f \sim 800 \text{ keV}. \quad (6.88)$$

From (6.85) it follows that the neutron–proton ratio at the end of this epoch is of order

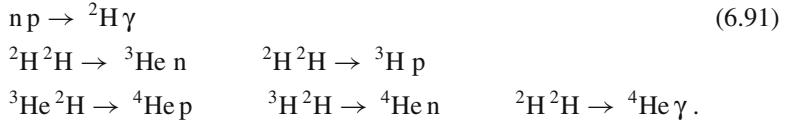
$$\frac{n_n}{n_p}(T_f) \sim 0.2 \quad T_f \sim 800 \text{ keV}. \quad (6.89)$$

We note that the curious “coincidence” $G_F^2 m_{pl} \Delta m^3 \sim 1$ is responsible for the fact that the number of neutrons is neither equal to nor much less than the number of protons. The consequence will be that the quantity of produced ${}^4\text{He}$ will be neither much greater than nor much less than the quantity of hydrogen.

2. $800 \text{ keV} > T > 60 \text{ keV}$. The neutrons decay freely. The duration of this period is $\Delta t = t(60 \text{ keV}) - t(800 \text{ keV}) \sim 3 \text{ min}$ (Exercise 6.11) so about half the neutrons decay leaving a neutron–proton ratio of

$$\frac{n_n}{n_p}(60 \text{ keV}) \sim 0.2 \exp(-\Delta t/\tau_n) \sim 0.1. \quad (6.90)$$

3. $T \sim 60 \text{ keV}$: nucleosynthesis. The remaining neutrons are rapidly incorporated into nuclei via a series of reactions, the most important being:



The nuclear abundances versus time are shown in Fig. 6.6. For $\eta \sim 5 \times 10^{-10}$, practically all the neutrons are incorporated into 4He , the most bound light nucleus.³ The number of available neutrons (6.90) therefore determines the quantity of helium:

$$\frac{\rho_{He}}{\rho_H} = \frac{2n_n/n_p}{1 - n_n/n_p} \sim 0.25 \quad (\eta \sim 5 \times 10^{-10}). \quad (6.92)$$

There are two obvious questions that we can ask about this scenario: why does nucleosynthesis start so late ($T \sim 70$ keV) and why does it stop so soon without the production of heavy elements. The first question is especially interesting because

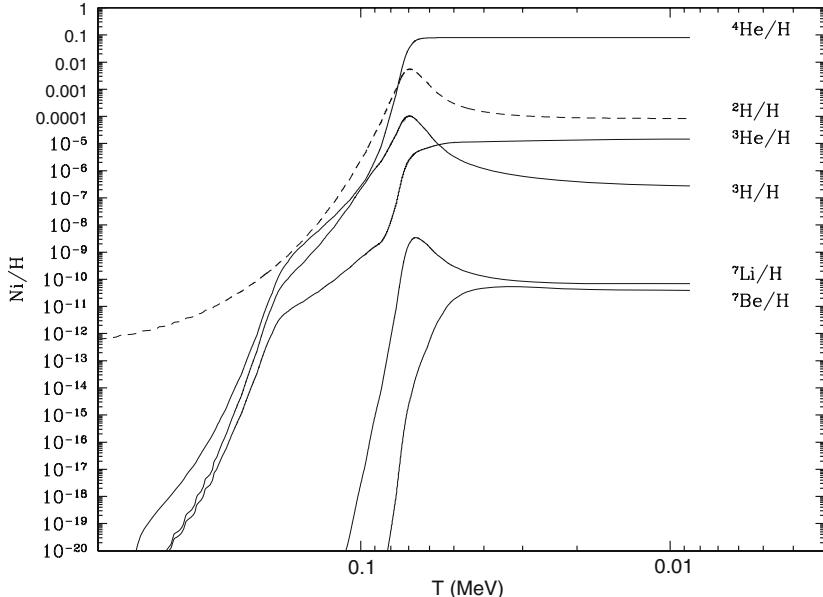


Fig. 6.6 The calculated abundances of the light elements as a function of temperature for $\eta = 3 \times 10^{-10}$ as calculated in [132]. The abundances are negligible until $T \sim 70$ keV, after which most of the available neutrons are incorporated into 4He . After $T \sim 30$ keV, nuclear reactions are frozen and the abundances are constant in time except for the later decays of 3H and 7Be . Courtesy of Elisabeth Vangioni-Flam

³ The most bound nucleus is not abundant because it is easy to produce but rather because it is very difficult to destroy at $T \sim 60$ keV.

the nuclear binding energies are all in the MeV range so it might be expected that nuclei would be produced when $T \sim \text{MeV}$. The reason for the late start is the tiny baryon–photon ratio $\eta \sim 5 \times 10^{-10}$ and its effect on the first step of nucleosynthesis, the formation of deuterium via reaction (6.91). For $\eta \sim 5 \times 10^{-10}$ and for $T \sim \text{MeV}$ the rate per neutron of this reaction is greater than the expansion rate (Exercise 6.10), from which it follows that there is approximate chemical equilibrium between n, p, and ^2H . Under these conditions, the formation rate of deuterium is equal to the destruction rate by photo-dissociation:

$$n_p n_n \langle \sigma_{np} v \rangle_T = n_2 n_\gamma \langle \sigma_{2\gamma} v \rangle_T, \quad (6.93)$$

where n_2 is the number density of deuterium. Since $n\gamma \gg n_p, n_n$, (6.93) can be satisfied only if $n_2 \ll n_p, n_n$ to compensate for the large number of photons. This situation persists until the temperature is sufficiently low that $\langle \sigma_{2\gamma} v \rangle_T$ becomes small because very few photons have energies above the threshold for photo-dissociation (2.2 MeV). Using the Saha equation, it can be shown (Exercise 6.10) that $n_2 \ll n_p, n_n$ for $T > 70 \text{ keV}$. Since heavier nuclei cannot be formed until deuterium is formed, it follows that nucleosynthesis cannot start until $T \sim 70 \text{ keV}$.

The end of nucleosynthesis so soon after its start is due to two effects. The first is the absence of stable or metastable elements with $A = 5$ or $A = 8$ which makes it impossible to form anything from the two primary species, ^4He and ^1H . The second is the increasing efficiency of the coulomb barrier between charged nuclei which strongly suppresses the cross-sections for $T < 60 \text{ keV}$. Nuclear reactions therefore freeze-out at a temperature of $T \sim 30 \text{ keV}$ with a non-equilibrium relic abundance of nuclei characterized by the complete lack of heavy nuclei.

The predicted abundances as a function of η are shown in Fig. 6.7. The abundance of ^4He , the primary product of primordial nucleosynthesis, is an increasing function of η . The abundances of the loosely bound intermediate nuclei ^2H and ^3He are decreasing functions of η . The reason for this behavior is quite simple. Nucleosynthesis can proceed only if the reaction rates between nuclei are greater than the expansion rate. The nuclear reaction rates are proportional to densities of initial state nuclei, which are themselves proportional to the total baryon density. For the first reaction (6.91), it is easy to show (Exercise 6.10) that its rate per neutron is smaller than the expansion rate for $\eta < 10^{-13}$, implying that there is essentially no nucleosynthesis for $\eta < 10^{-13}$. For $\eta > 10^{-13}$, nucleosynthesis proceeds with an efficiency that increases with η . More precisely, the temperature of the nuclear freeze-out is a decreasing function of η . The later the freeze-out, the more efficient the destruction of the intermediate nuclei and the production of ^4He .

We see in Fig. 6.7 that for $\eta \sim 6 \times 10^{-10}$ the abundance of ^4He is rather insensitive to η . This is simply because the great majority of the available neutrons are incorporated into this nucleus. On the other hand, the abundance of ^2H is very dependent on η , and before the WMAP measurements its observed abundance was the most reliable way to estimate η . Fortunately, the observed abundance agrees well with that predicted by the WMAP η value.

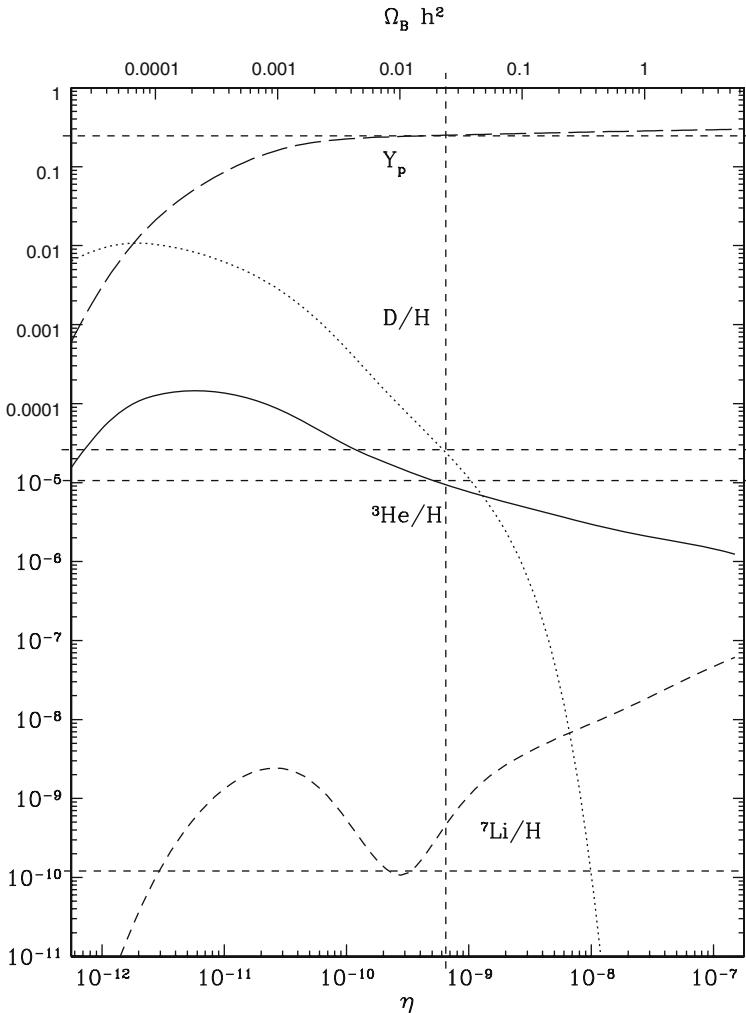


Fig. 6.7 The abundances of the light elements as a function of the assumed baryon-photon ratio η (Bottom Horizontal Axis) or of the assumed value of $\Omega_B h^2$ (top horizontal axis), as calculated in [132] (courtesy of Elisabeth Vangioni-Flam). For ${}^4\text{He}$, the abundance is given as the fraction $Y_p = \rho_{\text{He}}/\rho_b$ of the total baryonic mass that is in the form of ${}^4\text{He}$, while the other elements are reported as number densities normalized to ${}^1\text{H}$. The abundance by mass of ${}^4\text{He}$ is a slowly increasing function of η . The abundances of the loosely bound intermediate nuclei ${}^2\text{H}$ and ${}^3\text{He}$ are decreasing functions of η . The form of the curve for ${}^7\text{Li}$ is due to the fact that the production is mostly direct for $\eta < 3 \times 10^{-10}$ and mostly indirect via production and subsequent decay of ${}^7\text{Be}$ for $\eta > 3 \times 10^{-10}$. The dashed vertical line show the WMAP measurement of η [1] and the dashed horizontal lines show the recommended observed primordial abundances [130]

It is in principle simple to measure the quantity of deuterium at high redshift by measuring the relative absorption by deuterium and hydrogen of photons from background quasars. Figure 6.8 shows the spectrum of a high-redshift quasar. Blueward of the quasar's Lyman- α emission, we see the forest of lines corresponding to Ly- α absorption by hydrogen along the line-of-sight. The absorption due to hydrogen at a redshift z_H occurs at the wavelength satisfying $\lambda_1/\lambda_{\text{Ly}\alpha} = (z_H + 1)/(z_{\text{quasar}} + 1)$. The trick is to find an accumulation of matter (so-called "cloud") that has an optical depth that permits the observation of absorption by both hydrogen and deuterium. Such a cloud appears in this spectrum with absorption at 555.8 nm for hydrogen. The absorption is total at the center of the hydrogen line but the quantity of hydrogen in the cloud can be estimated from the width of the absorption profile. Also visible is the deuterium absorption line and the quantity of deuterium can be estimated from the total absorption at this line. The spectrum shows that only clouds within a

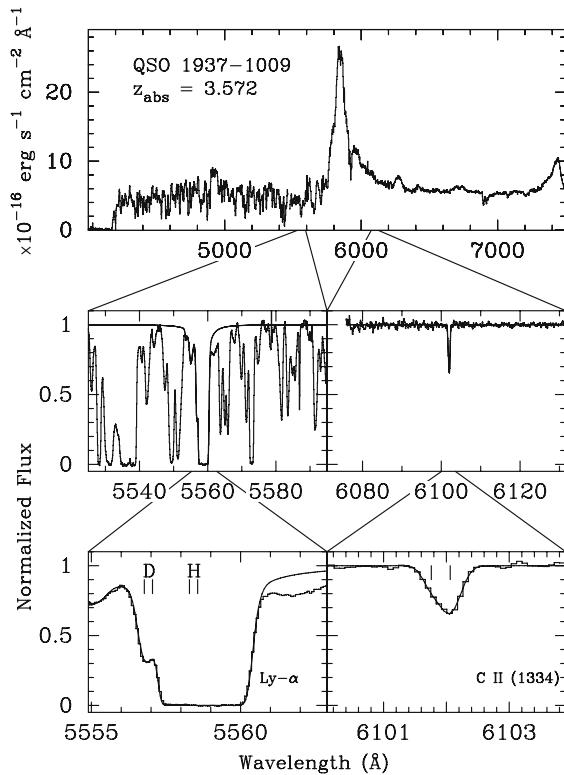


Fig. 6.8 The spectrum of a quasar at $z \sim 3.79$ showing Ly- α emission at 580 nm and, blueward of this line, the "forest" of Ly- α absorption lines by intervening atomic hydrogen [133]. The zoom on the left shows Ly- α hydrogen and deuterium absorption by a cloud at $z = 3.572$. The deuterium line is shifted with respect to the hydrogen line because the atomic energy levels are proportional to the reduced electron–nucleus mass. The ratio between the hydrogen and deuterium absorption can be used to determine the two abundances within the cloud. Courtesy of D. Tytler

narrow range of optical depth can be used to measure the deuterium–hydrogen ratio. The measurement would be impossible for clouds with less absorption (making the deuterium line too weak to be observed) or for clouds with more absorption (causing the hydrogen line to widen into the deuterium line).

Only a handful of appropriate absorption systems have been found in quasar spectra. The six best examples give results that are consistent with [133]:

$$n_2/n_1 = (2.68 \pm 0.08) \times 10^{-5}. \quad (6.94)$$

The quantity of heavy elements in the three high-redshift clouds is very small ($\sim 10^{-3}$ solar abundance) which suggests that the deuterium abundance might be unperturbed by stellar nucleosynthesis. If we suppose that (6.94) reflects the primordial abundance, the theory of nucleosynthesis gives a precise value of η and Ω_b :

$$\eta \sim (6.0 \pm 0.4) \times 10^{-10}. \quad (6.95)$$

As is often the case in astrophysics, the cited error is purely formal because the real uncertainty comes from the hypotheses necessary to interpret the data. In this case, it is necessary to suppose that the two lines in Fig. 6.8 near 555.8 nm are correctly identified and to suppose that the measured abundances are primordial. If either hypothesis is false, the measurement must be reinterpreted. For instance, the “deuterium” line could be a hydrogen line of a second cloud of a slightly different redshift. This would cause the observers to overestimate the deuterium and therefore underestimate Ω_b . On the other hand, if the measured deuterium is not primordial, the primordial deuterium is underestimated since stellar processes generally destroy deuterium. This would cause an overestimation of Ω_b . Debates over which clouds were the right ones to use were closed with by WMAP measurement.

The agreement between the observed primordial ^2H abundances and the prediction of nucleosynthesis theory using the WMAP value of η is a great triumph for cosmology. It means that the expansion rate at $T \sim 1\text{ MeV}$ is the rate predicted by the Friedmann equation. This is because the number of surviving neutrons available for nucleosynthesis depends directly on a comparison of the neutron decay rate (laboratory measurement) and the theoretical expansion rate.

The abundances of ^4He and ^3He are also in reasonable agreement with the predicted values. These abundances are derived from emission lines in hot ionized interstellar clouds. For ^4He normal atomic recombination lines are used while for ^3He the hyperfine ground state transition (analogous to the 21 cm hydrogen line) is used. These nuclei are produced in stellar nucleosynthesis and are then dispersed into the interstellar medium by supernova explosions. One must therefore use only clouds that have suffered little pollution from stellar nucleosynthesis as witnessed by very small abundances for heavier elements.

Finally, we must remember that the observed “primordial” ^7Li abundances is about a factor four lower than that predicted by theory. Because of its very low abundance, the observation of ^7Li is possible only in the atmospheres of very old

stars. It is suspected that the low observed abundance is due to settling of the lithium into the stellar interior. Few would say that the disagreement with predictions is sufficiently serious to call into question the Λ CDM model.

6.6 WIMPs

The three known neutrino species were relativistic when they decoupled ($T_f \gg m_\nu$). The consequence of this is that today the number density of neutrinos is of the same order of magnitude as that of photons $n_\gamma = (3/11)n_\nu$. If one of these neutrinos is sufficiently massive to be non-relativistic today, its present mass density would be $\rho_\nu = m_\nu n_\nu$. This gives $\Omega_M \sim 0.3$ for a neutrino mass of order 10 eV.

Any hypothetical stable particle with $m > 10$ eV that decoupled when it was relativistic would create problems for cosmology because the calculated mass density would be overcritical. A heavy weakly interacting particle can give an appropriate cosmological density only if it has an annihilation cross-section sufficiently large to keep it in equilibrium until the particle was non-relativistic, $T_f \ll m$. In this case its number density would be suppressed by the Boltzmann factor and the present density might not be too large. In fact, if the cross-section is chosen correctly, the particle can give a relic density near critical and constitute the desired non-baryonic dark matter.

Such a compensation between relic density and mass would seem a priori improbable, but stranger things have happened in cosmology. In fact, as it turns out, particles with weak interaction and masses in the GeV range naturally give relic densities within an order of magnitude or so of critical. Such particles are called “WIMPs” for “weakly interacting massive particle.” Numerically, we will find that

$$\Omega_{\text{wimp}} h_{70}^2 \sim \frac{1.4 \times 10^{-37} \text{cm}^2}{\langle \sigma v/c \rangle}. \quad (6.96)$$

The relic density is inversely proportional to the annihilation cross-section. This is reasonable since the higher the cross-section the more the particles would be destroyed in the early universe.

A stable WIMP is generally predicted by supersymmetric extensions of the standard model of particle physics. The particle is called the “LSP” (lightest supersymmetric particle) and is denoted by χ . Supersymmetric WIMPs are usually “Majorana” particles, i.e. they are their own antiparticle. Supersymmetric theories have many parameters that are not (yet) fixed by experiment so one can generally choose parameters that give a cross-section yielding the required relic density. The fact that they have not been seen at accelerators means that probably $m_\chi > 30$ GeV [62]. Searches for WIMPs trapped in the galactic halo were described in Sect. 2.5.

In order to understand (6.96), we have to consider the Boltzmann equation for n_χ :

$$\frac{dn_\chi a^3}{da^3} = 1/3 \frac{\Gamma_{\chi\chi \rightarrow x}}{\dot{a}/a} \frac{n_\chi(T)^2 - n_\chi^2}{n_\chi}, \quad (6.97)$$

where $\Gamma = n_\chi \langle \sigma v \rangle$ is the annihilation rate.

As with electrons and positrons, an approximate solution of (6.97) is

$$n_\chi \sim n_\chi(T) \quad T \gg T_f \quad (6.98)$$

$$n_\chi a^3 \sim \text{constant} \quad T \ll T_f, \quad (6.99)$$

where T_f is the freeze-out temperature corresponding to the moment when the annihilation rate was equal to the expansion rate:

$$n_\chi(T_f) \langle \sigma v \rangle = H(T_f) = \sqrt{8\pi g(T_f)/3} \frac{T_f^2}{m_{\text{pl}}}. \quad (6.100)$$

The numerical solution is shown for three values of the cross-section in Fig. 6.9. We see that, because of the exponential dependence of $n_\chi(T)$ for $T < m_\chi$, the

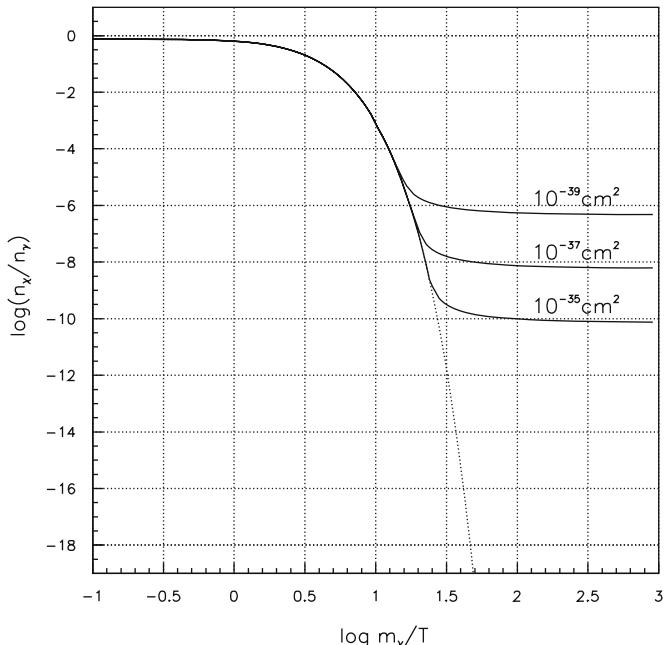


Fig. 6.9 $\log(n_\chi/n_\gamma)$ versus temperature for three values of the annihilation cross-section $\sigma v/c$. The WIMP mass was taken to be $m_\chi = 50$ GeV. The dotted line shows $\log(n_\chi/n_\gamma)$ in thermal equilibrium. Freeze-out occurs around $T_f \sim m_\chi/20$ where $g(T_f) \sim 50$. We see that to good approximation the relic density is inversely proportional to the cross-section. Using (6.104), it is easy to show that Ω_{wimp} is near unity for $\sigma v \sim 10^{-37} \text{ cm}^2$

freeze-out temperature is relatively insensitive to the cross-section, $T_f \sim m_\chi/20$. We can therefore easily estimate the χ relic density by equating the annihilation rate and the expansion rate. This gives:

$$\frac{n_\chi(a_f)}{n_\gamma(a_f)} \sim \frac{n_\chi(a_f)}{T_f^3} \sim \frac{\sqrt{g(T_f)}}{\langle \sigma v \rangle m_{\text{pl}} m_\chi / 20}. \quad (6.101)$$

After freeze-out, the expansion is adiabatic separately for the WIMPs and for the particles in thermal equilibrium:

$$n_\chi(a_0) = n_\chi(a_f) \left(\frac{a_f}{a_0} \right)^3 \quad (6.102)$$

$$n_\gamma(a_0) = n_\gamma(a_f) \left(\frac{g_f a_f^3}{g_0 a_0^3} \right), \quad (6.103)$$

which gives

$$\left[\frac{n_\chi}{n_\gamma} \right]_0 = \left[\frac{n_\chi}{n_\gamma} \right]_f \frac{g_0}{g_f}. \quad (6.104)$$

Combining with (6.101) we get the present day WIMP number density

$$n_\chi(a_0) \sim \left[\frac{n_\chi}{n_\gamma} \right]_f \frac{g_0}{g_f} n_\gamma(a_0) \sim \frac{n_\gamma(a_0)}{\sqrt{g(T_f)} \langle \sigma v \rangle m_{\text{pl}} m_\chi / 20}. \quad (6.105)$$

Multiplying by m_χ , we see that the relic density $\rho_\chi = m_\chi n_\chi$ is inversely proportional to the cross-section with a weak dependence on m_χ due to the factor $g_f = g(T_f)$. Dividing by the critical density we find numerically (6.96).

6.7 Baryogenesis

We have calculated the ratios n_V/n_γ and n_χ/n_γ with the Boltzmann equation and we might hope to also calculate the ratio

$$\eta = \frac{n_b}{n_\gamma} = \frac{n_b - \bar{n}_b}{n_\gamma} \sim 5 \times 10^{-10}. \quad (6.106)$$

The problem is fundamentally different from the two others. For neutrinos and WIMPs, the number of particles was hypothesized to be equal to the number of antiparticles, i.e. we supposed a vanishing chemical potential. To calculate η , the problem is to calculate the asymmetry or the chemical potential.

Of course, the baryon–antibaryon asymmetry can be considered to be simply an initial condition. If baryon number is conserved, the asymmetry will be conserved. Nevertheless, arguments of symmetry are very attractive and physicists originally considered the possibility that at large scale there is no asymmetry. The local asymmetry would be either a statistical fluctuation or the result of some hypothetical mechanism that separated baryons from antibaryons. Unfortunately, no plausible scenario has been found.

Since the invention of grand unified theories of particle physics in the 1970s, physicists have turned their attention toward mechanisms that can generate a global asymmetry starting with symmetric initial conditions. Three necessary conditions to achieve this were identified by Sakharov (in the 1960s). To generate an asymmetry, it is clearly necessary to have interactions that do not conserve baryon number and this is generally the case in GUTs. However, as we saw in Sect. 6.2, if the interactions take place in thermal equilibrium they would destroy rather than create an asymmetry. The second condition is therefore that the baryon non-conserving reactions take place out of equilibrium. This condition can be achieved because of the expansion of the universe. Finally, we note that if individual events violate baryon conservation, we cannot globally generate an asymmetry unless we favor production of baryons over antibaryons. We will see in a specific example (Exercise 6.9) that this can be achieved if the baryon-violating interactions also violate “C” and “CP,” i.e. charge conjugation symmetry and the product of charge conjugation and space inversion symmetry. The operation of charge conjugation changes particles to antiparticles (and vice versa) and the parity operation inverses all spatial coordinates.

In GUTs, baryon-number violating interactions are mediated by the exchange of super-heavy bosons $M > 10^{15}$ GeV. This explains the relative stability of the proton $\tau_p > 10^{32}$ yr. In these theories, the generation of the asymmetry occurs through asymmetric decays of heavy particles at high temperature $T \sim 10^{15}$ GeV. A specific model is the subject of Exercise 6.9.

As it turns out, things are a little more complicated because it was discovered that gauge theories that explain the observed low-energy interactions (e.g., the standard model of particle interactions with or without supersymmetry) contain subtle effects that violate baryon number via non-perturbative tunneling between degenerate vacuum states [134]. Thus, we expect baryon violation at temperatures $T \sim m_w$. If these interactions occur in thermal equilibrium they may destroy a pre-existing asymmetry. If they do not occur in thermal equilibrium, they may create a new one.

What exactly happens depends on the nature of the asymmetry created by the GUT interactions and on the nature of the electroweak baryon-number violating interactions. Depending on the type of Grand Unified Theory, interactions at the GUT temperature may or may not produce a lepton number asymmetry, e.g. $n_{e^-} + n_{\bar{\nu}_e} \neq n_{e^+} + n_{\bar{\nu}_e}$, and the fate of the baryon asymmetry depends on the lepton asymmetry. For example, in the standard model of particle physics *without* supersymmetry, interactions at $T \sim m_w$ drive the sum of the baryon and lepton numbers ($B + L$) to zero, while their difference ($B - L$) is conserved. If the GUT interactions set $B - L = 0$, then B is then driven to zero by the electroweak interactions. If the GUT interactions set $B - L \neq 0$, then B is driven to a new non-zero

value by the electroweak interactions. If supersymmetry is added to the standard model, the asymmetries will be treated in ways that depend on the parameters and symmetries of the model.

Further progress in this field would certainly benefit from some empirical input from particle physics concerning the nature of the GUT physics and the electroweak physics.

6.8 Irreversibility

We have seen that in a homogeneous universe, the entropy density evolves according to (6.23):

$$\frac{ds a^3}{da^3} = - \sum_i \frac{\mu_i}{T} \frac{dn_i a^3}{da^3}, \quad (6.107)$$

where the sum is over all particle species. This remarkable equation tells us that the entropy can increase only by the creation or destruction of particles *out of thermal equilibrium*. Without the creation or destruction of particles, $n_i a^3$ is constant for each species and each term in (6.107) vanishes. In thermal equilibrium, the condition for chemical equilibrium (6.46) requires that the sum vanish for all reactions.

In an *inhomogeneous* universe, there are other ways to increase the entropy. One thinks immediately of heat or particle diffusion between zones of differing temperature or density. An even more efficient way to increase the entropy is through the formation of gravitationally bound structures. This process results in the radiation of photons (Exercise 2.9). Since the entropy is proportional to the number of particles, this process directly increases the entropy.

In the early universe, there were few examples of entropy producing processes. One example that we have already seen is the free decay of particles, e.g. neutron decay during the period $800 \text{ keV} > T > 60 \text{ keV}$. The second example that we consider here is particle–antiparticle annihilation near the freeze-out temperature $T \sim T_f$.

We therefore come back to the case of a Majorana WIMP ($\chi = \bar{\chi}$) considered in Sect. 6.6. Substituting the WIMP Boltzmann equation (6.97) into (6.107), we find

$$\frac{ds a^3}{da^3} = - \frac{\mu_\chi}{T} \frac{\langle \sigma v \rangle}{3\dot{a}/a} \left(n_\chi(T)^2 - n_\chi^2 \right). \quad (6.108)$$

For $T > T_f$, the χ are almost at thermal equilibrium, $n_\chi \sim n_\chi(T)$ so the Boltzmann equation (6.97) gives

$$n_\chi \sim n_\chi(T) + \frac{3\dot{a}/a}{2\Gamma} \left| \frac{dn_\chi a^3}{da^3} \right| \quad T > T_f \Rightarrow \Gamma > \dot{a}/a. \quad (6.109)$$

We see that n_χ is a little larger than the equilibrium value. The excess is due to the fact that the χ cannot annihilate fast enough to follow the rapidly declining equilibrium abundance.

The χ excess given by (6.109) is equivalent to a chemical potential:

$$\mu_\chi = \frac{3\dot{a}/a}{2\Gamma} \frac{1}{n_\chi} \left| \frac{dn_\chi a^3}{da^3} \right| \quad T > T_f \Rightarrow \Gamma > \dot{a}/a. \quad (6.110)$$

The derivative of the entropy is therefore

$$\frac{ds a^3}{da^3} \sim \frac{3\dot{a}/a}{2\Gamma} \frac{1}{n_\chi} \left| \frac{dn_\chi a^3}{da^3} \right|^2 \quad T > T_f \Rightarrow \Gamma > \dot{a}/a. \quad (6.111)$$

Approximating n_χ by its equilibrium value we find

$$\frac{ds a^3}{da^3} \sim \frac{\dot{a}/a}{6\langle \sigma v \rangle} \frac{m^2}{T^2} \quad m_\chi > T > T_f. \quad (6.112)$$

The right-hand-side of this equation is constant to the extent that $\dot{a}/a \propto T^2$. In this approximation, we can replace all the quantities by their value at $T = T_f$:

$$\frac{ds a^3}{da^3} \sim \frac{n_\chi(T_f)}{6} \frac{m^2}{T_f^2} \quad m_\chi > T > T_f, \quad (6.113)$$

where we have used $\dot{a}/a = \Gamma$ at $T = T_f$. Integrating this equation, we find the entropy generated during the period of annihilation:

$$\frac{\Delta s a^3}{sa^3} \sim \left(\frac{m_\chi}{T_f} \right)^2 \left[\frac{n_\chi}{s} \right]_{T_f} \quad T_f < m_\chi. \quad (6.114)$$

Referring to Fig. 6.9, we see that the annihilation of a WIMP of mass $m_\chi \sim 50$ GeV and annihilation cross-section $\sigma v/c \sim 10^{-37}$ cm² will have a freeze-out temperature of $m_\chi/T_f \sim 20$ and a relic density of $n_\chi/s = n_\chi/g n_\gamma \sim 10^{-10}$. The quantity of entropy generated at freeze-out is therefore $\Delta s a^3/sa^3 \sim 10^{-8}$.

The immediate manifestation of the entropy increase would be the decrease by the same factor of the ratio between the number density of baryons and the number density of relativistic particles, n_{rel} . This comes about because baryon number is conserved to high precision ($n_b a^3 = \text{constant}$) while the number of relativistic particles follows the entropy which is not conserved ($n_{\text{rel}} a^3 \sim sa^3 \neq \text{constant}$).

A more dramatic consequence of the entropy increase is that the expansion is not reversible implying that the contraction epoch of a closed universe would not be the time reversal of the expansion epoch. Figure 6.10 shows the evolution of the ratio n_χ/n_γ during the two epochs. After freeze-out in the expansion epoch, annihilation continues for a short time without compensation by the inverse reaction. During the

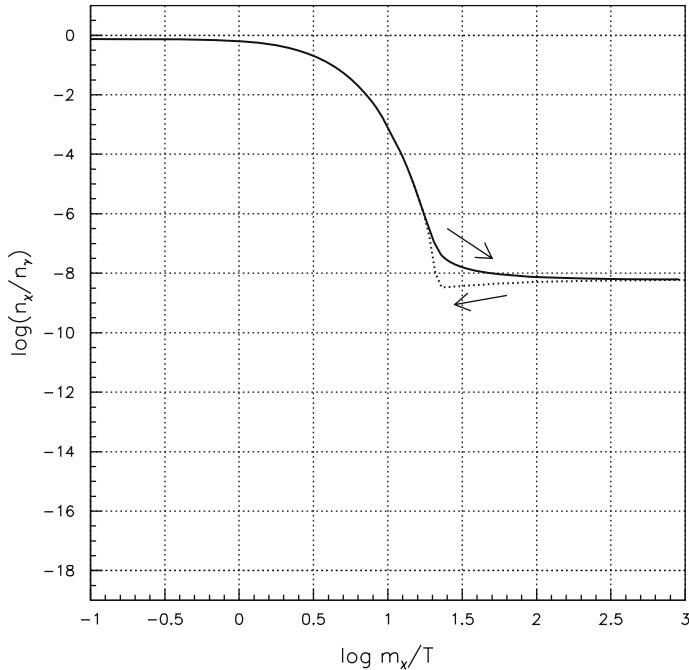


Fig. 6.10 $\log(n_\chi/n_\gamma)$ versus temperature for $m_\chi = 50 \text{ GeV}$ and $\sigma v/c = 10^{-37} \text{ cm}^2$. The *solid line* corresponds to the expansion phase and the *dotted line* corresponds to the contraction phase of a closed universe

contraction epoch, the annihilation starts up again before the temperature increases to the point where thermal production of χ can begin. A movie of the contraction phase is therefore not the movie of the expansion phase run in reverse since over a short period during the contraction, n_χ decreases while the temperature increases. For the expansion movie run in reverse, n_χ always increases while the temperature increases.

6.9 The Future

The present thermal activity of the universe is dominated by the production of photons, neutrinos, and heavy nuclei in nuclear fusion reactions in stars. Photons escape from stellar surfaces into space while neutrinos produced in the fusion of hydrogen to helium escape directly from stellar cores. The “hot” neutrinos and photons therefore add to the cold primordial neutrinos and photons to produce a non-equilibrium mixture.

The next few Hubble times will see a continuation of these processes. The cosmological formation of structure stopped (or will stop) when Ω_M dropped significantly below unity, fixing the quantity of material available for star formation. Heavy stars can recycle their material through supernova explosions but we can suppose that

after some time, most baryonic material will end up in low-mass stars. After burning the hydrogen in their cores, these stars simply cool down. In some number of Hubble times we can therefore expect that the universe will consist of extragalactic particles and galaxies consisting of non-baryonic dark matter and dead stars, i.e. white dwarfs, brown dwarfs, neutron stars, and black holes. It can be expected that galactic mass will become more and more concentrated in massive black holes in galactic centers as these objects accrete surrounding mass.

If the universe avoids a contraction epoch, as suggested by the present measurement of Ω_Λ , the far future will be determined by more exotic processes. The future is simple if the proton is unstable as predicted by many extensions of the standard model of particle physics. Proton decay will produce more photons, neutrinos, and positrons. The positrons from decays in stars will immediately annihilate. After a few proton lifetimes, dead stars will have evaporated because of proton decay. Positrons produced in intergalactic decays will be protected from annihilation by the expansion of universe with $\Gamma \ll H$.

The last surviving structures will be black holes, but these will also evaporate by Hawking radiation. After this happens, the universe will consist of the original CMB and neutrinos, of various non-thermal mixtures of stable particles, and of presumably a vacuum energy that dominates the total energy. What happens after this depends on the unknown nature of the vacuum energy. As an example, we note that the last time the universe was dominated by vacuum energy was during inflation. This period ended with a phase transition that started the epoch of particle domination.

Exercises

6.1 Consider a system of identical particles in thermal equilibrium at a temperature T and chemical potential μ . The phase-space distribution is given by (6.10).

- (a) Verify the thermal equilibrium formulas in Table 6.1.
- (b) The use of ideal gas formulas for a gas of charged particles is justified only if the mean particle kinetic energy is much greater than the mean potential energy due to coulomb interactions. For a globally neutral gas, the potential energy is of order $U \sim e^2 n^{1/3}$ where n is the number density of charged particles. Compare the two energies for a gas of electron–positron pairs at temperature $T \gg m_e$ and for $T \ll m_e$.
- (c) In thermal equilibrium, if the density of particles $n(T)$ is not equal to the density of antiparticles $\bar{n}(T)$, their chemical potentials do not vanish, $\mu = -\bar{\mu} \neq 0$. Calculate $n(T, \mu) - \bar{n}(T, -\mu)$ in the limit ($T \gg m$, $T \gg \mu$). Estimate the chemical potentials of electrons and positrons in our universe for $T \gg m_e$.
- (d) After the freeze-out, $T < T_f \ll m_e$, there are no electron–positron annihilations and the densities are fixed at $n_e \sim 4 \times 10^{-10} n_\gamma$ and $\bar{n}_e \sim 0$. In the absence of annihilations, we can no longer suppose $\mu = -\bar{\mu}$. Estimate the value of μ_e for $T \ll T_f$.

6.2 Consider an ideal gas in a volume V of total energy E and consisting of N particles. The temperature T , the pressure p , and the chemical potential μ are defined so that the differential of the entropy S is given by

$$dS = \frac{dE}{T} + \frac{pdV}{T} - \frac{\mu dN}{T}.$$

Use this relation to show that the entropy density $s(T, \mu)$ is given by

$$s(T, \mu) = \frac{\rho(T, \mu) + p(T, \mu) - \mu n(T, \mu)}{T}.$$

6.3 Demonstrate Liouville's equation for massive particles (6.26) by appropriately modifying Figs. 1.10 and 1.11.

6.4 Consider a reaction $i + j \rightarrow k + l$ with a rate

$$\Gamma_{ij \rightarrow kl} = n_j \langle \sigma v \rangle, \quad (6.115)$$

where $\langle \sigma v \rangle$ is the mean cross-section times velocity.

Consider a time t_1 after the freeze-out when the reaction rate is significantly less than the expansion rate

$$\Gamma(t_1) \ll H(t_1). \quad (6.116)$$

We suppose that after t_1 the number of j particles is conserved: $n_j \propto a^{-3}$ and that the cross-section has a simple form: $\langle \sigma v \rangle \propto T^\beta \propto a^{-\beta}$ ($\beta = 0$ for electron–positron annihilation and for Thomson scattering). Integrate the reaction rate to show that the reaction is really frozen:

$$\int_{t_1}^{\infty} \Gamma dt = \int_{a_1}^{\infty} \Gamma da/\dot{a} \ll 1, \quad (6.117)$$

i.e. that most i particles will never scatter again.

6.5 The binding energy of the ground state of atomic hydrogen is $B = 13.6$ eV and for $T < B$ we would expect that electrons would “recombine” with the protons to form atoms. Because the cross-section for photon–hydrogen scattering is less than that for photon–electron scattering, the recombination precipitates the decoupling of matter and photons making the universe effectively transparent.

In this exercise, we will estimate the temperature at which the universe becomes transparent and the number of electrons that remain unbound. To do this, we will

need the fraction of free electrons in thermal equilibrium and the cross-section for atom formation.

To simplify the problem, we will consider only the ground state of atomic hydrogen. In this case, in thermal equilibrium the densities of electrons, protons, and hydrogen atoms are related by the Saha equation:

$$\frac{n_{\text{H}}}{n_{\text{p}} n_{\text{e}}} = \left(\frac{2\pi}{m_{\text{e}} T} \frac{2\pi}{m_{\text{p}} T} \frac{m_{\text{H}} T}{2\pi} \right)^{3/2} e^{B/T} \sim \left(\frac{2\pi}{m_{\text{e}} T} \right)^{3/2} e^{B/T}.$$

- (a) Prove the Saha equation by using $\mu_{\text{e}} + \mu_{\text{p}} = \mu_{\text{H}}$ in thermal equilibrium. Neglecting the presence of helium, we can write:

$$\begin{aligned} n_{\text{e}} &= n_{\text{p}} = x n_{\text{b}} = x \eta n_{\gamma} \\ n_{\text{H}} &= (1 - x) \eta n_{\gamma}, \end{aligned}$$

where $\eta = n_{\text{b}}/n_{\gamma} \sim 5 \times 10^{-10}$ and x is the ionization fraction. The Saha equation becomes:

$$\frac{1 - x}{x^2} = \eta \frac{2.4}{\pi^2} \left(\frac{2\pi T}{m_{\text{e}}} \right)^{3/2} e^{B/T}.$$

We see that since $\eta \ll 1$, the temperature must be significantly smaller than the binding energy B before the density of atoms becomes significant. It follows that recombination happens much later than we would have naively expected. This is the same situation as we encountered in Sect. 6.5 where we found that the formation of nuclei started only at $T \sim 60 \text{ keV}$, i.e. much below the binding energy of deuterium. The explanation was that the density of photons was so large that deuterium (or atomic hydrogen) was photo-dissociated rapidly even for temperatures much less than the binding energy.

The Saha equation applies only in the case of thermal equilibrium which requires that the recombination rate be greater than the expansion rate. The cross-section for the formation of the ground state of atomic hydrogen is

$$\sigma_{\text{rec}}(\text{e}^- \text{p} \rightarrow \text{H}\gamma) = 4\pi \frac{\alpha^2}{m_{\text{e}}^2} \frac{B}{mv^2}.$$

Multiplying by the velocity and averaging over the thermal spectrum we get a reaction rate inversely proportional to \sqrt{T} :

$$\langle \sigma_{\text{rec}} v \rangle = 1.4 \left(\frac{1 \text{ eV}}{T} \right)^{1/2} 10^{-13} \text{ cm}^3 \text{ s}^{-1}.$$

The recombination rate is then this factor multiplied by the proton density, $\Gamma_{\text{rec}} = n_{\text{p}} \langle \sigma_{\text{rec}} v \rangle$

The transparency of the universe is determined by the scattering rate. Two processes should be considered: Thomson scattering on free electrons and Rayleigh scattering on atoms. The cross-section for Rayleigh scattering is of order $\sigma_R = \sigma_T(E\gamma/B)^4$ where $\sigma_T = 6.65 \times 10^{-25} \text{ cm}^2$ is the Thomson cross-section. The photon scattering rate is therefore

$$\Gamma_T + \Gamma_R \sim n_e \sigma_{TC} + n_H \sigma_T \left(\frac{T}{B} \right)^4 c. \quad (6.118)$$

- (b) For a series of temperatures between 0.3 eV and 0.2 eV calculate the expansion rate $H(T)$, the equilibrium hydrogen ionization fraction $x(T)$, and the equilibrium rates $\Gamma_R(T)$, $\Gamma_T(T)$, and $\Gamma_{\text{rec}}(T)$. We can suppose $\Omega_M = 0.3$ and $\eta = 5 \times 10^{-10}$.
- (c) Estimate the temperature when $\Gamma_T + \Gamma_R$ becomes smaller than $H(T)$. This is the temperature when the universe becomes transparent for photons.
- (d) Estimate the temperature when the recombination rate becomes smaller than the expansion rate. What is the fraction of hydrogen that remains ionized?

6.6 The photons from the reaction $e^+e^- \rightarrow \gamma\gamma$ are highly non-thermal for late annihilations $T \ll m_e$. In Thomson scattering with $E\gamma \ll m_e$, a photon loses on average a fraction $\sim E\gamma/m_e$ of its energy. Estimate the number of collisions necessary for an annihilation photon to reach a thermal energy. Compare the time necessary for thermalization with the Hubble time at the appropriate epoch.

6.7 Estimate the freeze-out temperature for the annihilation of positrons in a universe with $n_e/n_\gamma \sim 4 \times 10^{-10}$. Estimate the relic density of positrons.

6.8 Neutrinos (antineutrinos) produced in normal weak interactions have left handed (right handed) helicities. If neutrinos are “Dirac” particles, neutrinos and antineutrinos of the “wrong” helicity exist but have cross-sections smaller than that of normal neutrinos by a factor $\sim m_V^2/E_V^2$. Show that neutrinos of the wrong helicity and of mass less than $\sim 10 \text{ keV}$ never reach thermal equilibrium.

6.9 In this exercise, we will study a model [135] of baryogenesis based on certain ideas of Grand Unified Theories (GUTs). We suppose the existence of a super-heavy boson X of charge 4/3 and its antiparticle \bar{X} with $m_X \gg 1 \text{ TeV}$. We suppose that initially the universe is symmetric, $n_X = \bar{n}_X$, and that the X have decoupled from the other particles so that their only interactions are free decays of mean lifetime τ . A fundamental theorem of particle physics (CPT conservation) implies that the mean lifetimes of X and \bar{X} are equal, $\tau = \bar{\tau}$ (see below).

To simplify things, we suppose that there are only two decay modes of the X, both producing u or d quarks or antiquarks:

$$\begin{array}{ll} X \rightarrow uu & \text{branching ratio} = r \\ X \rightarrow \bar{d}e^+ & \text{branching ratio} = 1 - r . \end{array}$$

The decay modes for \bar{X} are

$$\begin{array}{ll} \bar{X} \rightarrow \bar{u}\bar{u} & \text{branching ratio} = \bar{r} \\ \bar{X} \rightarrow de^- & \text{branching ratio} = 1 - \bar{r} . \end{array}$$

If the symmetries “C” and “CP” are violated, we can have $r \neq \bar{r}$ (see below).

- (a) The baryon number of u and d quarks is $1/3$ while that of the antiquarks is $-1/3$. What is the mean baryon number generated per X decay and per \bar{X} decay?
- (b) We suppose that the decays of X and \bar{X} occur during a short period at a temperature T_1 and scale factor a_1 . Ignoring the expansion during the decay period, calculate the excess baryon number $\Delta n_b = n_b - \bar{n}_b$ generated by the decays as a function of the original X density $n_X = \bar{n}_X$.
- (c) What is the baryon–photon ratio $\eta(a_1) = \Delta n_b/n\gamma$ just after the decay period?

After the decay period, we suppose that there are no further baryon number violating interaction so that the excess is conserved:

$$\Delta n_b(a) = \Delta n_b(a_1) \left(\frac{a_1}{a} \right)^3 .$$

- (d) If $g_s(T_1)$ is the number of relativistic spin states at T_1 , calculate the present value of $\eta(a_0) = \Delta n_b/n\gamma$ by supposing that the expansion was adiabatic after t_1 .

A reaction is said to violate “C” invariance if the “antireaction” where all particles are replaced by their antiparticles does not, in reality, proceed at the same rate as the original reaction. A reaction is said to violate “P” invariance if the reaction viewed in a mirror does not, in reality, proceed at the same rate as the original reaction.⁴ Finally a reaction violates “CP” if the process viewed in a mirror with particles replaced by antiparticles does not, in reality, proceed at the same rate as the original reaction.

It is clear that if C is not violated, the rate of $X \rightarrow \bar{d}e^+$ must be the same as the rate of $\bar{X} \rightarrow de^-$, implying that $r = \bar{r}$.

- (e) The requirement that CP be violated to have $r \neq \bar{r}$ can be understood by considering the decay of a classical spinning particle. Draw a picture of $X \rightarrow \bar{d}e^+$

⁴ Strictly speaking, the P operation reverses all coordinates while mirror reflection reverses only the coordinate perpendicular to the mirror. However, mirror reflection followed by an appropriate spatial rotation can reverse all coordinates. Mirror reflection is therefore equivalent to P if rotation symmetry is respected.

placed in front of a mirror. Observe that the spins of all particles are reversed in the mirror image. Now imagine that all particles in the mirror are the antiparticles of the original particles. Argue that if the reaction in the mirror proceeds at the same rate as the original reaction for all possible spin combinations of the original reaction, i.e. if CP is not violated, then $r = \bar{r}$ after averaging over all possible spins.

- (f) A reaction violates “CPT” if the time reversal of a process seen in a mirror with particles replaced by antiparticles does not, in reality, proceed at the same rate as the original process. Draw a picture of the resonant elastic scattering $\bar{d}e^+ \rightarrow X \rightarrow \bar{d}e^+$ placed in front of a mirror. Argue that the CPT transformed reaction $de^- \rightarrow \bar{X} \rightarrow de^-$ would be shown in a film running backward of the reaction in the mirror with particles replaced by antiparticles. The two reactions will have a resonant peak at a center-of-mass energy equal to m_X for the first reaction and $m_{\bar{X}}$ for the second. If CPT invariance is respected, the cross-section for the first reaction must be equal to that for the second. Because of the resonant peaks in the energy-dependences of the two cross-sections, argue that the masses and total widths (lifetimes) of X and \bar{X} are equal.

- 6.10** The cross-section times velocity for the reaction $np \rightarrow {}^2H\gamma$ is

$$\sigma v \sim 7.4 \times 10^{-20} \text{ cm}^3 \text{ s}^{-1} \quad (v \ll c). \quad (6.119)$$

- (a) Show that the rate per neutron of this reaction is smaller than the expansion rate at $T \sim 60 \text{ keV}$ if $\eta < 4 \times 10^{-12}$. It follows that there is no nucleosynthesis if η is less than this value.
 (b) If $\eta > 4 \times 10^{-12}$, deuterium is in thermal equilibrium with neutrons and protons. By imposing $\mu_2 = \mu_p + \mu_n$ show that the abundances of deuterium, protons, and neutrons are governed by the Saha equation:

$$\frac{n_2}{n_p n_n} = \left(\frac{2\pi}{m_n T} \frac{2\pi}{m_p T} \frac{m_2 T}{2\pi} \right)^{3/2} e^{B/T}$$

where $B = 2.2 \text{ MeV}$ is the deuterium binding energy.

- (c) Show that for $\eta \sim 5 \times 10^{-10}$ the great majority of neutrons are free until $T \sim 60 \text{ keV}$. (Since the majority of baryons are protons, you can approximate $n_p \sim \eta n \gamma$.)

- 6.11** Calculate the duration of the epoch of neutron decay ($t(60 \text{ keV}) - t(800 \text{ keV})$). Estimate the entropy generated by this irreversible process. Discuss how “nucleo-destruction” in a contraction phase of a closed universe would differ from the time reversal of nucleosynthesis in the expansion phase.

6.12 Consider the freeze-out temperature (6.88) of the weak reactions transforming neutrons into protons. The value of T_f depends, via the expansion rate, on the number of relativistic spin states and therefore on the number N_V of neutrino species (or any other light particle species in thermal equilibrium). The number of neutrons available for helium production therefore depends on N_V .

To simplify things, we suppose that the observed light element abundances agree with the calculated abundances for $\eta \sim 5 \times 10^{-10}$ and $N_V = 3$. If the ${}^4\text{He}$ abundance n_4 is known with a precision of 10%, derive an upper limit on N_V .

Before the measurement of $N_V = 3$ at the LEP Collider using the Z^0 decay width, the cosmological limit on N_V was much stronger than the experimental limits. It should be emphasized that the LEP measurement is sensitive to neutral weakly interacting particles that couple to the Z^0 and that have masses $< M_Z$. On the other hand, the cosmological limit applies to any light particle species in thermal equilibrium at $T \sim \text{MeV}$.

Chapter 7

Structure Formation

The basic idea of structure formation is quite simple. During the not too distant past ($\hat{a} < (\Omega_M/\Omega_\Lambda)^{1/3} \sim 0.75$), the universe was matter-dominated with a near critical mean density, $\bar{\rho} \sim 3H^2/8\pi G$. Under these conditions, a “perturbed” region with a small positive density fluctuation will have an overcritical density and, following the Newtonian argument of Sect. 1.2.6, a negative total energy. As illustrated in Fig. 7.1, we would expect that the particles in the over-dense region will behave as though they were in a mini-closed universe. They will initially participate in the universal expansion but, because of the mass excess, they will decelerate with respect to the rate of universal expansion. The over-dense region will reach a maximal extent, after which the particles will fall back on themselves to form a gravitationally bound structure.

This is basically what happens in a universe dominated by CDM, as ours was during the epoch of matter domination. However, the presence of significant quantities of baryons, photons, neutrinos, and dark energy in our universe modifies this simple picture in the following ways:

- Before recombination, baryons and photons form a fluid so that density fluctuations of these two components propagate as acoustic waves with a sound speed $\sim c/\sqrt{3}$. The velocity of CDM particles near the edge of the over-dense region of radius $R(t)$ is HR and the sound speed is much greater than this if the size of the region is much less than the Hubble distance: $R < c/H$. For the configuration shown in Fig. 7.1 this means that the baryon–photon fluctuation will leave the zone of the initial fluctuation. During the radiation epoch, gravity was dominated by this component, so the CDM is perturbed, inhibiting the growth of the amplitude of the CDM fluctuations.
- Neutrinos, before the temperature drops below their mass, are collisionless relativistic particles. As such, they, like the photons and baryons, leave the perturbed region in Fig. 7.1.
- Dark-energy domination beginning at $\hat{a} = (\Omega_M/\Omega_\Lambda)^{1/3}$ causes the expansion to accelerate, overwhelming the gravitational effect of the CDM. This prevents the formation of structures from small density fluctuations.

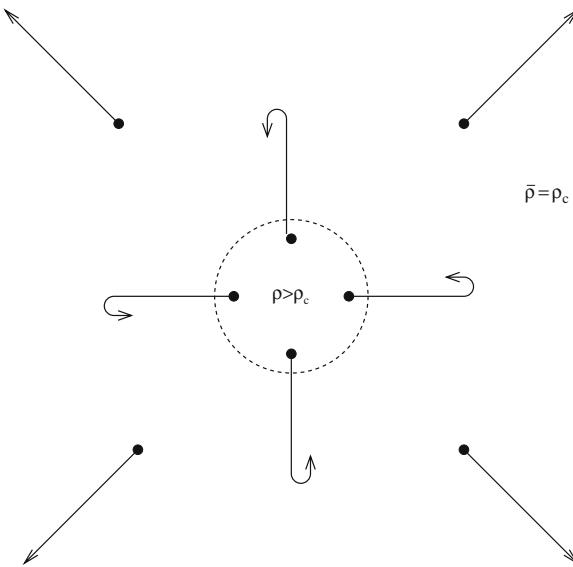


Fig. 7.1 Structure formation in a CDM universe with $\Omega_M = \Omega_T \sim 1$. In such a universe, an over-dense region ($\rho > \rho_c$) can evolve like a mini-closed universe. The region will expand to a maximum radius and then detach from the universal expansion by contracting to form a gravitationally bound object while the rest of the universe continues its expansion. In a universe with Ω_M significantly less than unity, a small over-density will not be sufficient to make the region supercritical and the region will expand eternally. In a universe with Ω_M significantly greater than unity, regions with a small over-density do not have time to form structures before the universe as a whole collapses

We will see that the effects of the baryons, photons, and neutrinos are seen today in the suppression of density fluctuations at small scale (Fig. 7.4). The effect of the dark energy is seen, e.g., in the number density of galaxy clusters (Fig. 2.15) increasing with decreasing redshift more slowly than what one would expect if structure formation were not inhibited by the dark energy.

The mathematical details of structure formation theory are very complex. One must solve the Boltzmann–Liouville equations for each particle species coupled to the Einstein equations for an inhomogeneous universe. The treatment of the Einstein equations requires a choice of coordinates that is not nearly as obvious as the comoving coordinates for a homogeneous universe. For nearly homogeneous mixtures of CDM and a baryon–photon fluid, solutions to the equations can be found by the usual technique of decomposing inhomogeneities in normal modes. As long as the inhomogeneities are small (the so-called “linear regime”) the modes develop in time independently via uncoupled ordinary differential equations. Once the inhomogeneities become large, numerical techniques must be used: N-body simulations for CDM and hydrodynamic simulations for fluids. The result of these calculations is to transform the nearly homogeneous early universe into the highly inhomogeneous universe of bound clusters and unbound filaments. A small part of a simulated uni-

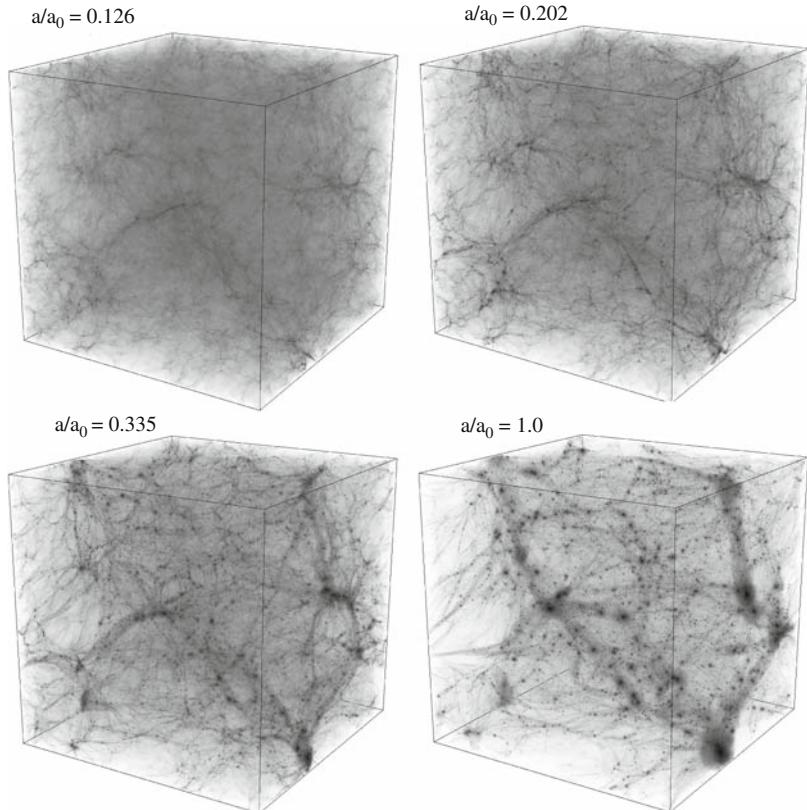


Fig. 7.2 A simulated $(20h^{-1} \text{ Mpc})^3$ volume at four epochs showing the formation of filamentary structures and galaxy clusters. The N-body simulation used 512^3 particles [137]. (Courtesy of Stéphane Colombi.)

verse is shown in Fig. 7.2 where one sees filaments and bound objects very much like those seen in redshift surveys (Fig. 2.18).

Because of the complicated nature of structure formation, the aim of this chapter is to give a qualitative introduction to the subject. (More quantitative treatments can be found in [138].) In this spirit, we begin in Sect. 7.1 by studying an isotropic universe that is homogeneous apart from a spherical over-dense region at the origin. The model will allow us to make some important conclusions about the growth of inhomogeneities in the linear regime, and about the nature of the objects formed after gravitational collapse.

In Section 7.2 we introduce the formalism to study more general density fields by expanding the density inhomogeneities and the associated Newtonian potential inhomogeneities in Fourier series. Besides neatly separating fluctuations according to length scales, this decomposition simplifies enormously the time dependence of the inhomogeneities: during the matter epoch, the potential fluctuations are constant

in time and the density fluctuations grow, while during the radiation epoch the density fluctuations are constant in time and the potential fluctuations decay.

The spectrum of fluctuations on large scales was unaffected by acoustic effects during the radiation epoch. The observed fluctuations of the Newtonian potential on these scales are observed to be nearly “scale-independent.” In Section 7.3 we will show how this is what is expected in models where fluctuations were seeded by quantum fluctuations of a scalar field during a early inflationary epoch.

In Section 7.4 we will show how the fluctuations result in CMB anisotropies. The measurement of these anisotropies has been a primary ingredient in the establishment of the Λ CDM model. Section 7.5 discusses CMB polarization and the information that has been obtained from its measurement.

CMB and large-scale structure data are interpreted within the framework of Λ CDM with adiabatic primordial fluctuations. Section 7.6 discusses models without CDM and/or adiabatic fluctuations and presents the difficulties such models have in accounting for observations.

Finally, in the last two sections we derive certain mathematical results anticipated in the preceding sections. Sections 7.7 presents the fluid-dynamic equations that allow one to derive the growth rates for Newtonian fluctuations in the linear regime. Section 7.8 shows how to calculate the trajectories and redshifts of photons in an inhomogeneous but isotropic universe.

7.1 A Spherical Collapse Model

In the standard CDM scenario, the formation of a gravitationally bound object from a region with a small over-density is a three-step process:

1. The expansion of the region to a maximal extent.
2. The contraction of the region until the formation of a bound (virialized) structure.
The process is “non-dissipative” in the sense that the Newtonian energy of the region is approximately conserved during the collapse.
3. The “dissipative” contraction of the baryonic component via radiation of photons.

Clusters of galaxies and galactic halos are primarily the results of the first two steps. The third step is essential in the formation of baryonic structures like galactic disks.

In this section we will follow the first two steps for a perturbed spherical region with an initial small over-density. Because of the spherical symmetry, it is plausible that the dynamics of the over-dense region will be independent of the universal dynamics. In general relativity, this independence is, in fact, guaranteed by Birkhoff’s theorem, the analog of Gauss’s theorem. In a critical universe, an initial infinitesimal over-density is sufficient to bind the region which must eventually detach from the expansion and contract.

As we have emphasized, the process of structure formation is, in reality, not spherically symmetric. Large N-body simulations [136] show that non-spherical collapse results in a “filamentary” distribution of matter with clusters of galaxies forming at the intersection of filaments. The structures seen in redshift surveys (Fig. 2.18) are a reflection of this density structure. The spherical collapse model is, however, useful for developing an intuition for structure formation and for estimating orders of magnitude.

The scale factor for the universe as a whole is assumed to obey the $k = 0$, matter-dominated Friedmann equation

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G \bar{\rho}}{3} \quad \bar{\rho} \propto a^{-3} \quad (7.1)$$

which has the solution

$$a(t) \propto t^{2/3} \quad \frac{\dot{a}}{a} = \frac{2}{3t} \quad \bar{\rho} = \frac{1}{6\pi G t^2} \quad (7.2)$$

Inside the spherical region of radius $R(t)$, we assume that the particles obey a Hubble law with a slightly lower expansion rate, $\dot{R}/R < \dot{a}/a$. The total mass in the region M_R is therefore time-independent. A particle at the edge of the region then obeys the standard Newtonian law

$$\ddot{R} = -\frac{GM_R}{R^2}. \quad (7.3)$$

The solution can be found by simply imposing energy conservation:

$$\dot{R}^2/2 - GM_R/R = -GM_R/R_{\max} \quad (7.4)$$

where the first and second terms on the l.h.s. are the kinetic and potential energies per unit mass and the r.h.s. is the total given by the potential energy at maximum radius, R_{\max} . The energy conservation equation can be integrated to give $t(R)$

$$t = \frac{1}{\sqrt{2GM_R}} \int_0^{R(t)} \frac{dR}{\sqrt{1/R - 1/R_{\max}}} \quad (7.5)$$

The function $R(t)$ calculated numerically is shown in Fig. 7.3.

We are mostly interested in early times when the density contrast is small and the expansion rate of the spherical region is close to the Universal rate. During this time, we can evaluate the integral (7.5) by treating R/R_{\max} as a small parameter:

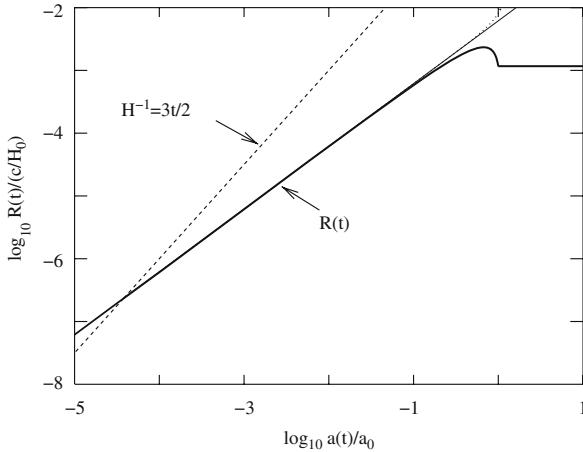


Fig. 7.3 The radius of an over-dense region, $R(t)$, as a function of the universal scale parameter, $a(t)$, in a universe with only CDM. At early times $R(t) \propto a(t) \propto t^{2/3}$ before reaching a maximum value and then decreasing by a factor two (virialization). The over-density $\Delta\rho/\bar{\rho}$ was chosen to be equal to 10^{-4} at the moment, $t = t_{\text{enter}}$, when $R(t)$ is equal to the Hubble distance (shown as the dashed line). The collapse of the region thus occurs at a time when $a(t) \sim 10^4 a(t_{\text{enter}})$

$$t \sim \frac{1}{\sqrt{2GM_R}} \int_0^{R(t)} R^{1/2} dR (1 + R/(2R_{\max})), \quad (7.6)$$

$$\Rightarrow R(t)^{3/2} \sim 3t \sqrt{\frac{GM_R}{2}} \left[1 - \frac{3R(t)}{10R_{\max}} \right]. \quad (7.7)$$

As expected, this gives $R(t) \propto t^{2/3} \propto a(t)$ for $R \ll R_{\max}$. The density in the perturbed region is

$$\rho_R(t) = \frac{M_R}{4\pi R^3/3} \sim \frac{1}{6\pi G t^2} \left[1 + \frac{3R(t)}{5R_{\max}} \right]. \quad (7.8)$$

The pre-factor is just the mean Universal density given by (7.2). The density contrast is thus

$$\delta(t) \equiv \frac{\rho_R - \bar{\rho}}{\bar{\rho}} = \frac{3R(t)}{5R_{\max}} \propto a(t) \quad \delta(t) \ll 1. \quad (7.9)$$

Since in the linear regime $R(t) \propto a(t)$ we make the important conclusion that $\delta(t) \propto a(t)$ for $\delta(t) \ll 1$.

To understand the evolution of small amplitude density fluctuation, it is useful to consider the associated fluctuations of the Newtonian potential. Very loosely, potential fluctuations can be defined as proportional to the difference, ΔM_R , between M_R and the mass expected in an unperturbed region of the same volume, $\bar{\rho}4\pi R^3/3$:

$$\Delta\phi(t) \equiv \frac{G\Delta M_R}{R} = \frac{4\pi G}{3}\bar{\rho}(t)\delta(t)R(t)^2 \quad (7.10)$$

The second form tells us that the potential fluctuation is time-independent because $\bar{\rho} \propto a^{-3}$, $\delta \propto a$, and $R^2 \propto a^2$. It is useful to evaluate the potential fluctuation explicitly at t_{enter} , the moment when $R(t)$ equals the Hubble distance, $(\dot{a}/a)^{-1}$:

$$R(t_{\text{enter}})^{-2} = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\bar{\rho}}{3} \quad (7.11)$$

giving

$$\Delta\phi = \delta(t_{\text{enter}})/2 \quad (\text{independent of } t) \quad (7.12)$$

The potential perturbation is of the order of the density contrast at t_{enter} . We will see in Sect. 7.3 that in inflationary models, it is the fluctuations at t_{enter} that are determined by the fluctuations of the scalar fields driving inflation.

Finally, by setting $M_R = \rho(t_{\text{enter}})(4\pi/3)R(t_{\text{enter}})^3$ and using (7.11) and (7.2) we get:

$$t_{\text{enter}} = \frac{4GM_R}{3} , \quad (7.13)$$

i.e. the more massive the perturbation, the later it enters the Hubble radius. Substituting this into (7.9) we get a relation between the density contrast at t_{enter} and R_{max} :

$$\delta(t_{\text{enter}}) = (6/5) \frac{GM_R}{R_{\text{max}}} , \quad (7.14)$$

which tells us that the energy of the perturbation is of the order of the density contrast at t_{enter} or, via (7.12), the potential fluctuation in the linear regime.

We now leave the linear regime to study the formation of a bound structure. The time of maximum expansion is:

$$t_{\text{max}} = \frac{1}{\sqrt{2GM_R}} \int_0^{R_{\text{max}}} \frac{dR}{\sqrt{1/R - 1/R_{\text{max}}}} = \frac{R_{\text{max}}^{3/2}}{\sqrt{2GM_R}} . \quad (7.15)$$

At this time, the density contrast is

$$\frac{M_R/(4\pi R_{\text{max}}^3/3)}{\bar{\rho}} - 1 = \left(\frac{3\pi}{4}\right)^2 = 5.55 , \quad (7.16)$$

where the mean density is given by (7.2). After t_{max} , the matter in the central region starts to contract. In the absence of random velocities inside the perturbed region,

the collapse would be to a black hole. The random velocities cause the production of a virialized object in which the mean kinetic energy is one half the magnitude of the mean gravitational potential energy (virial theorem). From (7.4), this occurs when $R = R_{\max}/2$. The time required for contraction is of the order t_{\max} . (If the object collapsed to a black hole, symmetry would require that the expansion time be exactly the contraction time.) During the contraction from t_{\max} to $2t_{\max}$, $a(t)$ is multiplied by a factor $\sim 2^{2/3}$ so $\bar{\rho}$ is reduced by a factor 2². The density contrast when the bound object forms at $\sim 2t_{\max}$ is therefore roughly

$$\frac{\rho_R(2t_{\max})}{\bar{\rho}(2t_{\max})} \sim 5.55 \times 2^3 \times 2^2 = 177. \quad (7.17)$$

The depth of the potential well of the object are of the order

$$\langle -\phi \rangle \sim \frac{GM_R}{R_{\max}/2} \sim \delta(t_{\text{enter}}) \quad (7.18)$$

In a CDM universe, the potential depth of bound objects and the associated velocity dispersions are directly related to the density contrast at Hubble entry.

Equations (7.9), (7.12), (7.17), and (7.18) guide our intuition for the formation of structure in a CDM-only universe:

- linear regime: density contrasts $\delta(t) \propto a(t)$ and potential fluctuations independent of time $\Delta\phi \sim \delta(t_{\text{enter}})$.
- structure formed with potentials wells given by $\phi \sim \delta(t_{\text{enter}})$ and density contrast (at time of formation) $\Delta\rho/\rho \sim 200$.

7.1.1 A CDM Universe with Scale Invariant Fluctuations

The universe does not have just one spherical perturbed region and to make contact with the observed features of the universe we have to describe the spectrum of perturbations. Very qualitatively, one can think of the universe as a superposition of many perturbations, each characterized by two quantities: its time independent mass, M_R , and its density contrast or potential fluctuation at the time when it entered the Hubble radius, $\delta(t_{\text{enter}})$. How to describe mathematically this spectrum of perturbations will be explained in Sect. 7.2.

In a “scale-invariant universe,” perturbations of all sizes enter the Hubble radius with, on average, the same density contrast: $\langle \delta(t_{\text{enter}})^2 \rangle^{1/2}$ independent of M_R . The time-independent potential fluctuations are therefore also independent of scale. As such, this spectrum of fluctuations is referred to as a “scale-invariant” spectrum, or Harrison-Zel’dovich spectrum. At large scales, this closely approximates our universe which has potential fluctuations of $\sim 10^{-5}$:

$$\langle \delta(t_{\text{enter}})^2 \rangle^{1/2} \sim \langle \Delta\phi^2 \rangle^{1/2} \sim 10^{-5} \quad (7.19)$$

From (7.10), this implies

$$\langle \delta(t)^2 \rangle^{1/2} \sim 10^{-5} \frac{d_H(t)^2}{R(t)^2} \quad (7.20)$$

The universe is homogeneous ($\delta \ll 1$) on scales $R \gg d_H/\sqrt{\Delta\phi}$. For our universe, this corresponds at t_0 to $R \gg 15h_{70}^{-1}\text{Mpc}$, in qualitative agreement with observations.

We will see that the evidence for scale-invariant fluctuations comes from the fluctuations of the CMB temperature on scales > 1 deg. where one observes $\Delta T/T \sim 2 \times 10^{-5}$ independent of angular scale. Loosely speaking, this comes about because the photons that are liberated at recombination suffer a gravitational redshift as they climb out of potential wells or fall off potential hills. The first observation these temperature fluctuation in 1991 by the COBE satellite [148] was a decisive step in the study of structure formation giving the “initial conditions” at t_{enter} .

In the scale-invariant CDM universe, bound structures of all sizes would consist of potential wells of the same depth, $\langle \delta(t_{\text{enter}})^2 \rangle^{1/2}$. Since the time of formation (7.15) is proportional to $M_R/\delta(t_{\text{enter}})^{3/2}$, small structures form first, with larger structures involving the mergers of smaller structures. This is called “bottom-up” structure formation. Formation of larger and larger bound structures continues as long as $\Omega_M \sim 1$.

7.1.2 Effects of Baryons and Photons

The simple conclusions of the previous section are modified in a very interesting way by the presence of baryons, photons, and neutrinos. Before recombination, photons and baryons are coupled via Compton and Rutherford scattering on electrons to form a fluid that supports acoustic waves. As long as the photons dominate the pressure of the plasma, the speed of sound is $c_s \sim c/\sqrt{3}$. As illustrated in Fig. 5.5, the perturbation will initiate a sound wave that will propagate away from the initial perturbation at the speed of sound. The sound wave stops at recombination, leaving the excess baryons on a shell of radius equal to the sound horizon at recombination $a_{\text{rec}}\chi_s$. As for neutrinos, they cease to scatter on other particles after decoupling at $T \sim 1 \text{ MeV}$, after which they can “free stream” away from the initial perturbation at velocity c .

The effect of photon–baryon wave propagation and neutrino free-streaming is to suppress the growth of the CDM density contrast if the neutrino and photon–baryon perturbations have the time to separate spatially from the initial perturbation before matter-radiation equality, t_{eq} . This effect comes about because before t_{eq} the photon–baryon plasma dominates the CDM gravitationally. The CDM particles in the perturbed region were initially given Hubble velocities to compensate the extra gravitation of the over-density. With the exit of the gravitationally dominant photons and neutrinos, the CDM velocities no longer act to maintain the co-moving radius

of the perturbation. The CDM perturbation thus starts to disperse. The CDM density contrast therefore does not grow in proportion to $a(t)$ as it would in the absence of the photons and neutrinos.

The requirement that there be enough time before t_{eq} for the CDM the photon–baryon perturbations to separate is simply

$$R(t_{\text{eq}}) < a(t_{\text{eq}})\chi_s(t_{\text{eq}}), \quad (7.21)$$

where χ_s is given by (5.28) with the upper limit of integration \hat{a}_{rec} replaced with \hat{a}_{eq} . For $(\Omega_M, \Omega_\Lambda) = (0.27, 0.73)$ this gives $a_0\chi_s(t_{\text{eq}}) \sim 70 \text{ Mpc}$ so the necessary requirement for suppression of density-contrast growth is

$$R(t_0) < 70 \text{ Mpc}. \quad (7.22)$$

At scales below this, we can expect density contrasts and potential fluctuations to be less than one would expect for a pure CDM universe. Note that since the speed of sound is of the same order as the speed of light, the effects of the free-streaming neutrinos act on the same distance scale.

7.2 The Spectrum of Density Fluctuations

In the previous section, we discussed the time evolution of simple spherical perturbations in an otherwise homogeneous universe. The discussion allowed us to understand certain characteristics of our universe. However, in order to describe the details of the complicated density distribution, it is convenient to Fourier decompose the density field.

We consider the energy density, $\rho(\mathbf{r}, t)$, in a volume V of mean density $\bar{\rho}$. We suppose that the universe is flat ($k = 0$) so as to avoid complications of spatial curvature. The space coordinates \mathbf{r} are thus co-moving Cartesian coordinates. The variations of the density around $\bar{\rho}$ are described by a function $\delta(\mathbf{r}, t)$:

$$\rho(\mathbf{r}, t) = \bar{\rho}(t) (1 + \delta(\mathbf{r}, t)). \quad (7.23)$$

We expand $\delta(\mathbf{r}, t)$ in a Fourier series:

$$\delta(\mathbf{r}, t) = V^{-1/2} \sum_{\mathbf{k}} \delta_{\mathbf{k}}(t) \exp(i \mathbf{k} \cdot \mathbf{r}), \quad (7.24)$$

Note that $\delta_{\mathbf{k}}$ has the dimension of volume $^{1/2}$. The sum is over all waves that average to zero over the region, $\langle \exp(i \mathbf{k} \cdot \mathbf{r}) \rangle = 0$. This is equivalent to imposing periodic boundary conditions. The reality of ρ imposes $\delta_{\mathbf{k}}^* = \delta_{-\mathbf{k}}$. Compared to normal Fourier series, the only peculiar thing is that since \mathbf{r} is co-moving, the wave vector \mathbf{k} must be co-moving, in the sense that the associated wavelength λ_k expands with the universe:

$$\lambda_k(t) = \frac{2\pi}{k} \hat{a}(t) . \quad (7.25)$$

The wave number k has dimensions of inverse length and is related to the present-day wavelength, $k = 2\pi/\lambda(t_0)$. As with the radius $R(t)$ of the perturbed region of Sect. 7.1, at early times $\lambda_k(t) < d_H(t)$ and then ‘‘enters’’ the Hubble radius at $t_{\text{enter}}(k)$.

The use of $\delta_{\mathbf{k}}(t)$ has three important advantages over direct use of $\delta(\mathbf{r}, t)$:

1. Large-scale (small k) effects are separated neatly from small-scale (large k) effects. This is important because we will see that different physics acts on different scales.
2. As long as the density contrast is small, the $\delta_{\mathbf{k}}(t)$ evolve in time independently of each other obeying linear ordinary differential equations, some of which will be derived in Sect. 7.7. This will lead to simple behavior of the $\delta_{\mathbf{k}}$ as a function of $a(t)$. For instance, using the linear growth of density contrast seen in Sect. 7.1, we can anticipate linear growth of $\delta_{\mathbf{k}}(t)$ as long as the density contrast is small:

$$\delta_{\mathbf{k}}(t) \propto a(t) \quad \Omega_{\text{CDM}} = 1 \quad (7.26)$$

3. The initial values of the real and imaginary parts of the $\delta_{\mathbf{k}}$ in inflationary models are Gaussian random variables distributed about zero. Measurement $\langle |\delta_{\mathbf{k}}|^2 \rangle$ thus gives us directly information on the fluctuations of the scalar field driving inflation.

The last point emphasizes the fact that individual $\delta_{\mathbf{k}}$ are of little or no interest. Rather, we are interested in the variance of the $\delta_{\mathbf{k}}$ given by the power spectrum, $P(k)$:

$$P(k) \equiv \langle |\delta_{\mathbf{k}}|^2 \rangle , \quad (7.27)$$

where the average is over \mathbf{k} with $|\mathbf{k}|$ near k . Note that $P(k)$ has dimensions of volume. The power spectrum is directly related to the variance of $\delta(\mathbf{r}, t)$:

$$\langle [\delta(r, t)]^2 \rangle = V^{-1} \int_V d^3r \delta^2(r, t) = V^{-1} \sum_{\mathbf{k}} |\delta_{\mathbf{k}}|^2 , \quad (7.28)$$

where we used the orthogonality of the $\exp(i\mathbf{k} \cdot \mathbf{r})$. It is useful to express this as an integral over $k = |\mathbf{k}|$. This can be done by using the fact that the number of modes with $k < k_{\text{max}}$ is $(4\pi/3)V(k_{\text{max}}/2\pi)^3$. We thus make the usual replacement

$$V^{-1} \sum_{\mathbf{k}} \rightarrow \frac{1}{(2\pi)^3} \int d^3k \rightarrow \frac{1}{2\pi^2} \int_0^\infty k^2 dk \quad (7.29)$$

$$\Rightarrow \langle \delta^2(r, t) \rangle = \frac{1}{2\pi^2} \int_0^\infty k^2 dk P(k) , \quad (7.30)$$

where we have replaced $|\delta_{\mathbf{k}}|^2$ by its average.

Since the power spectrum has dimensions of volume, its physical meaning is perhaps not immediately obvious. A more easily understood function is the dimensionless function $\Delta(k)$:

$$\Delta^2(k) \equiv \frac{k^3 P(k)}{2\pi^2} \quad \Rightarrow \quad \langle \delta^2(r, t) \rangle = \int_0^\infty \frac{dk}{k} \Delta^2(k), \quad (7.31)$$

The function $\Delta^2(k)$ gives the contribution per unit $\ln k$ to the density fluctuations. It is thus a measure of the density fluctuations on the scale k^{-1} .

Figure 7.4 shows the theoretical present-epoch power spectrum $P(k)$ and density fluctuations $\Delta(k)$ for $(\Omega_M, \Omega_\Lambda) = (0.27, 0.73)$ and for $(1, 0)$. The method of calculation will be explained below in Sect. 7.2.1. For small k , $P(k) \propto k$ until reaching a maximum near $k \sim 0.02 h \text{Mpc}^{-1}$, for $(0.27, 0.73)$, or $k \sim 0.08 h \text{Mpc}^{-1}$, for $(1, 0)$.¹ After the maximum, they fall roughly like $1/k^3$. The density fluctuations $\Delta(k)$ are small at large scale (small k) increasing like $\sim k^2$ before reaching unity at $k \sim 0.1 \text{ Mpc}^{-1}$. This value of k , corresponding to the length scale $k^{-1} \sim 10 \text{ Mpc}$ is the dividing line between the homogeneous universe at large scale and the very clumpy universe at small scale.

Note that $\Delta(k)$ does not decrease with increasing k so the integral (7.31) for the density variance shows no sign of converging. This has no practical importance because we cannot measure the density at individual points. We must always average over regions of non-vanishing volume. In this case, the effective variance of the density function is a function σ_r giving the fluctuations of mass within spheres of radius $r h^{-1} \text{Mpc}$. It is relatively straightforward (Exercise 7.1) to show that

$$\sigma_r^2 = \int_0^\infty \frac{dk}{k} \left[\frac{3j_1(kr)}{kr} \right]^2 \Delta^2(k), \quad (7.32)$$

where $j_1(x) = \sin(x)/x - \cos(x)$. The factor $(3j_1(kr)/kr)^2$ is unity for $kr \ll 1$ and $\rightarrow 0$ for $kr \gg 1$. The mass fluctuation (7.32) is thus the same as the density fluctuation (7.31) except that the integral is cut off for $k > 1/r$.

We saw in Sect. 7.1 that the potential depth of a spherical perturbation in a CDM universe remained constant in time. As such, we can expect that the fluctuation of the Newtonian potential will be very useful in understanding the time evolution of the power spectrum. As with the density, we can expand the potential in plane waves

$$\phi(\mathbf{r}, t) = V^{-1/2} \sum_{\mathbf{k}} \phi_{\mathbf{k}} \exp(i \mathbf{k} \cdot \mathbf{r}), \quad (7.33)$$

The $\phi_{\mathbf{k}}$ can be related to the $\delta_{\mathbf{k}}$ by using the Poisson equation, $\nabla^2 \phi = 4\pi G \rho$:

¹ In discussing the power spectrum, we will follow the convention of the literature by using h rather than h_{70} .

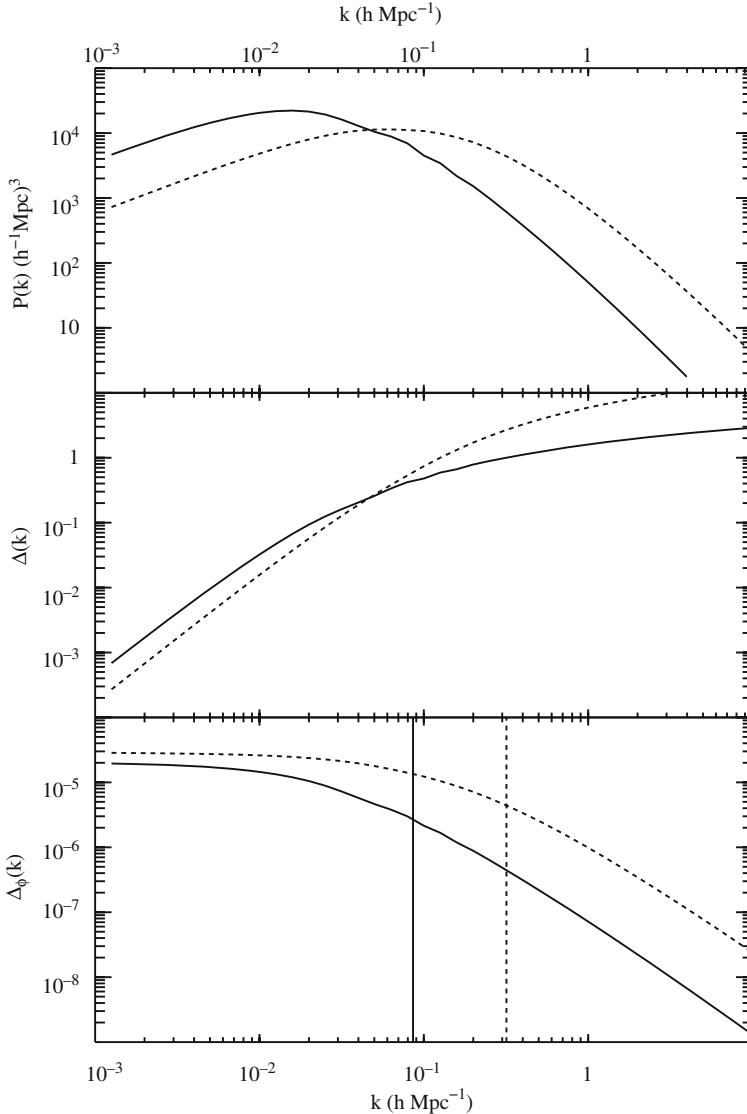


Fig. 7.4 The present-day power spectrum $P(k)$, the mass density fluctuation spectrum $\Delta(k)$, and the gravitational potential fluctuation spectrum $\Delta_\phi(k)$, calculated following the method of [139] with the normalization (small k) taken from CMB fluctuations [1], and with the transfer function of [140]. The solid lines are for the standard cosmological parameters $(\Omega_M, \Omega_A) = (0.27, 0.73)$ of [1]. The dashed lines are for $(\Omega_M, \Omega_A) = (1, 0)$. The two spectra are shifted horizontally because of the differing values of k_{eq} (7.41) shown as the vertical lines on the potential spectrum plot. The calculation gives the “linear $P(k)$ ” assuming independent evolution of each mode k and, as such, underestimates $P(k)$ for the non-linear regime, $k > 0.1 h \text{Mpc}^{-1}$

$$\phi_{\mathbf{k}}(t) = \frac{4\pi G \bar{\rho}(t) \hat{a}(t)^2}{k^2} \delta_{\mathbf{k}}(t) = \frac{3\Omega_M(t)}{2} \frac{\hat{a}(t)^2}{k^2 d_H(t)^2} \delta_{\mathbf{k}}(t) \quad (7.34)$$

where the \hat{a}^2 factor comes from the fact that the space derivatives in the Poisson equation are with respect to laboratory coordinates. In a $\Omega_{\text{CDM}} = 1$ universe, $\delta_{\mathbf{k}} \propto a(t)$ and $d_H(t)^2 \propto a(t)^3$, so the $\phi_{\mathbf{k}}$ are time-independent, as expected.

Calculating $\phi(\mathbf{r})^2$ and then averaging over \mathbf{r} we get the fluctuations of the potential

$$\langle \phi^2 \rangle = \frac{1}{2\pi^2} \int \frac{dk}{k} k^3 \langle |\phi_{\mathbf{k}}|^2 \rangle = \frac{9\Omega_M^2}{4} \frac{\hat{a}^4}{d_H^4} \int_0^\infty \frac{dk}{k} \frac{\Delta^2(k)}{k^4} \quad (7.35)$$

In analogy with $\Delta^2(k)$ defining the mass fluctuation on the scale k^{-1} , we define the potential fluctuations as

$$\Delta_\phi^2(k, t) \equiv \frac{9\Omega_M^2}{4} \frac{\hat{a}^4 \Delta^2(k)}{k^4 d_H^4} \quad (7.36)$$

We see again that the potential is time independent for $\Omega_{\text{CDM}} = 1$ since $\Delta^2(k) \propto a^2$ and $d_H^2 \propto a^3$. The potential is scale invariant if $\Delta^2(k) \propto k^4$, i.e. $P(k) \propto k$.

Figure 7.4 shows the potential fluctuation derived from the power spectrum shown in the same figure. The spectrum is flat at small k as expected for a scale-invariant spectrum. For $k > 0.1$ the potential fluctuations fall roughly like k^{-4} . Because of this fall-off, the integral for the potential fluctuations (7.35) appears to converge, consistent with the observation that, apart from the surfaces of blackholes, the gravitational potentials in the universe are $\ll 1$.

7.2.1 The Power Spectrum for a CDM-Baryon Universe

The two power spectra in Fig. 7.4 were calculated by choosing an initial condition for the $\delta_{\mathbf{k}}(t)$ (for each particle species) and then following their time development with the appropriate differential equations. Examples of this time development for three modes are shown in Fig. 7.5. The $\delta_{\mathbf{k}}(t)$ have time dependences that are different according to whether t is in the radiation or matter epoch and to whether the mode is inside or outside the Hubble radius. We can summarize the behavior as follows.

- Before Hubble entry, $t < t_{\text{enter}}(k)$, the modes grow like a^2 in the radiation epoch and a^1 in the matter epoch. This behavior outside the Hubble radius depends strongly on the choice of coordinates and the simple behavior in Fig. 7.5 is that for the so-called synchronous gauge corresponding to comoving coordinates that are not in uniform expansion.

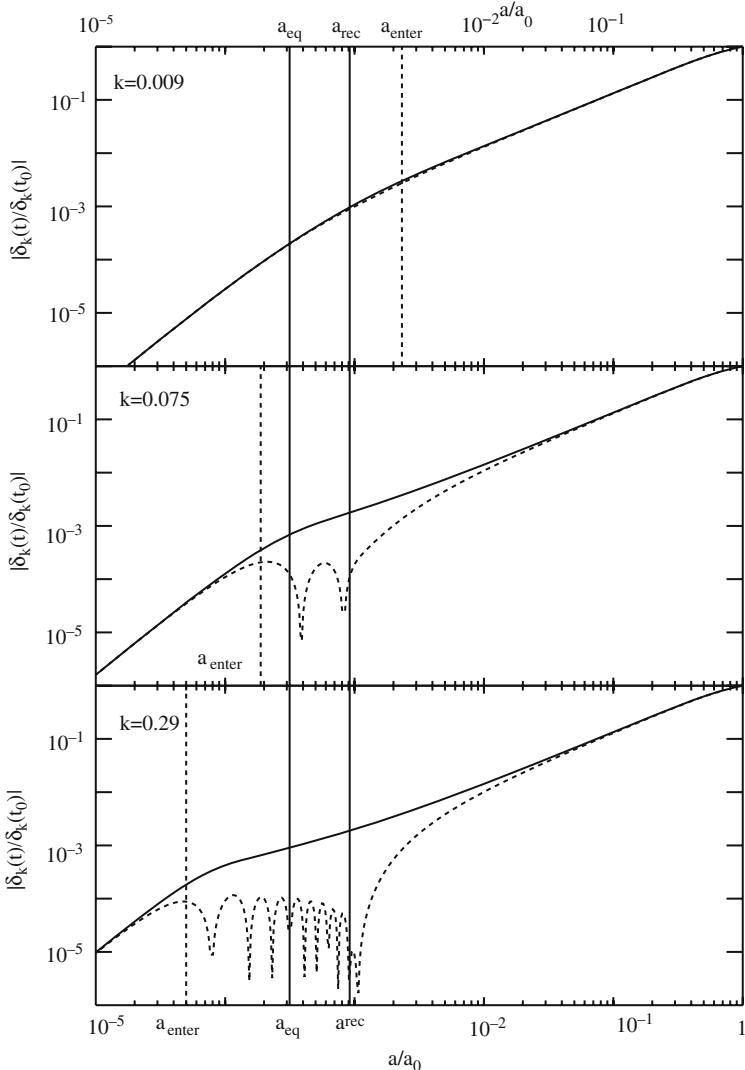


Fig. 7.5 The time evolution of $\delta_k(t)$ for the CDM component (solid curves) and for the baryonic component (dashed curves) for three values of k with $(\Omega_M, \Omega_A) = (0.27, 0.73)$ [152]. The three dashed vertical lines show $a_{\text{enter}}(k)$ while the solid vertical lines show a_{eq} and a_{rec} . The longest wavelength mode, $k = 0.009 h \text{Mpc}^{-1}$, enters the Hubble radius during the matter epoch. The shortest wavelength mode, $k = 0.29 h \text{Mpc}^{-1}$, enters during the radiation epochs and consequently has its growth retarded until t_{eq} . A zoom is shown in Fig. 7.13

After Hubble entry, the modes have the following behavior.

- CDM: for $t > t_{\text{eq}}$ $\delta_{\mathbf{k}} \propto a(t)$. For $t < t_{\text{eq}}$ the $\delta_{\mathbf{k}}$ grow more slowly because of the gravitational dominance of the baryon–photon plasma.
- Baryons: for $t < t_{\text{rec}}$, the baryonic component is coupled to the photons and oscillates as an acoustic wave. For $t > t_{\text{rec}}$ the baryons decouple from the photons and fall toward the CDM.

This general behavior allows us to understand the form of the power spectrum. It is simplest to consider first the potential fluctuations because in a scale-invariant CDM-only universe, potential fluctuations are time-independent:

$$\Delta_{\phi}(k, t) = (2.97 \pm 0.06) \times 10^{-5} \quad \text{CDM only} \quad (7.37)$$

where the numerical values is taken from the large-scale CMB anisotropies of WMAP [1] via the Sachs–Wolfe (SW) effect (Sect. 7.4).² The real universe is neither scale-invariant nor CDM-only and we take into account these facts with three factors:

$$\Delta_{\phi}(k, t) = 2.97 \times 10^{-5} \left(\frac{ck}{H_0} \right)^{(n_s-1)/2} T(k, t) G(\Omega_M, \Omega_A, t) \quad (7.38)$$

The first correction $(ck/H_0)^{(n_s-1)/2}$ allows for the possibility that the potential fluctuations are not scale invariant by introducing the “tilt,” n_s . The WMAP measurements combined with SNIa and BAO measurements give [1]

$$n_s = 0.960 \pm 0.013 , \quad (7.39)$$

This value of n_s close to unity means that the primordial potential fluctuations vary only by $\sim 10\%$ when k varies by a factor 10.

The second factor, $T(k)$, is called the transfer function. It takes into account the fact that during the radiation epoch, the amplitudes of the dominant components did not grow with time because they either propagated as acoustic waves (photon–baryon plasma) or free streamed (neutrinos). This has the effect of suppressing the amplitude of short wavelength modes, $k > k_{\text{eq}}$, where k_{eq} is given by (7.41). The transfer function can be calculated efficiently with publicly available computer programs [152] but we will estimate it below.

Finally, the growth factor, $G(\Omega_M, \Omega_A, t)$, takes into account the fact the density fluctuations stop growing when Ω_M drops significantly below unity. This causes the potential fluctuations to decay. By definition, $G(\Omega_M = 1, \Omega_A = 0, t) = 1$. For the standard cosmological parameters, one finds $G(0.27, 0.73, t_0) \sim 0.76$ [144].

² We assume that there are no large CMB anisotropies generated by gravitational waves.

We will now estimate the transfer function for $k > k_{\text{eq}}$. During the radiation epoch, perturbations in the dominant component oscillate as acoustic waves with time-independent amplitudes. This gives $P(k, t) \sim P(k, t_{\text{enter}})$ so the potential fluctuations decay like $\Delta_\phi^2(k) \propto a(t)^{-4}$. This decay continues until a_{eq} at which point matter domination starts and $\Delta_\phi^2(k)$ becomes time-independent again.

To calculate the factor by which Δ_ϕ decays, we need $a(t_{\text{enter}}(k)) = a_{\text{enter}}(k)$. It is the solution of

$$\frac{2\pi}{k} \hat{a}_{\text{enter}} = \frac{c/H_0}{(\Omega_M \hat{a}_{\text{enter}}^{-3} + 1.68 \Omega \gamma \hat{a}_{\text{enter}}^{-4})^{1/2}}. \quad (7.40)$$

Modes with large k have $a_{\text{enter}} < a_{\text{eq}}$. The dividing line between small and large k is called k_{eq} and is found by setting $a_{\text{enter}} = a_{\text{eq}} = 1.68 \Omega \gamma / \Omega_M$ in (7.40)

$$k_{\text{eq}} = 2\pi \frac{\Omega_M}{c/H_0} \sqrt{\frac{2}{1.68 \Omega \gamma}} = 0.06 h_{70}^{-1} \text{Mpc}^{-1} \frac{\Omega_M}{0.27}. \quad (7.41)$$

We can now calculate the amount by which the potential fluctuations decay. Modes entering during the radiation epoch have a_{enter} given by

$$\hat{a}_{\text{enter}} = \frac{2\pi}{H_0^{-1} k} \sqrt{1.68 \Omega \gamma} \quad (7.42)$$

The total decay of Δ_ϕ^2 between a_{enter} and a_{eq} is

$$\left(\frac{a_{\text{enter}}}{a_{\text{eq}}} \right)^4 = \left(\frac{2\pi}{H_0^{-1} k} \right)^4 \frac{\Omega_M^4}{(1.68 \Omega \gamma)^2}, \quad (7.43)$$

which implies

$$T^2(k \gg k_{\text{eq}}, t > t_{\text{eq}}) \propto 1/k^4. \quad (7.44)$$

This gives the following behavior for the functions in Fig. 7.4 for $k > k_{\text{eq}}$:

$$\Delta_\phi(k) \propto k^{-2} \quad P(k) \propto k^{-3} \quad \Delta(k) \sim \text{constant} \quad (7.45)$$

This is qualitatively what is seen. The power spectrum does not fall quite as rapidly as our estimate because the CDM contrast does, in fact, grow logarithmically with time during the radiation epoch.

7.2.2 Measurements of the Power Spectrum

The power spectrum can be measured with three types of measurements covering different ranges of k . For $k < 0.01h_{70}\text{Mpc}^{-1}$, $P(k)$ is tightly constrained by the CMB measurements which give, via the Sachs–Wolfe effect (Sect. 7.4), the primordial potential fluctuations.

For $k > 0.01h_{70}\text{Mpc}^{-1}$, $P(k)$ can be measured by estimating density contrast, $\delta(\mathbf{r})$, representing it as a sum of plane waves (7.24) and then averaging the coefficients (7.27). The density cannot be directly measured so one must measure the density of something else whose density is expected to be a function of the matter density. For $0.01h_{70}\text{Mpc}^{-1} < k < 0.001h_{70}\text{Mpc}^{-1}$ the simplest such density “tracers” are galaxies, which can be easily counted in redshift surveys. An example of a power spectrum derived for SDSS galaxies is shown in Fig. 7.6.

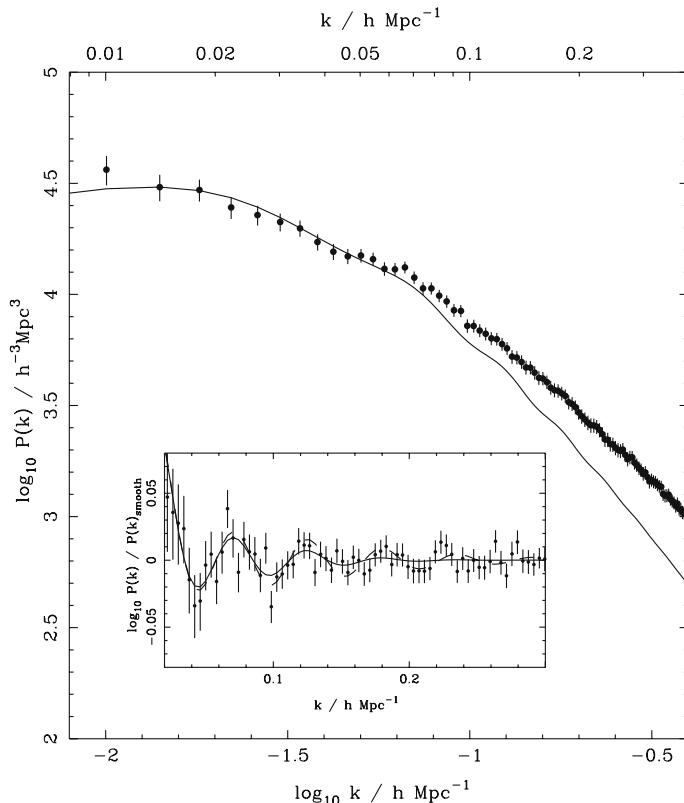


Fig. 7.6 The power spectrum measured with SDSS galaxies [141]. The *line* shows the calculated $P(k)$ using parameters similar to those of WMAP [1]. The excess power observed for $k > 0.1$ may be due to k -dependent bias in the galaxy sample that is not included in the analysis. The *inset* shows the baryon wiggles caused by BAO (Sect. 7.2.3)

The problem, of course, is that fluctuations in numbers of galaxies do not necessarily faithfully trace fluctuations in total density. One generally supposes that the density contrast is proportional to the contrast of the number density of galaxies:

$$\frac{\Delta n_{\text{gal}}}{n_{\text{gal}}} = b \frac{\Delta \rho}{\rho}, \quad (7.46)$$

where b is the “bias” parameter. This parameter can, a priori, be a complicated function of scale and environment, so it is the difficulty in estimating b that dominates the systematic uncertainties in measurements of $P(k)$.

One possible route to measuring b proceeds as follows. Cosmological N-body simulations are performed yielding a density map and catalog of simulated bound objects of known mass to be used as density tracers. The bias between the number density of objects and the mass density can be read off from the simulation output. The next step is to relate the objects of known mass in the simulation to the objects of known luminosity observed in redshift surveys. This can be done, e.g., by using weak-lensing data (Fig. 2.14) [142]. From this, one matches the observed objects used in (7.46) to the objects of known bias in the simulation, thus determining the bias of the observed objects.

For $k > 0.001 h_{70} \text{Mpc}^{-1}$, the power spectrum has been measured by using Lyman- α absorption of quasars to measure the density of hydrogen atoms along the line of sight [143]. A typical spectrum used to do this is shown in Fig. 6.8 and the technique to deduce the density of hydrogen atoms are discussed in Section 5.7. As with the number density of galaxies, the difficulty is to relate the density of hydrogen atoms to the density of matter.

Finally, we mention that the spectrum of density fluctuations can be measured through the gravitational lensing effect of large-scale structure. As discussed in Sect. 3.8, density inhomogeneities generate correlations between the apparent orientation of galaxies that are nearby on the sky. This method has the advantage of being independent of biasing.

Since lensing comes about because of density inhomogeneities, $\Delta\rho$, we can expect that measurement of large-scale lensing will constrain the product of ρ and $\Delta\rho/\rho$, i.e. the product of Ω_M and $\sqrt{P(k)}$. Up to now, measurements of this effect [113] have constrained the product of $\Omega_M^{-0.64}$ and σ_8 (7.32), the integral of $P(k)$ giving the mass fluctuations in spheres of radius $8h^{-1}\text{Mpc}$:

$$\sigma_8 = (0.825 \pm 0.05) \left(\frac{\Omega_M}{0.27} \right)^{0.64} \quad (7.47)$$

to be compared with the value derived [1] from CMB anisotropies (WMAP), SNIa (SNLS), and BAO (SDSS) assuming the Λ CDM model:

$$\sigma_8 = 0.812 \pm 0.026. \quad (7.48)$$

7.2.3 The Correlation Function

The power spectrum is directly related to the (two-point) correlation function $\xi(r)$. Consider two points \mathbf{r}' and $\mathbf{r}' + \mathbf{r}$. Using the orthogonality of the functions $\exp(i\mathbf{k} \cdot \mathbf{r})$, it is simple to show that the product of their density contrasts, averaged over \mathbf{r}' , is

$$\langle \delta(\mathbf{r}') \delta(\mathbf{r}' + \mathbf{r}) \rangle_{\mathbf{r}'} = V^{-1} \sum_{\mathbf{k}} |\delta_{\mathbf{k}}|^2 e^{i\mathbf{k} \cdot \mathbf{r}}. \quad (7.49)$$

Changing the sum to an integral and integrating over the angle between \mathbf{r} and \mathbf{k} we find the correlation function:

$$\xi(r) = \frac{1}{2\pi^2} \int_0^\infty k^2 dk P(k) \frac{\sin kr}{kr}. \quad (7.50)$$

Similarly

$$P(k) = \frac{1}{2\pi^2} \int_0^\infty r^2 dr \xi(r) \frac{\sin kr}{kr}. \quad (7.51)$$

The correlation function for SDSS galaxies was shown in Fig. 5.7 with its peak at the sound horizon. The peak in $\xi(r)$ is reflected in “wiggles” in $P(k)$ superimposed upon the smooth form. From (7.51) we expect peaks in the wiggles at values of k for which the function $\sin(kr)/kr$ has a peak at $r = a_0 \chi_s$:

$$k_n = \frac{(n + 1/2)\pi}{a_0 \chi_s} = 0.075h \text{ Mpc}^{-1} \frac{n + 1/2}{2.5} \quad n = 2, 4, 6\dots \quad (7.52)$$

The expected wiggles have an amplitude of only $\sim 3\%$ of $P(k)$ but have been seen (e.g., Fig. 7.6) in the power spectra of the 2dfGRS and SDSS redshift surveys [4].

7.2.4 Redshift Distortions

“Redshift distortions” of the power spectrum result from the use of redshifts to determine radial positions. While the angular coordinates of a galaxy are, apart from small gravitational lensing effects, relatively unambiguous, the use of the redshift to estimate the radial coordinate has the obvious objection that the contribution of peculiar velocities to measured recession velocities has no relation to the distance from us. This is most spectacularly seen in galaxy clusters where the line-of-sight orbital velocities add or subtract randomly from the cluster recession velocity. This so-called “Finger-of-God” effect is illustrated in the simulation shown in Fig. 7.7. In the right-hand figure, galaxy clusters appear as elongated structures pointing toward the observer below the figure. This is simply because the random virial velocities in the bound structure smear out the redshift derived distances by a factor $\Delta v / H_0$.

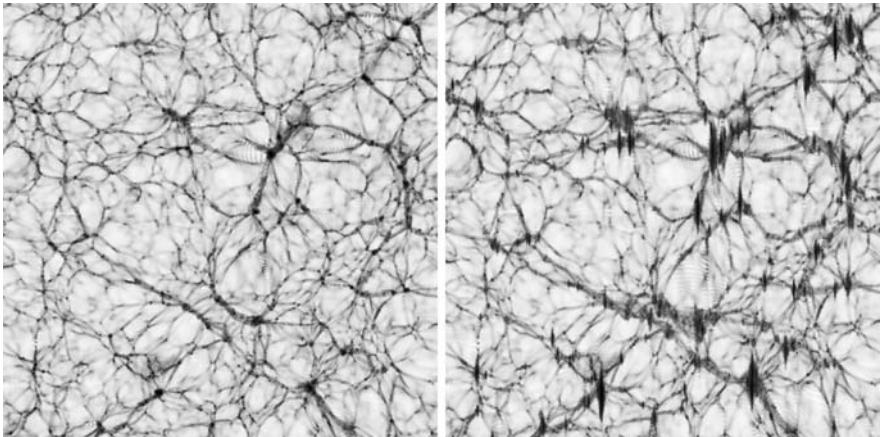


Fig. 7.7 The “finger-of-god” effect in a simulated redshift survey [145]. The two figures show the density distribution of a slice of a $(200h^{-1}\text{Mpc})^3$ volume. The *left figure* shows the true density field while the *right figure* shows the redshift distortions when the slice is observed from below the figure. The galaxy clusters appear as fingers pointing toward the observer because the random line-of-sight velocities in the galaxy clusters smear out the distances derived from the use of the Hubble law

For a velocity dispersion of $\Delta v = 3 \times 10^3 c$ for a large cluster, this corresponds to a distance dispersion of 12 Mpc. The fingers must be removed from the galaxy catalog in order to derive a reliable $P(k)$.

A more interesting effect, illustrated in Fig. 7.8, occurs at larger scale $> 10\text{ Mpc}$ where bound structures have not formed. In this case, objects are coherently attracted toward mass concentration giving them peculiar velocities. This has the effect of compressing the over-dense regions along the line of sight since far-side (near-side) objects have their apparent Hubble velocities reduced (increased).

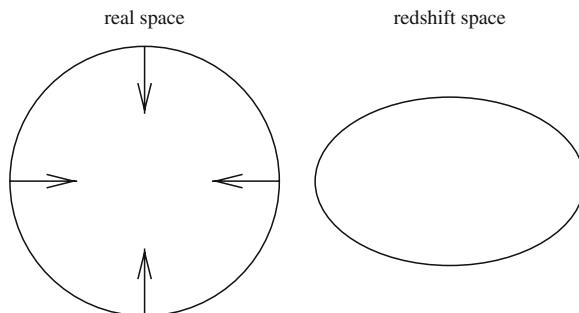


Fig. 7.8 Large-scale redshift distortions. The figure shows on the *left* a spherical over-dense region. Galaxies at the edge are attracted toward the center of the region generating peculiar velocities. When the line-of-sight peculiar velocities are added to the Hubble velocity, the spherical region is compressed along the line of sight

The size of the effect can be estimated using the spherical collapse model of Sect. 7.1. The peculiar velocity, v_p , of an object at the edge of the perturbed region of radius R is the difference between its velocity, \dot{R} , and the velocity it would have if it followed the universal expansion law:

$$v_p = \dot{R} - \frac{\dot{a}}{a} R \quad (7.53)$$

It is straight forward to show that in the linear regime the peculiar velocity is given by

$$v_p(t) \sim \delta(t) \frac{\dot{a}}{a} R(t) \quad (7.54)$$

Using (7.20), this gives a present day peculiar velocity about a mass concentration of radius R of

$$v_p(t_0)/c \sim 10^{-5} \frac{c}{H_0 R} \quad (7.55)$$

For $R \sim 40$ Mpc, this gives $v_p \sim 10^{-3} c$. If the mass concentration is at a redshift of $z = 0.1$, it is therefore compressed along the line of sight by a factor 10^{-2} . This small effect has been recently observed [25].

If uncorrected for, the large-scale redshift distortions make the observed power spectrum anisotropic:

$$P(k) \rightarrow P(\mathbf{k}_t, k_r), \quad (7.56)$$

where \mathbf{k}_t and k_r are the transverse (angular) and radial components of \mathbf{k} . While this may seem like an annoying complication, the anisotropies encode information on peculiar velocities and therefore on the rate at which large-scale structure is forming. This is redundant information since the peculiar velocities are determined by the density inhomogeneities. A consistency check can then be made by comparing the observed redshift distortions from those expected from the observed inhomogeneities. It has been suggested that this information may allow tests of the laws of gravitation at cosmological scale by determining if the Λ CDM mixture that explains the expansion rate can also explain the rate of structure formation [26].

7.3 The Primordial Spectrum from Inflation

Our universe is characterized by nearly scale-invariant potential fluctuations where all modes enter the Hubble radius with density fluctuations $\Delta\rho/\rho \sim 10^{-5}$. We need to explain this apparent miracle that all modes know what they should look like at t_{enter} . The explanation is suggested by Fig. 7.9 which shows the Hubble distance, $d_H(t)$, in a Λ CDM universe preceded by an inflationary epoch. The Hubble length

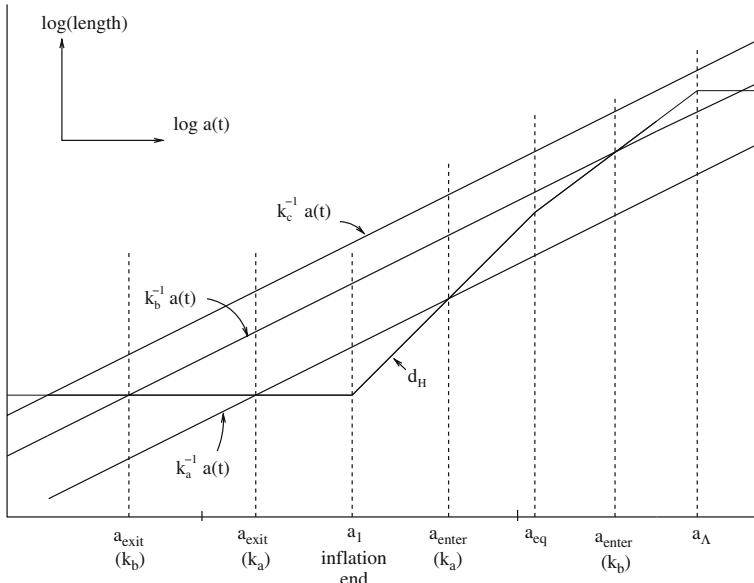


Fig. 7.9 The Hubble distance d_H in a model with inflation. A physical scale starts inside the Hubble radius, leaves at a_{exit} and then re-enters at a_{enter} . Small scales exit last and enter first. Mode k_a entered during the radiation epoch and mode k_b during the matter epoch. Mode k_c has not yet entered and never will if the vacuum energy continues to dominate

is approximately constant during inflation, increases like $a(t)^2$ during the radiation epoch and then like $a(t)^{3/2}$ during the matter epoch. Superimposed on the curve for $d_H(t)$ are the wavelengths associated with three wave vectors, $\lambda(t) = 2\pi \hat{a}(t)/k$. We see that modes that enter the Hubble radius during the radiation or matter epochs had previously been inside the Hubble radius during the inflationary epoch. Each mode is therefore characterized by an epoch of Hubble exit and Hubble entry, $a_{\text{exit}}(k)$ and $a_{\text{enter}}(k)$. This suggests that k independent fluctuations at $a_{\text{enter}}(k)$ could then be explained by, first, k independent fluctuation at $a_{\text{exit}}(k)$ and, second, time independence of the fluctuation outside the Hubble radius because disturbances cannot propagate over distances $>c/H$. This is basically what happens in inflationary models. The approximate time invariance of d_H during the inflationary epoch is translated into an approximate scale invariance of $\Delta_\phi(k)$ at t_{enter} . Unfortunately, the rigorous implementation of this simple idea involves advanced general relativity [126] and here we will give only a qualitative discussion.

The k independence fluctuation at $a_{\text{exit}}(k)$ is easy to arrange. In inflationary models, the density fluctuations can be associated with quantum fluctuations of the inflation field ϕ which dominates the energy density, $V(\phi)$. The fluctuations have a Hubble-volume averaged value of

$$\Delta\phi \sim \frac{2\pi}{d_H} = 2\pi H \quad H^2 = \frac{8\pi G V(\phi)}{3}, \quad (7.57)$$

which can be interpreted as an uncertainty principle relating the uncertainty in the field and the size of the region over which it is averaged. To the extent that $V(\phi)$ is constant, inflation then gives naturally a nearly scale-invariant spectrum of fluctuations at t_{exit} with

$$\frac{\Delta\rho}{\rho} \sim \frac{V' \Delta\phi}{V}. \quad (7.58)$$

Between Hubble exit and Hubble entry, the modes are frozen and we can relate the fluctuations at t_{enter} to those at t_{exit} . The question is, the fluctuations of what? It turns out [146] that it is the so-called curvature fluctuations, $\Delta_{\mathcal{R}}$, that have the simplest behavior with $\Delta_{\mathcal{R}} \sim \Delta\rho/(\rho + p)$ at t_{exit} and t_{enter} . At t_{enter} , $\Delta\rho/(\rho + p) \sim \Delta\rho/\rho$ but during inflation we have $(\rho + p) = \dot{\phi}^2$ (Eq. 4.77). The curvature fluctuations $\Delta_{\mathcal{R}}$ at Hubble entry and exit are then given by

$$\Delta_{\mathcal{R}}(a_{\text{enter}}) \propto \left[\frac{\Delta\rho}{\rho + p} \right]_{a_{\text{exit}}} \sim \left[\frac{V' H}{\dot{\phi}^2} \right]_{a_{\text{exit}}} \sim \left[\frac{H^2}{\dot{\phi}} \right]_{a_{\text{exit}}}, \quad (7.59)$$

where in the last form we use the slow-roll condition, $\dot{\phi} \sim -V'(\phi)/(3H)$. This mathematical result has an appealing “physical” explanation described in Fig. 7.10.

After Hubble entry, the curvature fluctuations will map into the more familiar potential and CMB Sachs–Wolfe temperature fluctuations with

$$\Delta_{\phi} = \frac{3}{5} \Delta_{\mathcal{R}} \quad \left(\frac{\Delta T}{T} \right)_{\text{SW}} = \frac{1}{5} \Delta_{\mathcal{R}} \quad (7.60)$$

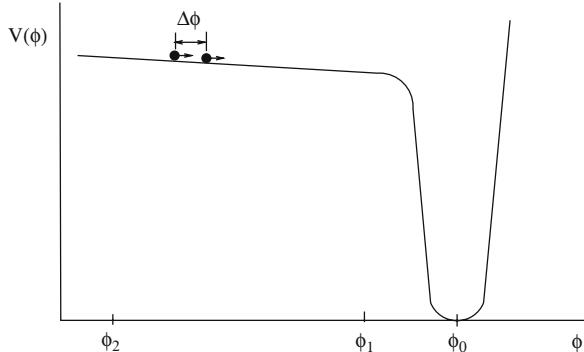


Fig. 7.10 An inflationary potential. The field slowly rolls down the potential hill until it reaches $\phi \sim \phi_1$ at which point inflation ends with the field falling into the well transforming its energy to radiation. A perturbed region with a field fluctuation $\Delta\phi$ that places it higher up on the potential will stop inflating at a time later than the mean. For a short time $\Delta t_1 \sim \Delta\phi/\dot{\phi}$ while the perturbed region is still inflating with a constant energy density, the rest of the universe is in the radiation epoch with a density $\rho \propto a^{-4}$. When the perturbed region stops inflating, it will therefore have a radiation density greater than the mean, $\Delta\rho/\rho \sim 4H_1 \Delta t_1 \sim H_1^2/\dot{\phi}$

The curvature fluctuation can be expressed directly in terms of the inflationary potential. Using $H^2 = 8\pi G V(\phi)/3$ and the slow-roll condition, (7.59) becomes

$$\Delta_{\mathcal{R}}^2(k) = \frac{V}{12\pi^2 M_{\text{pl}}^6} \left(\frac{V}{V'} \right)^2. \quad (7.61)$$

where $M_{\text{pl}} = 1/\sqrt{8\pi G}$ is the reduced Planck mass. The quantities on the r.h.s. are to be evaluating at t_{exit} . Since $V(\phi)$ is not independent of ϕ , the fluctuations depend on $t_{\text{exit}}(k)$ resulting in a scale non-invariant spectrum with n_s in (7.38) proportional to the derivative of V^3/V'^2 (Exercise 7.6):

$$n_s - 1 = M_{\text{pl}}^2 \left[3 \left(\frac{V'}{V} \right)^2 - 2 \left(\frac{V''}{V} \right) \right] \quad (7.62)$$

The WMAP measurement, $n_s = 0.960 \pm 0.013$, places constraints on the form of the inflationary potential. For example, if $V'' = 0$ the WMAP result implies

$$\frac{V'}{V} \sim \frac{0.1}{M_{\text{pl}}}. \quad (7.63)$$

The fluctuations leading to $\Delta_{\mathcal{R}}$ or Δ_ϕ are called “scalar” fluctuations. (The “s” in n_s means “scalar”, not “sound.”) In inflationary models, they are expected to be Gaussian since, like the harmonic oscillator ground state wave function, the quantum fluctuations of the inflation field are Gaussian. They are also expected to be “adiabatic” in the sense that the particle ratios, photons/baryons, wimbs/baryons, etc., are constants, independent of position. This is because different regions have different energy densities only because they stopped inflating at different epochs so the particle number ratios should be uniform. So-called “isocurvature fluctuations” can be generated in more complicated scenarios involving more than one inflationary field. In this case, fluctuations in the ratios of the two fields can lead, e.g., to fluctuations in the baryon–photon ratio.

Inflation also generates a spectrum of gravity waves leading to so-called “tensor” mode fluctuations with amplitude Δ_h determined by V :

$$\Delta_h^2 = \frac{2}{3\pi^2} \frac{V}{M_{\text{pl}}^4}. \quad (7.64)$$

The ratio between the tensor and scalar modes in inflationary models is

$$r \equiv \frac{\Delta_h^2}{\Delta_{\mathcal{R}}^2} = 8M_{\text{pl}}^2 \left(\frac{V'}{V} \right)^2 \quad (7.65)$$

The WMAP limit is $r < 0.2$, derived from the fact that CMB temperature anisotropies due to gravity waves do not have the same scale (multipole) dependence as

those due to scalar modes. We will see in Sect. 7.5 that gravity waves generate characteristic “B-mode” polarization patterns of the CMB and future experiments will aim to determine r by identifying these modes. Note that a measurement of Δ_h would give a measurement of the inflationary potential V and therefore on the energy scale of inflation $E_{\text{inf}} \sim V^{1/4}$. For the gravity waves to give a measurable effect, we need $\Delta_h > 10^{-2} \Delta_R$ which requires $E_{\text{inf}} > 10^{-3} M_{\text{pl}} \sim 10^{15} \text{ GeV}$.

The prediction of a scale-invariant-adiabatic spectrum of density fluctuations can be considered a success of inflationary theory. On the other hand, the amplitude of the fluctuations, determined by $V(\phi)$ and $V'(\phi)$, remains somewhat of a mystery since, as discussed in Sect. 5.6, inflationary models would be expected to have a steeper slope, giving larger fluctuations.

In view of this, it is natural to ask if Δ_R could be a variable determined by anthropic selection effects. In certain models, Δ_R depends on position at super-Hubble scales in which case the probability to observe a given Δ_R must include anthropic considerations in the Bayesian manner.

It is easy to show that the local conditions are strongly influenced by the value of Δ_R [147]. If Δ_R was smaller than the observed value, the velocity dispersion of virialized objects would be smaller. If this dispersion is too small, there can be no inelastic atomic collisions. Such collisions are necessary for the radiative cooling in the formation of compact objects, so it appears that the formation of stars and planets would not be possible for $\Delta_R < 10^{-6}$.

On the other hand, if Δ_R was larger than the observed value, the virialized objects would be denser. It is estimated that for $\Delta_R > 10^{-4}$, galaxies would be sufficiently dense that planetary systems would be unstable due to close stellar encounters. This would eliminate the stable conditions that led to the evolution of life on Earth. For $\Delta_R > 10^{-2}$, large numbers of black holes are formed, limiting the number of baryons available for life.

It thus appears that Δ_R and therefore the inflationary potential might be subject to anthropic selection effects.

7.4 CMB Temperature Anisotropies

The temperature of the CMB radiation is a function of the direction on the sky defined by the polar coordinates (θ, ϕ) . The description of the anisotropies of the CMB parallels that of the density fluctuations of Sect. 7.2. The expansion in plane waves (7.24) is replaced with an expansion in spherical harmonics

$$T(\theta, \phi) = \sum_{\ell, |m| \leq \ell} a_{\ell m} Y_{\ell m}(\theta, \phi) = \sum_{\ell, m} a_{\ell m} P_{\ell m}(\cos \theta) e^{im\phi}. \quad (7.66)$$

The reality of the temperature implies $a_{\ell m}^* = a_{-\ell -m}$. Just as the ϕ_k of Sect. 7.2 described density fluctuations on the scale $2\pi/k$, the $a_{\ell m}$ describe temperature variations on angular scales $\Delta\theta$ with

$$l \sim \frac{\pi}{\Delta\theta} = 300 \frac{0.01 \text{ rad}}{\Delta\theta} . \quad (7.67)$$

We note that if one considers only a small patch of the sky, a Fourier expansion is sufficient to describe the temperature (Exercise 7.2). Fourier expansions are often simpler to use than spherical harmonic expansions. It is, e.g., easy to show that for such a small patch near $\phi = \pi/2$, temperature patterns like $T(\theta, \phi) \propto \sin \pi\theta/\Delta\theta$ or $\propto \sin \pi\phi/\Delta\theta$ give $a_{\ell m}$ that are big only if $\ell \sim \pi/\Delta\theta$.

Four terms in (7.66) are special. The $\ell = 0$ term is just the mean temperature [5]

$$a_{00} = (2.725 \pm 0.002) \text{ K} \quad (7.68)$$

The $\ell = 1$ (dipole) terms are mostly generated by the Doppler effect of the Solar System's movement with respect to the mean cosmological frame (the frame where all the $a_{\ell=1,m} \sim 0$). The velocity [149] is $370 \text{ km s}^{-1} = 1.23 \times 10^{-3}c$ which gives (Exercise 7.12)

$$a_{10} = (3.358 \pm 0.02) \text{ mK} \quad a_{1\pm 1} = 0 \quad (7.69)$$

where the z axis is taken to be aligned with the solar system's movement. The dipole term has no cosmological information and we will ignore it in what follows.

The variance of the temperature, $\langle \Delta T^2 \rangle = \langle (T - \langle T \rangle)^2 \rangle$, takes a form analogous to that for the variance of the density contrast (7.28). Using the orthogonality of the $Y_{\ell m}$, one finds

$$\langle \Delta T^2 \rangle = \frac{1}{4\pi} \sum_{l>1,m} |a_{lm}|^2 = \frac{1}{4\pi} \sum_{l>1} (2\ell + 1) C_\ell \quad (7.70)$$

where C_ℓ , the analog of the power spectrum $P(k)$ (7.27), is the average over m :

$$C_\ell = \langle |a_{\ell m}|^2 \rangle_m . \quad (7.71)$$

Comparing (7.70) with (7.31) we see that the quantity $\ell(\ell + 1/2) C_\ell / 2\pi$ gives the contribution per interval of $\ln l$ to the temperature fluctuations in the same way that $\Delta(k)^2$ gives the density fluctuations per interval of $\ln k$. However, one generally uses a nearly equivalent quantity that is more closely related to the potential fluctuations (Exercise 7.5):

$$\Delta_T^2 \equiv \frac{\ell(\ell + 1) C_\ell}{2\pi} . \quad (7.72)$$

The measurements of Δ_T are shown in Fig. 7.11. The spectrum is relatively flat for $l < 20$ with a value of $\sim 10^{-5}$ of the mean temperature. We also note that there is a prominent peak near $l \sim 200$ followed by a series of smaller peaks. The spectrum

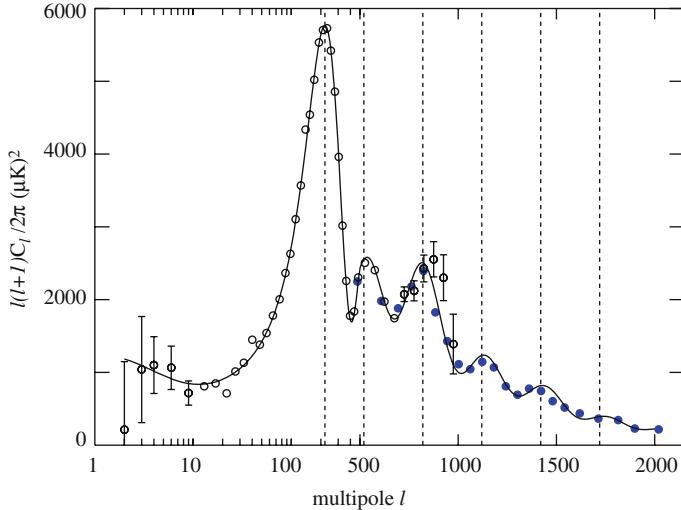


Fig. 7.11 Measurements of the spectrum of CMB anisotropies. The open circles are from WMAP satellite [150] and the filled circles are from ACBAR observations from Antarctica [151]. The curve correspond to calculations [152] using cosmological parameters derived [1] from the WMAP CMB anisotropies combined with SNIa and BAO measurements. The dotted vertical lines correspond to the approximate peak positions, $\ell_n = 300n - 80$. (Note the change of scale at $\ell = 500$.)

declines rapidly for $\ell > 1000$ suggesting that the sum (7.70) converges, unlike the integral for the density variance (7.31).

It is natural to suppose that the CMB anisotropies are due to inhomogeneities present at t_{rec} . As illustrated in Fig. 7.12, temperature anisotropies can be generated by potential wells, peculiar velocities, and intrinsic temperature variation on our “last-scattering surface,” (LastSS) defined by $\chi(z_{\text{rec}})$. Anisotropies on angular scales $\Delta\theta$ are mostly the projections of inhomogeneities on distances scales R (at t_{rec}) given by

$$\Delta\theta = \frac{R}{d_A(z_{\text{rec}})} . \quad (7.73)$$

To be more precise, a sinusoidal temperature variation on the LastSS with a wavelength (on the LastSS) $\lambda(z_{\text{rec}}) = 2\pi\hat{a}_{\text{rec}}/k$ gives $a_{\ell m}$ concentrated near

$$\ell = \frac{2\pi}{\lambda(z_{\text{rec}})/d_A(z_{\text{rec}})} = ka_0r(z_{\text{rec}}) , \quad (7.74)$$

where the angular distance is $d_A(z) = a_0r(z)/(1+z)$. The most interesting angular scale is that of the sound horizon at recombination whose angular size $\Delta\theta_s$ is shown in Fig. 5.8 as a function of $(\Omega_M, \Omega_\Lambda)$. For $(\Omega_M, \Omega_\Lambda) = (0.27, 0.73)$ we have $a_0\chi_s = 150$ Mpc and $a_0r(z_{\text{rec}}) = 14.3$ Gpc giving $\Delta\theta_s = 0.0105$. The corresponding wavelength is $\propto 2\Delta\theta_s$, i.e. the diameter of a baryonic ring surrounding a CDM

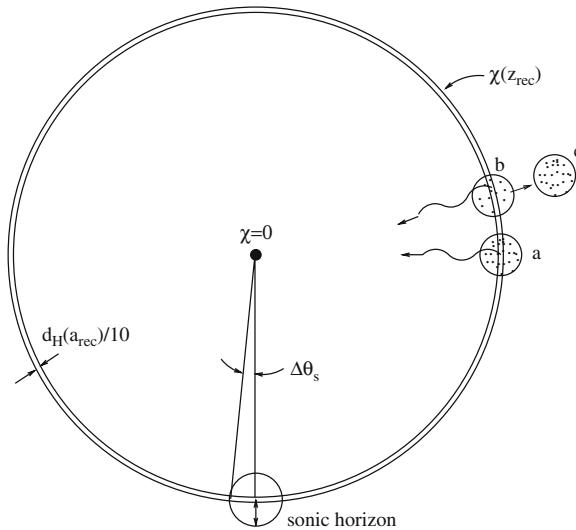


Fig. 7.12 Our “last-scattering surface” (LastSS) defining the origin of the CMB photons we observe. It is defined, on average, by $\chi(z_{\text{rec}})$ but has a thickness of about 1/10 the Hubble distance at t_{rec} . The sound horizon is shown (not to scale) and corresponds to an angle $\Delta\theta_s = 0.0105$. Density inhomogeneities on our LastSS generate temperature anisotropies. Photons from the dense region *a* will be redshifted climbing out of their potential well, but will have a higher initial temperature because of the greater compression. (The first effect dominates on large scales.) Photons coming from the under-dense region *b* will be blueshifted as they fall off their potential hill but will be redshifted since the plasma is being accelerated toward the over-dense, *c*, region behind our LastSS

concentration (Fig. 5.5). For the standard Λ CDM model we can then associate a characteristic multipole, ℓ_s , with $\Delta\theta_s$:

$$\Delta\theta_s \sim 0.0105 \text{ rad} \quad \Rightarrow \quad \ell_s \equiv \frac{\pi}{\Delta\theta_s} \sim 300. \quad (7.75)$$

We will see that on large angular scales, $\Delta\theta \gg \Delta\theta_s$, the anisotropies depend mostly on the primordial spectrum of potential fluctuations. On small angular scales, $< \Delta\theta_s$, the anisotropies depend on the nature of the acoustic oscillations of the baryon–photon plasma.

7.4.1 The Sources of Anisotropy

The calculation of the CMB anisotropies begins with a calculation of the density and potential inhomogeneities on the LastSS. The calculations involve expanding the inhomogeneities in normal modes and then solving the coupled differential equations that determine their time evolution. One must then calculate redshifts at points on the LastSS which translate the temperature inhomogeneities on the LastSS into temperature anisotropies observed by us. Publicly available computer codes [152]

can do this as a function of the cosmological parameters. In this section, we will try to give a qualitative explanation for the observed anisotropies. More quantitative treatments were reviewed in [153].

Most of the anisotropies are due to varying conditions on our LastSS. One might expect that these effects can be separated into those proportional to potential fluctuations, to peculiar velocities of the plasma, and to density fluctuations. A simple argument suggests that at large angular scales, $\ell < \ell_s$, the effects proportional to the potential fluctuations are most important. Consider a perturbed region on the LastSS of radius R . Its potential fluctuation is of order 10^{-5} . From (7.20), the density contrast will be $\sim 10^{-5}(d_H/R)^2$ and from (7.55) the peculiar velocities will be $\sim 10^{-5}(d_H/R)$. Extending these relations into the region $R > d_H$ suggests that at large scale the effects proportional to the potential fluctuations should dominate. This turns out to be true, but the constant of proportionality is surprising:

$$\left(\frac{\Delta T}{T} \right)_{\text{SW}} = \frac{\Delta \phi}{3} \quad (7.76)$$

where the SW refers to Sachs and Wolfe who first calculated this effect. The factor $(1/3)$, derived in Sect. 7.8, shows that this is not quite the familiar gravitational redshift effect.

Note that there is only a numerical factor between $d_H(t_{\text{rec}})$ and the sound horizon, so we expect the Sachs–Wolfe effect to dominate for $\ell \ll \ell_s$. Since the potential fluctuations are scale-independent, we can expect that the temperature fluctuations are scale-independent. This is what is seen in the observations of Fig. 7.11. For $3 < \ell < 30$ the Sachs–Wolfe effect dominates the anisotropies and we can “read off” the potential fluctuations from the C_ℓ (Exercise 7.5):

$$\frac{\Delta \phi}{3} \sim \frac{1}{2.725 \text{ K}} \left(\frac{\ell(\ell+1)C_\ell}{2\pi} \right)^{1/2} \sim (0.99 \pm 0.02) \times 10^{-5} \quad (7.77)$$

where the numerical value is the fitted WMAP value ($\Delta_R/5$, Table 7.1). This measurement provides the normalization of the calculation of the matter homogeneities presented in Sect. 7.2. It assumes that there are no large anisotropies generated by gravitational waves.

At small angular scales, $\Delta\theta < \Delta\theta_s$, the corresponding distance scales are $< d_H$ and we must take into account the pressure forces on the photon–baryon plasma. The photons and baryons remain tightly coupled as a perfect fluid until recombination. While the dynamical equations are a little complicated, the resulting phenomenology is simple [154, 153].

We consider one mode consisting of a spatially periodic potential whose source is primarily the CDM. As illustrated in Fig. 7.5 and Fig. 7.13, the amplitude of the baryon component at early times follows the growth of the CDM component. The baryon compression continues until the fluid pressure resists further compression. At this point the baryons start to rebound from the potential well until gravity forces the baryons back. This series of compressions and rarefactions continues until

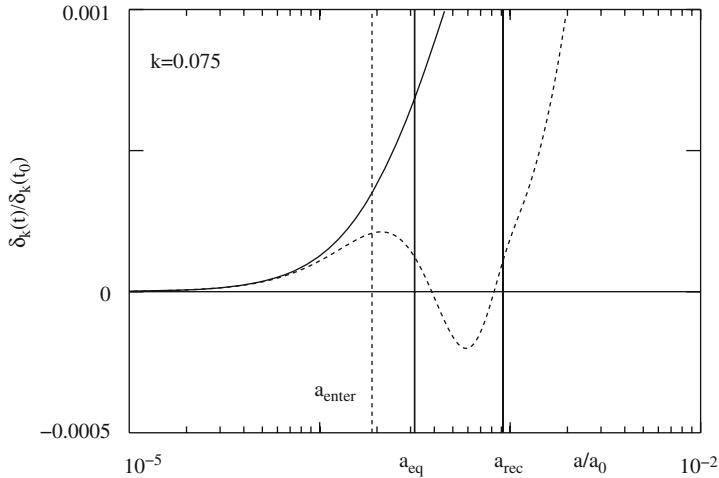


Fig. 7.13 The density contrast ($k = 0.075 h \text{Mpc}^{-1}$) for CDM (solid line) and for the baryons (dashed line) as in Fig. 7.5 except on a linear scale. Initially, the plasma contrast follows the CDM contrast until t_{enter} when the plasma starts to oscillate. At recombination, the photons are liberated and the baryons fall back into the CDM potential wells

recombination when the photons escape. With no pressure to resist collapse, the baryons then fall back into the CDM potential wells.

A time sequence of the spatial pattern of the baryon oscillations is shown in Fig. 7.14. Time is parameterized by $\Theta(t)$, the phase of the oscillations:

$$\Theta(t) = \int_0^t \omega_k(t) dt \quad (7.78)$$

with the oscillation frequency inversely proportional to the wavelength of the mode, $2\pi \hat{a}/k$:

$$\omega_k(t) = \frac{k c_s(t)}{\hat{a}(t)} . \quad (7.79)$$

The sound speed, $c_s(t)$, is given by (5.25). Compressions occur at $\Theta = n\pi$ $n = 1, 3, 5\dots$ and rarefactions occur at $\Theta = n\pi$ $n = 0, 2, 4\dots$. Midway between the compressions and rarefactions, movement of the plasma into or out of the wells leads to plasma velocities indicated by the single-tipped arrows in the Fig. 7.14.

The nature of the anisotropies created by a given oscillation mode depends on the phase of the oscillation at recombination:

$$\Theta_{\text{rec}}(k) = k a_0 \int_0^{\text{rec}} \frac{c_s(t) dt}{a(t)} = k a_0 \chi_s , \quad (7.80)$$

where $a_0 \chi_s \sim 150 h_{70}^{-1} \text{Mpc}$ is the sound horizon at recombination.

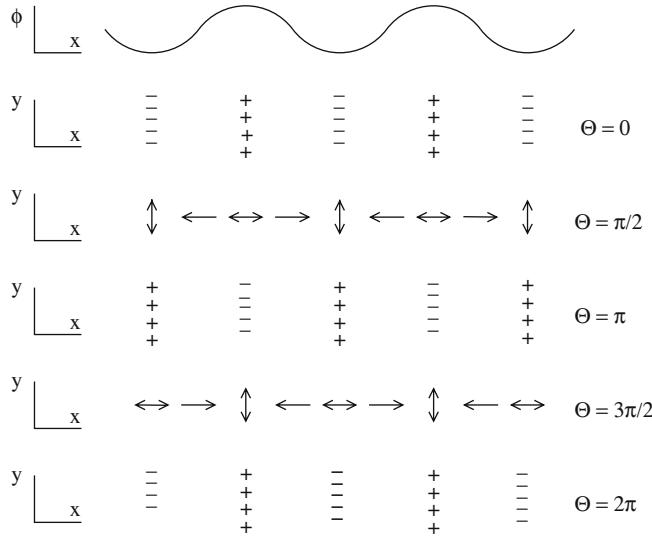


Fig. 7.14 The behavior of one oscillation mode of the baryon–photon plasma on a small patch of our LastSS. The figure shows a time sequence where the time is parameterized by the phase angle $\Theta(t)$ given by (7.78). The CDM potential is shown at the top. The + and – signs signify the hot and cold regions that would be observed if recombination occurs at the moment in question. The single-tipped arrows correspond to plasma peculiar velocities and the double-tipped arrows to CMB polarization that would be observed if recombination occurs at the moment in question

The phase $\Theta_{\text{rec}}(k)$ determines for each mode the relative importance of the gravitational redshift, the Doppler effect, and the intrinsic temperature effect. We can distinguish three extreme cases in Fig. 7.14:

- $\Theta_{\text{rec}} = n\pi$, $n = 1, 3, 5, \dots$ Compression. The observed photon temperature is *maximal* for photons coming out of the wells because of the intrinsically high temperature inside the wells (+ signs in Fig. 7.14).
- $\Theta_{\text{rec}} = n\pi$, $n = 2, 4, 6, \dots$ Rarefaction. The observed photon temperature is *minimal* for photons coming out of the wells because of the gravitational redshift (– signs in Fig. 7.14).
- $\Theta_{\text{rec}} = (n + 0.5)\pi$, $n = 1, 2, 3, \dots$ The intrinsic and gravitational redshift effects cancel. Because the peculiar velocity of the plasma is maximal at this phase, the temperature variations follow the line-of-sight plasma velocity.

While the origin of the anisotropies depends on the phase $\Theta_{\text{rec}}(k)$, it is not clear from this analysis which values of k give the largest anisotropies. It turns out that gravitational effect of the baryons on the plasma has the effect of enhancing the gravitational effects over the Doppler effects, especially for the compressional (odd n) modes. We therefore expect enhanced anisotropies for wave numbers on the LastSS of

$$k_n \sim \frac{n\pi}{a_0 \chi_s} \quad (7.81)$$

Using (7.74) we expect a series of peaks in the spectrum at the associated multipoles, $\ell \sim n\ell_s \sim 300n$. As explained below, because of the non-negligible radiation density at recombination, the position of the peaks are shifted to lower multipoles. However, the spacing of the peaks is maintained to good approximation and one observes

$$\ell_n \sim n\ell_s - \Delta\ell \quad (\ell_s, \Delta\ell) \sim (300, 80), \quad n = 1, 2, 3 \dots \quad (7.82)$$

where ℓ_s is given by (7.75). As seen in Fig. 7.11, at least five peaks are apparent in the anisotropy spectrum.

We will see in Sect. 7.5 that the large peculiar velocities of the plasma for $\Theta_{\text{rec}} = (n + 0.5)\pi$, $n = 1, 2, 3 \dots$ generate peaks in the spectrum of the CMB polarization. These peaks, shown in Fig. 7.17, have been seen by the QUAD collaboration [157].

The shift $\Delta\ell$ in (7.82) depends on the ratio of the radiation and matter densities at recombination because of the “early-time integrated Sachs–Wolf (ISW) effect” [154]. This effect is due to the decay of the gravitational potential between t_{enter} and t_{eq} . The potential perturbation for a mode of wavelength λ is proportional to $\rho(\Delta\rho/\rho)\lambda^2$. During the matter epoch, the potential is constant because $\rho \propto a^{-3}$, $\Delta\rho/\rho \propto a$, and $\lambda^2 \propto a^2$. During the radiation epoch, perturbations inside the Hubble radius oscillate ($\Delta\rho/\rho$ constant) so the potential decays. Potential decay occurs mostly during the first compression. It is timed so as to give a kick to the mode, increasing its amplitude and decreasing the time to reach the first compression. This has the effect of pushing the first peak to larger angular scales. An empirical formula is [156]:

$$\Delta\ell \sim \ell_s \times 0.267 \left(\frac{\rho_R(a_{\text{rec}})}{\rho_M(a_{\text{rec}})} \right)^{0.1} \sim 72 \frac{\ell_s}{300} \left(\frac{0.27}{\Omega_M h_{70}^2} \right)^{0.1} \quad (7.83)$$

Besides the series of peaks for $\ell > 200$, the spectrum in Fig. 7.11 exhibits two interesting effects, the late time integrated Sachs–Wolfe effect at $\ell < 5$ and damping at $\ell > 1000$. Like the early-time ISW effect, its late-time analog comes about through potential decay, this time due to the onset of vacuum energy domination. A photon that enters a potential fluctuation will leave the fluctuation later when the depth is smaller. This causes the photon energy to be boosted by the change in the potential between entry and exit. This adds temperature fluctuation at large scale (because the change in potential is proportional to the time spent in the potential). It is responsible for the rise in the C_ℓ for $\ell < 5$ seen in Fig. 7.11. It also generates a correlation between temperature and the number density of foreground galaxies. The correlation with SDSS galaxies has been seen [155]. All these measurements confirm that $\Omega_M < 1$ providing independent evidence for dark energy.

The rapid decrease of the C_ℓ for $\ell > 1000$ is due to the fact that recombination is not an instantaneous event but takes place over a non-zero time interval. Photons

observed by us therefore come from a range of positions that can be estimated by solving the Boltzmann equation for the photon-baryon plasma at recombination. It turns out that the effective “size” of a point on our LastSS is about 1/10 the Hubble distance at recombination. A temperature in a given direction is therefore an average of the temperatures over this scale. We therefore expect little anisotropy on scales of less than 1/10 the Hubble distance or about 0.1 deg. corresponding to multipoles $l > 2000$.

7.4.2 The Cosmological Parameters

Figure 7.15 shows the calculated CMB spectrum for different combinations of Ω_M , Ω_A , Ω_b . It is clear that the spectrum depends strongly on the parameters, and this explains why the parameters of the Λ CDM model are so tightly constrained. While the best values that are reported in Table 7.1 were found by a rather impersonal fit of the C_ℓ , we can nevertheless try to point out which features in the spectrum that influence most directly which cosmological parameters. Most of the following points can be confirmed by inspection of the four curves in Fig. 7.15 (Exercise 7.8).

- Δ_R determines the potential fluctuations on the LastSS giving the C_ℓ in the flat part of the Sachs–Wolfe zone, $10 < \ell < 40$.

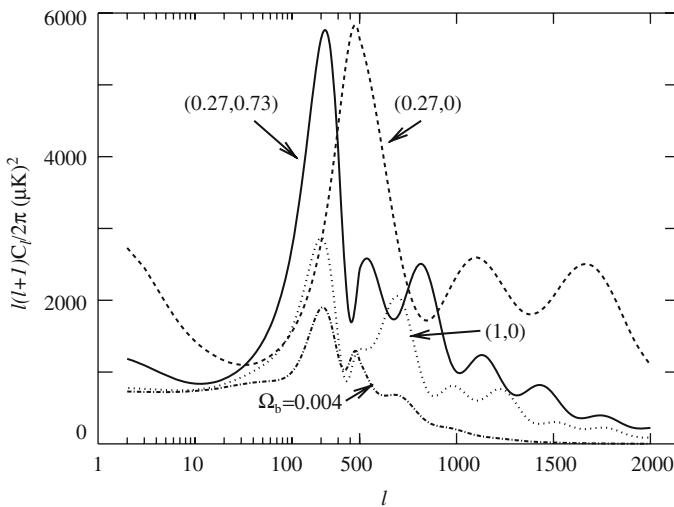


Fig. 7.15 Dependence of the predicted spectrum of CMB anisotropies on the cosmological parameters [152]. The *solid curve* uses the parameters derived from the WMAP CMB anisotropies combined with SNIa and BAO measurements. The *dashed* and *dotted curves* correspond to $(\Omega_M, \Omega_A) = (0.27, 0)$ and $(1, 0)$. The *dash-dotted curve* corresponds to $(\Omega_M, \Omega_A, \Omega_b) = (1, 0, 0.004)$

Table 7.1 The WMAP best fit cosmological parameters for a Λ CDM model [1]. All parameters were determined by assuming $\Omega_T = 1$ except for Ω_T itself. The WMAP-only value of Ω_T assumes $H_0 = (72 \pm 8) \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Hubble Key Project [9])

Parameter	WMAP only	WMAP+BAO+SN
$100\Omega_b h^2$	2.273 ± 0.062	$2.267^{+0.058}_{-0.059}$
$\Omega_{\text{CDM}} h^2$	0.1099 ± 0.0062	0.1131 ± 0.0034
Ω_Λ	0.742 ± 0.030	0.726 ± 0.015
n_s	$0.963^{+0.014}_{-0.015}$	0.960 ± 0.013
τ	0.087 ± 0.017	0.084 ± 0.016
$\Delta_R^2(k_0)$	$(2.41 \pm 0.11) \times 10^{-9}$	$(2.445 \pm 0.096) \times 10^{-9}$
$\Omega_k \equiv 1 - \Omega_T$	$-0.063 < \Omega_k < 0.017$	$-0.0179 < \Omega_k < 0.0081$

- n_s determines the change in primordial fluctuations when going to higher multipoles with $\Delta C_\ell / C_\ell \sim 0.01(\Delta \ell / \ell)(n_s - 1)/0.04$.
- $\Delta\theta_s$ determines the peak spacings. In turn $\Delta\theta_s$ is mostly a function of $(\Omega_M, \Omega_\Lambda)$ shown in Fig. 5.8. If one assumes $\Omega_T = 1$, the peak spacing constrains Ω_Λ and Ω_M .
- $\Omega_T \neq 1$ requires a late-time ISW effect at low ℓ . A modest effect is also present for $\Omega_\Lambda \neq 0$.
- $\Omega_M h^2$ determines the early-time ISW effect through (7.83). It therefore impacts on the position of the first peak and on the peak amplitudes.
- $\Omega_b h^2$ determines the baryon-radiation ratio with high ratios enhancing the compression peaks.
- τ , the optical depth to the LastSS is determined by the large-scale polarization discussed in Sect. 7.5.

Note that other cosmological parameters can be derived from those in Table 7.1. For example, H_0 can be derived from Ω_Λ and $(\Omega_{\text{CDM}} + \Omega_b)h^2 = (1 - \Omega_\Lambda)h^2$. The redshift of reionization can be derived from τ by using (5.72). Finally, we note in Table 7.2 the sound horizon parameters derived from the WMAP data.

Table 7.2 Sound horizon scales determined by WMAP [1]. The recombination redshift, z_{rec} , corresponds to the mean liberation epoch of the photons. Used as the upper limit of integration in (5.28), it gives the corresponding value of $a_0\chi_s$. For a short time after photon decoupling, the baryons continue to be dragged along by the large number of photons. The baryon decoupling redshift, z_d , used as the integration limit in (5.28) gives the slightly larger $a_0\chi_s$ used for BAO measurements

z_{rec} (photon decoupling)	1090.51 ± 0.95
$a_0\chi_s$ (photon decoupling)	$146.8 \pm 1.8 \text{ Mpc}$
ℓ_s	302.45 ± 0.86
$\Delta\theta_s = \pi/\ell_s$	0.01039 ± 0.00003
z_d (baryon decoupling)	1020.5 ± 1.6
$a_0\chi_s$ (baryon decoupling)	$153.3 \pm 2.0 \text{ Mpc}$

7.5 CMB Polarization

Important information about the conditions on our LastSS can be obtained from measurements of the CMB polarization. This polarization is a bit paradoxical because there cannot be any polarization in an ensemble of photons in thermal equilibrium because photons of both polarization states have the same energy. Therefore, if recombination took place instantaneously with a state of thermal equilibrium being translated to a state of free propagation, the polarization would vanish. Polarization can only be generated during the short time that the photons are decoupling.

Polarization can be generated by Compton scattering on our LastSS, as illustrated in Fig. 7.16. The figure shows an electron on our LastSS that happens to see a large flux of photons coming from the $\pm y$ direction and a smaller flux coming from the $\pm x$ direction. In such a case, an observer in the z direction will necessarily see a net polarization because low-energy Compton scattering at 90 deg. conserves the linear polarization of the scattered photon. It is worth emphasizing that this is a classical effect: an oscillating electro-magnetic field “shakes” an electron in the direction of the electric field, causing the electron to radiate. The re-radiated field at any point far from the electron will be polarized in the direction of the shaking as seen from that point.

It is possible for electrons on the LastSS to see such an anisotropic flux of photons during recombination because of density and velocity inhomogeneities in the surrounding plasma. It is not simple to guess the dominant effect because this is a non-equilibrium phenomenon, sensitive to the duration of recombination. It turns out that the dominant effect causing polarization is that due to gradients in plasma velocity. Consider an electron on the LastSS shown in Fig. 7.16. Suppose that the plasma at positive y has negative v_y while that at negative y has positive v_y , i.e.

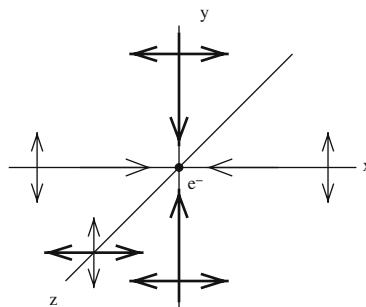


Fig. 7.16 Generation of CMB polarization. Our LastSS is the xy plane and the observer is in the z direction. The figure shows an electron at the origin that sees a photon flux coming from the $\pm x$ and $\pm y$ directions. Photons from the $\pm x$ ($\pm y$) directions must be polarized in the $\pm y$ ($\pm x$) directions to be scattered toward the observer. If the flux from the $\pm x$ directions is less than that from the $(\pm y)$ directions (as indicated in the figure), the photons scattered by the observer will have a net polarization in the $(\pm x)$ directions

that the plasma is converging toward the xz plane. In this case, the electron will see a higher photon temperature from the $\pm y$ directions (relative to the $\pm x$ directions) because of the Doppler effect. The higher temperature implies a larger photon flux. This generates a polarization along the $\pm x$ direction for an observer on the z axis.

The generation of polarization from individual modes is illustrated in Fig. 7.14. Between plasma compressions and rarefactions, the plasma converges toward or diverges away from the minima of the periodic potential. For a given mode, if recombination occurs at a time midway between compression and rarefaction, the polarization pattern (double-tipped arrows) shown in the figure is seen by the observer. We can thus expect large polarization for ℓ values roughly midway between the peaks in the temperature anisotropy spectrum. The theoretical spectrum and the QUAD data are shown in Fig. 7.17 confirming that this is indeed the case.

For modes between polarization maxima and rarefactions or compressions, there are both polarization and temperature anisotropies. This leads to peak in “TE” spectrum shown in Fig. 7.18. The C_ℓ^{TE} are defined as with the “TT” C_ℓ in (7.71) except that $|a_{\ell m}|^2$ is replaced with the product of the $a_{\ell m}$ for the temperature and the $a_{\ell m}$ for the (E-mode) polarization (Exercise 7.3, Eq. 7.150). Note also the large signal for $\ell < 5$. This is interpreted as due to scattering on electrons that were reionized at $\hat{a} \sim 0.1$ (Exercise 7.11). The polarization that WMAP has observed at $\ell < 10$ allows one to calculate the optical depth (Table 7.1) for Compton scattering on the reionization electrons and thereby estimate the epoch of reionization.

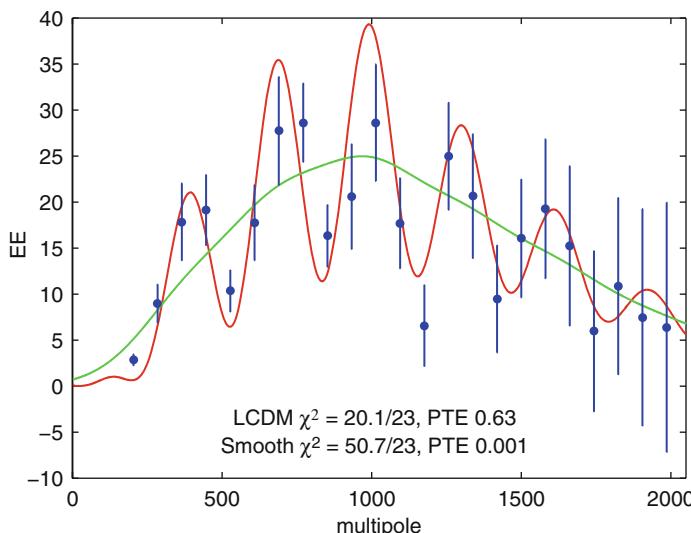


Fig. 7.17 The polarization (E-mode) spectrum from QUAD [157]. The oscillating curve shows their Λ CDM fit to the data with peaks at $\ell_n = \ell_s(n + 1/4 - p)$ with $\ell_s = 306 \pm 10$ and $p = 0.036 \pm 0.092$

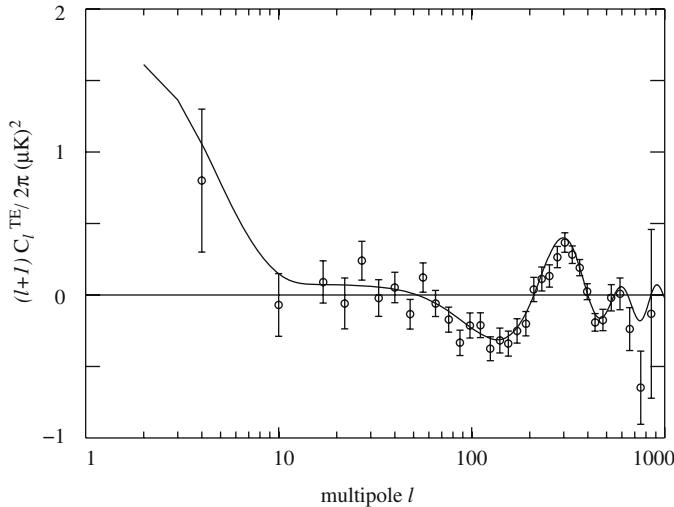


Fig. 7.18 The temperature-polarization power spectrum defined by (7.150). The *open circles* are from WMAP satellite [150]. The *curve* correspond to calculations [152] using cosmological parameters derived [1] from the WMAP CMB anisotropies combined with SNIa and BAO measurements

When one describes mathematically the polarization pattern seen on the sky [158], one finds that there are two types of polarization modes shown in Fig. 7.19. E-modes have a periodically varying polarization that is alternately perpendicular to and parallel to the polarization gradient. The polarization generated by the periodic

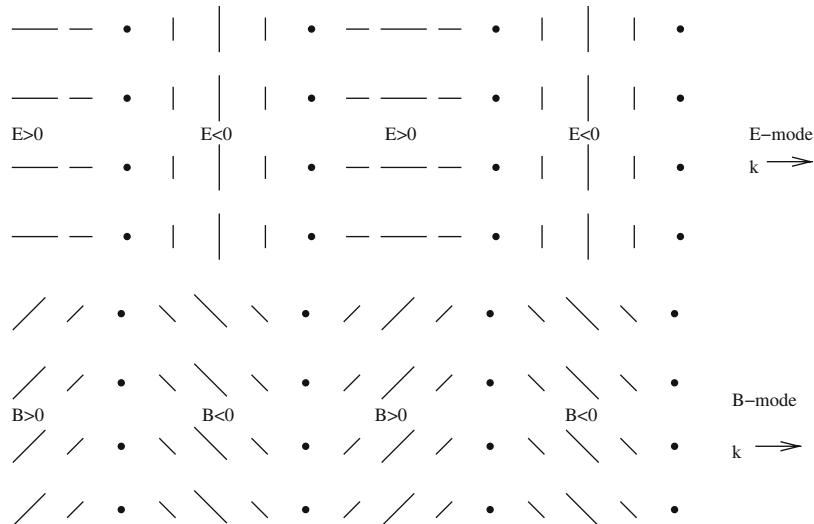


Fig. 7.19 E-mode (*top*) and B-mode (*bottom*) polarization patterns on a small patch of the LastSS. The direction and length of the lines give the size and direction of the polarization with *dots* corresponding to no polarization

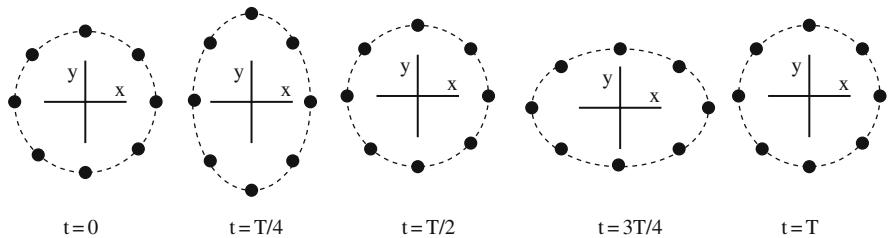


Fig. 7.20 The response of a ring of free particles to a gravitational wave. The ring is in the xy plane and the wave is incident from the z direction. The ring is periodically stretched and compressed alternately in the y and x directions. This generates relative movements of the particles in the x and y directions. This behavior corresponds to one of two possible polarizations of the wave, the second generating stretching and compressing at ± 45 deg. to the x and y axes

potential of Fig. 7.14 is pure E-mode. B-modes have polarization that is alternately at ± 45 deg. from the polarization gradient. Such a pattern cannot be generated by density gradients but they can be generated by gravitational waves on the LastSS at recombination. This is because they generate peculiar velocities that are perpendicular to rather than parallel to k . How this is done is illustrated in Figs. 7.20 and 7.21.

Primordial gravitation waves are predicted to be produced during inflation [153]. At the LastSS, these waves generate “tensor mode” temperature fluctuations. These add to the standard “scalar mode” temperature fluctuations that we have assumed up to now are the dominant cause of the temperature anisotropies. In fact, the present limits from WMAP for their contribution to the temperature anisotropy are rather weak, <20% of the total. It is hoped that future polarization measurements will

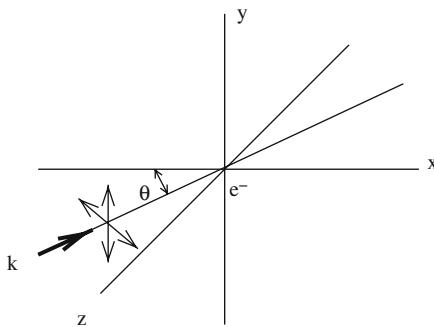


Fig. 7.21 Generation of B-mode polarization with a gravitational wave. The LastSS is in the xy plane and a gravitational wave is incident with its wave vector k in the xz plane and at an angle θ with respect to the xy plane. The wave is polarized as shown with the double-tipped arrows with one of the axes in the y direction. The disturbance on the LastSS will have a wave vector $k_{\text{LastSS}} = k \cos \theta$ and will point in the x direction. With the polarization as shown, the wave will generate LastSS peculiar velocities perpendicular to k_{LastSS} corresponding to an E-mode. If the polarization is rotated by 45 deg. about k , the generated peculiar velocities will have a component at 45 deg. with respect to k_{LastSS} corresponding to a B-mode

identify the B-modes generated by the waves, thus detecting what is perhaps a direct relic from the inflationary epoch.

Besides gravitational waves, B-modes can be generated by gravitational lensing of the E-mode polarization by foreground large-scale structure [153]. This signal will certainly dominate over the gravitational wave signal for $\ell > 100$. While this effect may appear less interesting than that due to inflationary gravity waves, it could give an important bias-free method to determine the matter inhomogeneities. Such a measurement may be useful in testing gravity at cosmological distances [26].

7.6 Alternatives to Λ CDM

The Λ CDM model with scale-invariant adiabatic primordial fluctuations describes well the CMB anisotropy data as well as other observations, e.g. those with SNIa and BAO. Before these observations became possible in the late 1990s, there were a large number of models on the market for generating the inhomogeneities in the universe. Among them were models with no CDM, using only baryonic dark matter or “hot” dark matter in the form of light neutrinos. Alternatives to adiabatic primordial fluctuations include primordial “isocurvature” fluctuations and those using “causal” fluctuations from, e.g., the decay of topological defects. In this section, we will discuss why they are all strongly disfavored by the data.

Whereas adiabatic perturbations are fluctuations in the energy density, isocurvature perturbations are fluctuations in particle number ratios, e.g. photon/baryon, photon/CDM, etc. Lacking fluctuations in the energy density, the expansion rate must be initially uniform and fluctuations cannot grow until they enter the Hubble radius. At this point inhomogeneities can begin to grow because of the pressure gradients caused by pressure differences between the components. Since the acoustic oscillations are delayed, the first compression peak comes at a smaller scale $\sim 3\ell_s/2$. While such a shift might be compensated by changing cosmological parameters, the temperature-polarization power spectrum is inconsistent with any isocurvature-dominated scenario. This is because of the negative peak at $\ell \sim 150$ which is due to the early time plasma velocities seen in Fig. 7.14. This corresponds to a super-horizon temperature-polarization correlation (Exercise 7.3) that cannot exist with purely isocurvature initial conditions. Quantitative limits on the fraction of the anisotropies due to isocurvature seeded modes are model dependent but WMAP gives upper limits of $\sim 6\%$ for certain models.

Inflationary models plant seeds of the inhomogeneities in the distant past. They thus appear to generate inhomogeneities outside of our (classical) horizon ($\sim d_H$). It is also possible to seed fluctuations within the classical horizon in an apparently causal manner. One class of models [159] does this by the decay of “topological defects.” Such models appears to be ruled out because they cannot normally create the observed acoustic peaks in the anisotropy spectrum. This is because decays occur randomly in time so there is no temporal coherence between different modes. In inflationary models, the initial conditions are established in the very early universe. Each \mathbf{k} has, in general, a growing mode and a decaying mode (Sect. 7.7) but

the decaying modes have long since vanished by t_{enter} . Modes of neighboring $|\mathbf{k}|$ are therefore oscillating with a common phase at t_{rec} so acoustic peaks can be generated.

We now turn to models that do away with CDM. Because of the failure for the moment, to identify any other dark matter candidate, models with only baryonic dark matter have a natural appeal. However, apart from the lack of a convincing idea on how to hide the baryons (Sect. 2.5), such models have always had problems with structure formation. This is due to the fact that perturbations of the baryon density continue to oscillate as acoustic waves until t_{rec} compared to CDM which can gravitate freely after t_{eq} . This gives CDM more time to form structures than baryonic dark matter. At galactic scales, the problem is even worse because of “Silk damping.” At the time of recombination, as the photon mean free path becomes large, the photons stream out of over-dense regions. Because of residual Compton scattering, they push the electrons along with them, who in turn pull the baryons by coulomb interactions. This destroys fluctuations on galactic scales. In models with CDM, the baryon fluctuations on galactic scales are initially erased by Silk damping but the baryons then fall back into the CDM potential wells.

The problem of Silk damping can be avoided by supposing that the primordial fluctuations are isocurvature fluctuations of the photon–baryon ratio. In this case, the regions of baryon excess do not correspond to regions of photon excess so Silk damping is less important. This possibility appears to be ruled out by the WMAP limits on isocurvature perturbations.

Finally, we consider “hot dark matter” in the form of neutrinos. A neutrino with a mass $\sim 10 \text{ eV}$ would seem to be a natural dark-matter candidate. Experiments (Sect. 1.1.5) do not encourage the belief that neutrinos have masses this large but the possibility is not completely ruled out. Like CDM particles, neutrinos do not participate in the acoustic oscillations of the photon-charged particle plasma, having decoupled from the other particles at $T \sim \text{MeV}$. They can therefore start to freely create gravitational structures as soon as they dominate the energy density at t_{eq} without being bothered by the presence of the relativistic photon–baryon plasma.

The problem with neutrinos is that they are relativistic between decoupling $T \sim \text{MeV}$ and $T \sim m_V$. During this time, small-scale density fluctuations of neutrinos are destroyed as they freely propagate from regions of large density to regions of small density. This phenomenon is called “free-streaming.” Because neutrinos were relativistic when they decoupled, neutrino dark matter is called hot dark matter. CDM particles were, by definition, non-relativistic when they decoupled so free-streaming is unimportant.

We can calculate the scale λ_{fs} over which the fluctuations are destroyed by calculating the distance a neutrino will travel between $T \sim \text{MeV}$ ($t = t_f$) and $T \sim m_V$ ($t = t_V$). The coordinate distance is

$$\chi_{\text{fs}} = \int_{t_f}^{t_V} \frac{v(t)dt}{a(t)} , \quad (7.84)$$

where $v(t)$ is the neutrino velocity. In the approximation $v(t) \sim 1$, this is of order the Hubble distance at t_V :

$$a(t_V)\chi_{\text{fs}} \sim \frac{H_0^{-1}}{\sqrt{1.68\Omega_V}} \hat{a}_V^2 , \quad (7.85)$$

where $\hat{a}_V \sim T_\gamma(t_0)/m_V$. This corresponds to a present-day distance $a_0\chi_{\text{fs}}$ of

$$\lambda_{\text{fs}} \sim 60 \text{ Mpc} \frac{10 \text{ eV}}{m_V} . \quad (7.86)$$

Fluctuations on scales smaller than this are erased by free-streaming. For larger scales, $\lambda_k > \lambda_{\text{fs}}$, the fluctuation spectrum is basically the same as that for CDM.

All the observed virialized structures correspond to scales that are smaller than λ_{fs} for neutrino masses in the eV range. In HDM models, the only way to form such structures is by fragmentation of larger scale structures.³ This generally leads to rather late galaxy formation, around $z \sim 1$, which is difficult to reconcile with the known existence of galaxies and quasars at $z \sim 5$. Furthermore, the early formation of galaxy cluster leads to hotter and more massive clusters than those observed. We can conclude that models with only HDM are ruled out for many reasons.

The fact that neutrinos cannot make up most of the dark matter implies

$$\Omega_V \ll 0.27 \Rightarrow \sum m_V \ll 10 \text{ eV} , \quad (7.87)$$

where the sum is over the three neutrino species. WMAP has reported a more strict limit by using the fact that the position of the peaks in the temperature anisotropy spectrum depends on the shift (7.83) due to the early-time integrated Sachs–Wolfe effect. Neutrinos that were not relativistic at recombination would modify the radiation energy density and therefore the peak positions. WMAP reports the limit [1]:

$$\sum m_V < 0.67 \text{ eV (95\% CL)} . \quad (7.88)$$

7.7 Newtonian Evolution

In this section, we will perform the standard Jeans analysis for a non-relativistic fluid in an expanding universe. This will allow us to derive the growth rates for the $\delta_{\mathbf{k}}(t)$ that we supposed in Sect. 7.2.

In laboratory coordinates, the three equations that determine the behavior of a gravitating perfect fluid are the conservation of mass

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 , \quad (7.89)$$

³ Hot dark matter models are therefore called “top-down” models as opposed to “bottom-up” CDM models where the smallest structures form first.

the gravitational and hydrodynamic force on a volume element

$$\rho \left[\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right] = -\nabla p - \rho \nabla \Phi , \quad (7.90)$$

and the Poisson equation for the gravitational potential

$$\nabla^2 \Phi = 4\pi G \rho . \quad (7.91)$$

The solution of these three equations that corresponds to a matter-dominated homogeneous universe is:

$$\rho = \rho(t_0) \hat{a}^{-3} \quad \mathbf{v} = \frac{\dot{a}}{a} \mathbf{R} \quad \Phi = \frac{4\pi G \rho}{6} |\mathbf{R}|^2 , \quad (7.92)$$

where $a(t)$ satisfies the Friedmann equation and \mathbf{R} and \mathbf{v} are laboratory coordinates and velocities. The fact that the velocity and potential become of order unity when the distance from the origin becomes of order $d_H \sim 1/\sqrt{G\rho}$ means that this non-relativistic formalism can only treat perturbations on scales much less than d_H .

To the above solution, we now add a small density perturbation in the form of a wave of wavelength proportional to the scale factor:

$$\rho = \rho(t_0) \hat{a}^{-3} \left[1 + \delta_{\mathbf{k}}(t) \exp\left(\frac{-i \mathbf{k} \cdot \mathbf{R}}{\hat{a}(t)}\right) \right] , \quad (7.93)$$

$$\lambda_k(t) = \frac{2\pi}{k} \hat{a}(t) \dots \quad (7.94)$$

The wave vector \mathbf{k} is fixed. The perturbation describes the density contrast on the scale of $1/k$ (today). A perturbation in the density generally perturbs the pressure:

$$p = \bar{p}(t) + c_s^2 \rho(t_0) \hat{a}^{-3} \delta_{\mathbf{k}}(t) \exp\left(\frac{-i \mathbf{k} \cdot \mathbf{R}}{\hat{a}(t)}\right) \quad (7.95)$$

where we have used the definition (5.25) of the sound speed as the derivative of the pressure with respect to the energy density for adiabatic variations. The perturbations in the velocity field and potentials are defined by

$$\mathbf{v} = \frac{\dot{a}}{a} \mathbf{R} + \mathbf{v}_{\mathbf{k}}(t) \exp\left(\frac{-i \mathbf{k} \cdot \mathbf{R}}{\hat{a}(t)}\right) , \quad (7.96)$$

and

$$\Phi = \frac{4\pi G \rho}{6} r^2 + \Phi_{\mathbf{k}}(t) \exp\left(\frac{-i \mathbf{k} \cdot \mathbf{R}}{\hat{a}(t)}\right) . \quad (7.97)$$

Substituting ρ , p , \mathbf{v} , and Φ into (7.89), (7.90), and (7.91), we find

$$\dot{\delta}_{\mathbf{k}} - \frac{i\mathbf{k} \cdot \mathbf{v}_{\mathbf{k}}}{\hat{a}(t)} = 0 , \quad (7.98)$$

$$\frac{d(\hat{a}\mathbf{v}_{\mathbf{k}})}{dt} - i\mathbf{k}(c_s^2\delta_{\mathbf{k}} + \Phi_{\mathbf{k}}) = 0 , \quad (7.99)$$

and

$$\Phi_{\mathbf{k}} = -\frac{4\pi G\bar{\rho}}{(k/\hat{a})^2} \delta_{\mathbf{k}} . \quad (7.100)$$

If $\mathbf{v}_{\mathbf{k}} \perp \mathbf{k}$, it follows from (7.99) that $d(\hat{a}\mathbf{v}_{\mathbf{k}})/dt = 0$ implying that the amplitude of the perturbation decreases with time, $\mathbf{v}_{\mathbf{k}} \propto 1/a$. We can therefore generally ignore these “transverse” modes and suppose $\mathbf{v}_{\mathbf{k}} \parallel \mathbf{k}$. In this case, (7.98) becomes

$$\dot{\delta}_{\mathbf{k}} = -i(k/\hat{a})v_{\mathbf{k}} . \quad (7.101)$$

Substituting (7.101) and (7.100) into (7.99), we find a second-order differential equation for $\delta_{\mathbf{k}}$:

$$\ddot{\delta}_{\mathbf{k}} + 2\frac{\dot{a}}{a}\dot{\delta}_{\mathbf{k}} + (c_s^2(k/\hat{a})^2 - 4\pi G\bar{\rho})\delta_{\mathbf{k}} = 0$$

or, equivalently,

$$\ddot{\delta}_{\mathbf{k}} + 2\frac{\dot{a}}{a}\dot{\delta}_{\mathbf{k}} + c_s^2[(k/\hat{a})^2 - k_j^2]\delta_{\mathbf{k}} = 0 , \quad (7.102)$$

where $k_j = 2\pi/\lambda_j$ with the “Jeans wavelength” given by

$$\lambda_j = 2\pi \frac{c_s}{\sqrt{4\pi G\bar{\rho}}} = \frac{2\pi}{\sqrt{3/2}} c_s t_H , \quad (7.103)$$

The solutions of (7.102) depend on the value of the wavenumber k/\hat{a} compared to the Jeans wavenumber k_j . For short wavelengths, $\lambda_k \hat{a} \ll \lambda_j$, (7.102) takes the form

$$\ddot{\delta}_{\mathbf{k}} + 2\frac{\dot{a}}{a}\dot{\delta}_{\mathbf{k}} + c_s^2(k/\hat{a})^2\delta_{\mathbf{k}} = 0 \quad c_s k/\hat{a} \gg 4\pi G\bar{\rho} \sim \dot{a}/a . \quad (7.104)$$

This is the equation of an acoustic wave that is damped by the expansion of the universe. For $\dot{a}/a \rightarrow 0$, the solution is

$$\delta_{\mathbf{k}} \sim \exp(\pm i\Theta) \quad \Theta(t) = k \int^t \frac{c_s dt}{\hat{a}} . \quad (7.105)$$

For long wavelengths, $\lambda_k \hat{a} \gg \lambda_j$, (7.102) takes the form

$$\ddot{\delta}_{\mathbf{k}} + 2\frac{\dot{a}}{a}\dot{\delta}_{\mathbf{k}} - 4\pi G\bar{\rho}\delta_{\mathbf{k}} = 0 \quad c_s k/\hat{a} \ll 4\pi G\bar{\rho}. \quad (7.106)$$

As usual, the situation is clarified by using $\hat{a}(t)$ as the independent variable. In the case $\Omega_M = \Omega_T = \Omega$, we find

$$\begin{aligned} [\Omega + (1 - \Omega)\hat{a}] \hat{a}^2 \frac{d^2\delta_{\mathbf{k}}}{d\hat{a}^2} + \\ [3\Omega/2 + 2(1 - \Omega)\hat{a}] \hat{a} \frac{d\delta_{\mathbf{k}}}{d\hat{a}} - [3\Omega/2] \delta_{\mathbf{k}} = 0. \end{aligned} \quad (7.107)$$

There are two solutions corresponding to growing and decaying modes. For $\Omega = 1$, the solutions are

$$\delta_{\mathbf{k}} \propto a, \quad \delta_{\mathbf{k}} \propto a^{-3/2} \quad (\Omega_M = \Omega_T = 1). \quad (7.108)$$

For $\Omega = 0$ the solutions are

$$\delta_{\mathbf{k}} \propto a^0, \quad \delta_{\mathbf{k}} \propto a^{-1} \quad (\Omega_M = \Omega_T = 0). \quad (7.109)$$

The growing modes confirm the large time behavior found with the spherical collapse model, e.g. $\Delta\rho/\rho \propto a$ for $\Omega_M = \Omega_T = 1$.

Finally, we note the solution of (7.106) in the limit $\dot{a}/a \rightarrow 0$:

$$\delta_{\mathbf{k}} \propto \exp(\pm t/\tau) \quad \tau^{-2} = 4\pi G\bar{\rho} \quad \left(\frac{\dot{a}}{a} = 0 \right). \quad (7.110)$$

We see that the gravitational collapse of the fluid in this case is exponential. In an expanding universe, the collapse follows only a power law because the expansion works against collapse.

7.8 Photon Propagation

The primary source of information on homogeneities in the universe is the spectrum of CMB anisotropies. In order to understand the anisotropies, it is first necessary to understand the propagation of photons in an inhomogeneous universe. To simplify the problem, we consider only the spherically symmetric universe introduced in Sect. 7.1 with a small over-dense region at the center of a critical universe. The primary result will be the demonstration of the Sachs–Wolfe formula, $\Delta T/T = \Delta\phi/3$.

We consider a zero-pressure universe with a radially dependent density:

$$\rho = \rho(r, t) \quad p = 0. \quad (7.111)$$

For the coordinate system, the isotropy allows us to use the angular coordinates (θ, ϕ) in the same way as in a homogeneous universe. If the density inhomogeneity is small, we can also use co-moving clocks with fixed spatial coordinates to measure the time coordinate. Such clocks are not useful as soon as bound structures are formed since clock trajectories will eventually cross making the coordinates extremely complicated.

With an arbitrary radial coordinate, r , the isotropy and the use of co-moving clocks implies that the metric must take the following form:

$$d\tau^2 = dt^2 - A(r, t) dr^2 - B(r, t) (d\theta^2 + \sin^2 \theta d\phi^2). \quad (7.112)$$

From this metric, the Einstein equation can be written [160] and we invite the reader to do this in Exercise 7.13. The solutions to the equations are not unique because we still have the liberty to choose the radial coordinate. Nevertheless, all solutions will satisfy the conservation of energy:

$$\frac{\partial \rho \sqrt{AB}}{\partial t} = 0. \quad (7.113)$$

This relation guarantees the conservation of the number of particles in a fixed co-moving volume, $dV = \sqrt{A} dr \sqrt{B} d\theta \sqrt{B} \sin \theta d\phi$.

It is possible (Exercise 7.13) to choose the radial coordinate r so that the metric takes a form that is very similar to the Robertson–Walker metric:

$$d\tau^2 = dt^2 - b(r, t)^2 \left[\frac{dr^2}{1 - \kappa(r)r^2} \left(1 + \frac{r}{b} \frac{\partial b}{\partial r} \right)^2 - r^2 (d\theta^2 + \sin^2 \theta d\phi^2) \right]. \quad (7.114)$$

With this metric, the Einstein equations are such that the “scale function” $b(r, t)$ satisfies a “Friedmannesque” equation:

$$\left(\frac{\dot{b}}{b} \right)^2 + \frac{\kappa(r)}{b^2} = \frac{8\pi G}{3} \rho \left(1 + \frac{r}{b} \frac{\partial b}{\partial r} \right), \quad (7.115)$$

and the conservation of energy (7.113) takes the form

$$\rho(r, t) = \frac{\text{const.}}{b^3 (1 + (r/b)(\partial b / \partial r))}. \quad (7.116)$$

Like the curvature parameter k of the Robertson–Walker metric, the “curvature function” $\kappa(r)$ is determined by the conditions at an arbitrary time which we will call t_1 :

$$\frac{\kappa(r)}{b(r, t_1)^2} = - \left(\frac{\dot{b}}{b} \right)_{t_1}^2 + \frac{8\pi G}{3} \rho(r, t_1) \left(1 + \frac{r}{b} \frac{\partial b}{\partial r} \right)_{t_1}. \quad (7.117)$$

We will take as initial conditions

$$\rho(r < r_c, t_1) = \bar{\rho}(t_1)(1 + \delta_1) \quad |\delta_1| \ll 1 \quad (7.118)$$

$$\rho(r > r_c, t_1) = \bar{\rho}(t_1), \quad (7.119)$$

where r_c is the radius of the perturbed region. The parameters of the metric (7.114) must satisfy the Friedmannesque equation (7.115) and the conservation of energy (7.116). For a constant density, (7.116) becomes

$$b^3 \left(1 + \frac{r}{b} \frac{\partial b}{\partial r} \right) = \text{const.}, \quad (7.120)$$

with the general solution

$$b(r, t) = \left(\alpha(t) + \frac{\beta(t)}{r^3} \right)^{1/3}. \quad (7.121)$$

It is straightforward to verify that the solutions (to first order in δ_1) of (7.115) and (7.116) that satisfy (7.118), (7.119), and (7.121) are, for $r > r_c$

$$\rho(r > r_c, t) = \bar{\rho}(t) = \bar{\rho}(t_1) \tilde{a}^{-3}, \quad \tilde{a}(t) \equiv \frac{a(t)}{a_1}, \quad (7.122)$$

$$b(r > r_c, t) = a(t) \left[1 - \tilde{a} \frac{r_c^3}{r^3} (\delta_1/3) \right], \quad (7.123)$$

$$\kappa(r > r_c) r^2 = \frac{r_c^2 a^2}{d_H^2} \frac{r_c}{r} (5\delta_1/3), \quad (7.124)$$

and for $r < r_c$

$$\rho(r < r_c, t) = \bar{\rho}(t)[1 + \tilde{a}\delta_1], \quad (7.125)$$

$$b(r < r_c, t) = a(t)[1 - \tilde{a}(\delta_1/3)], \quad (7.126)$$

$$\kappa(r < r_c) r^2 = \frac{r^2 a^2}{d_H^2} (5\delta_1/3). \quad (7.127)$$

As expected, the perturbations to the density, to the scale factor, and to the curvature parameter are all proportional to the initial density perturbation δ_1 . Also as expected, the perturbations to the density and scale factor grow linearly with the scale factor $\tilde{a} = a/a_1$. We see that κ is negligible if the size $r_c a$ of the over-dense region is negligible compared to d_H .

Neglecting the curvature, the radial metric for $r > r_c$ is

$$dt^2 = dt^2 - a(t)^2 dr^2 (1 - h(r, t)), \quad (7.128)$$

where

$$h(r, t) = (4/3) \tilde{a} \frac{r_c^3}{r^3} \delta_1 = \frac{M(t)}{\pi \bar{\rho} a^3 r^3}, \quad (7.129)$$

where the mass excess in the region $r < r_c$ is

$$M(t) = \frac{4\pi}{3} \bar{\rho} a^3 r_c^3 \tilde{a} \delta_1. \quad (7.130)$$

We can now calculate the propagation of photons using the metric (7.128) in the same way that we calculated the propagation of photons with the Robertson–Walker metric in Chap. 3. Inverting the spatial configuration in Chap. 3, we consider a photon with wavelength λ_1 emitted at $r_1 > r_c$ and detected with wavelength λ_0 at $r_0 \gg r_1$. Treating the photon as a classical wave, we impose $d\tau = 0$ for each wave crest. The first crest is emitted at t_1 and detected at t_0 , implying

$$\int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_{r_1}^{r_0} [1 - (1/2)h(r, t(r))] dr. \quad (7.131)$$

The second crest is emitted at $t_1 + \lambda_1$ and detected at $t_0 + \lambda_0$:

$$\int_{t_1 + \lambda_1}^{t_0 + \lambda_0} \frac{dt}{a(t)} = \int_{r_1}^{r_0} \left[1 - \frac{h(r, t(r))}{2} - \frac{\partial h}{\partial t} \lambda_1 \frac{a(t(r))}{2a_1} \right] dr. \quad (7.132)$$

In this expression, we have used the fact that to a first approximation the second crest follows the first by a time $\lambda_1 a(t)/a_1$, i.e. the time delay for a homogeneous universe.

Subtracting (7.131) and (7.132), we find the redshift:

$$z + 1 = \frac{\lambda_0}{\lambda_1} = \frac{a_0}{a_1} \left[1 - (1/2) \int_{r_1}^{r_0} \frac{\partial h}{\partial t} a(t(r)) dr \right]. \quad (7.133)$$

In this expression we see the redshift for a homogeneous universe, a_0/a_1 , and the first-order correction

$$\frac{\Delta \lambda_0}{\lambda_0} = -(1/2) \int_{r_1}^{r_0} \frac{\partial h}{\partial t} a(t(r)) dr = \frac{M(t_1)}{2\pi \bar{\rho} a^3 a_1} \int_{r_1}^{r_0} \frac{a \dot{a} dr}{r^3}, \quad (7.134)$$

where we have used (7.129).

To evaluate the correction, we can use the unperturbed ($\delta_1 = 0$) trajectory:

$$a(t) = a_1 \left(\frac{t}{t_1} \right)^{2/3} \Rightarrow t(r)^{1/3} = t_1^{1/3} + \frac{a_1}{3t_1^{2/3}} (r - r_1). \quad (7.135)$$

Substituting this in (7.134), we find

$$\frac{\Delta\lambda_0}{\lambda_0} = (1/3) \frac{GM_1}{a_1 r_1} + \frac{GM_1}{(a_1 r_1)^2} t_1 . \quad (7.136)$$

The two terms in this expression have simple interpretations. The first is called the Sachs–Wolfe term and is proportional to the gravitational potential of the emission point

$$(1/3) \frac{GM_1}{a_1 r_1} = \Phi_1/3 . \quad (7.137)$$

The factor of 1/3 is perhaps surprising in view of its absence in the case of a static gravitational field (Fig. 3.16). We will return to this later.

The second term in (7.136) has a simple interpretation as simply the peculiar velocity of the source projected in the direction of the observer:

$$\frac{GM_1}{(a_1 r_1)^2} t_1 = \mathbf{v}_1 \cdot \hat{\mathbf{n}} . \quad (7.138)$$

This term is therefore called the “Doppler” term.

In a general treatment, it is not surprising that one finds

$$\frac{\Delta\lambda_0}{\lambda_0} = (1/3)(\Phi_1 - \Phi_0) + \hat{\mathbf{n}} \cdot (\mathbf{v}_1 - \mathbf{v}_0) , \quad (7.139)$$

where the gravitational potential at the point i is

$$\Phi_i = G\bar{\rho}a^3 \int \frac{\delta(\mathbf{R}, t)d^3r}{a|\mathbf{R} - \mathbf{R}_i|} , \quad (7.140)$$

and where the peculiar velocities are determined by

$$\nabla \cdot \mathbf{v}_p = -\dot{\delta}(\mathbf{R}, t) \sim -H \Omega_M^{0.6} \delta(\mathbf{R}, t) . \quad (7.141)$$

Returning to the factor (1/3) in the SW term, we note that some people prefer to use a “Newtonian” coordinate system [161] where the SW term is the sum of two terms:

$$\Phi/3 = \Phi - (2/3)\Phi . \quad (7.142)$$

The first term $= \Phi_1$ is the usual gravitational redshift in a static gravitational field. The second term $= -(2/3)\Phi_1$ is due to the fact that in these coordinates, the time of emission differs from the mean Robertson–Walker time by a position dependent factor $\Delta t/t = \Phi$. The normal cosmological redshift a_0/a_1 is thus modified by a factor $-da/a = -(2/3)dt/t$ where we have assumed $a \propto t^{2/3}$. The sum of these two effects then gives the SW effect which is, of course, the physically meaningful result.

We also note that in Newtonian coordinates, the density contrast is dominated by the fact that in the dense region, the universe at time t is “younger” by a factor $\Delta t/t = \Phi$ and therefore denser by a factor

$$\delta \sim -2\Phi \quad \text{Newtonian coordinates.} \quad (7.143)$$

The use of Newtonian coordinates has become popular in discussion of the CMB anisotropies [154].

Exercises

7.1 Using (7.24), calculate the mass in a sphere of comoving radius r and centered at \mathbf{r}' . Square the mass and subtract the mean squared mass and then average over \mathbf{r}' to verify the formula for σ_r (7.32).

7.2 Consider a small patch of the sky of solid angle $\Delta\Omega = (\Delta\theta)^2 \ll 1$. The patch is covered by “flat sky” coordinates (θ_x, θ_y) with $0 < \theta_x < \Delta\theta$ and $0 < \theta_y < \Delta\theta$. Consider a Fourier expansion of the temperature inside the patch:

$$T(\boldsymbol{\theta}) = \frac{1}{\Delta\theta} \sum_{\mathbf{k}} a_{\mathbf{k}} e^{i\mathbf{k}\cdot\boldsymbol{\theta}} \quad (7.144)$$

with the $\mathbf{k} = 2\pi\mathbf{n}/\Delta\theta$ with $\mathbf{n} = (n_x, n_y)$, integer n_x, n_y .

(a) Show that the variance of the temperature is given by

$$\langle (T - \langle T \rangle)^2 \rangle = \frac{1}{\Delta\Omega} \sum_{\mathbf{k}} |a_{\mathbf{k}}|^2 \sim \frac{1}{2\pi} \int_0^\infty k \, dk \langle |a_{\mathbf{k}}|^2 \rangle , \quad (7.145)$$

where the average in the last form is over directions of \mathbf{k} .

The temperature correlation function is

$$\xi(\theta) \equiv \langle (T(\boldsymbol{\theta}') - \bar{T})(T(\boldsymbol{\theta}' - \boldsymbol{\theta}) - \bar{T}) \rangle \quad (7.146)$$

where the average is over $\boldsymbol{\theta}'$ and orientations of $\boldsymbol{\theta}$.

(b) Show that for $\theta \ll 1$, the correlation function is given by

$$\xi(\theta) = \frac{1}{2\pi} \int_0^\infty k \, dk J_0(k\theta) \langle |a_{\mathbf{k}}|^2 \rangle , \quad (7.147)$$

where the Bessel function is

$$J_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{ix \cos \phi} d\phi . \quad (7.148)$$

$J_0(x)$ resembles the function $\sin x/x$ with $J_0(0) = 1$, a zero at $x = 2.4$, a minimum at $x = 3.8$, a zero at $x = 5.5$, etc.

- (c) Draw qualitatively $J_0(k\theta)$ and $C_\ell(\ell = k)$ for $\theta = \Delta\theta_s/2$, $\theta = \Delta\theta_s$, and $\theta = 5\Delta\theta_s$. Over what angular range is the CMB temperature strongly correlated?

7.3 The E-mode polarization on a small patch of the sky can be expanded as with (7.144):

$$E(\boldsymbol{\theta}) = \frac{1}{\Delta\theta} \sum_{\mathbf{k}} a_{\mathbf{k}}^E e^{i\mathbf{k}\cdot\boldsymbol{\theta}} \quad (7.149)$$

Note that $E > 0$ ($E < 0$) corresponds to polarization parallel to (perpendicular to) \mathbf{k} (see Fig. 7.19).

The C_ℓ^{TE} in Fig. 7.18 are then defined by

$$C_\ell^{TE}(\ell \sim k) \equiv \langle a_{\mathbf{k}}^E a_{\mathbf{k}}^T \rangle \quad (7.150)$$

where the $a_{\mathbf{k}}^T$ are simply the $a_{\mathbf{k}}$ in the temperature expansion (7.144) and the average is over orientations of \mathbf{k} .

- (a) Show that the TE correlation function is

$$\xi(\theta) \equiv \langle E(\boldsymbol{\theta}') T(\boldsymbol{\theta}' - \boldsymbol{\theta}) \rangle \sim \frac{1}{2\pi} \int_0^\infty k dk J_0(k\theta) C_\ell^{TE}(k = \ell) \quad (7.151)$$

where the average is over $\boldsymbol{\theta}'$ and orientations of $\boldsymbol{\theta}$.

- (b) Show that the observed TE spectrum (Fig. 7.18) implies that the temperature and polarization are correlated on super-horizon distances ($\theta > \Delta\theta_s$). This is characteristic of adiabatic perturbations. What must be the form of the TE spectrum to avoid such a correlation?

7.4 Consider a universe that is completely homogeneous except for a single initial adiabatic perturbation that happens to lie at $\theta = 0$ on our LastSS. According to Fig. 5.5, at recombination there will be a CDM cold spot on the LastSS surrounded by a hot ring at a distance $a_{\text{rec}}\chi_s$ from the cold spot.

- (a) Show that the symmetry about $\theta = 0$ means that the $a_{\mathbf{k}}$ in (7.144) are given by

$$a_{\mathbf{k}} \sim \frac{2\pi}{\Delta\theta} \int_0^\infty \theta d\theta T(\theta) J_0(k\theta) \quad (7.152)$$

where the Bessel function J_0 is given by (7.148).

- (b) By drawing qualitatively $T(\theta)$ and $J_0(k\theta)$ estimate the position of the first peak in the anisotropy spectrum.
- (c) If the initial perturbations were isocurvature with a CDM deficit replacing the CDM excess, then the central spot would be hot (potential hill) rather than cold. Estimate the position of the first acoustic peak.

7.5 Consider an observer at \mathbf{r}' who sees Sachs–Wolfe temperature fluctuations given by

$$T(\theta, \phi) = \sum_{\ell m} a_{\ell m}(\mathbf{r}') Y_{\ell m}(\theta, \phi) = \frac{\phi(\mathbf{r})}{3} = \frac{1}{3} \sum_{\mathbf{k}} \phi_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \quad (7.153)$$

where \mathbf{r} is on the observers LastSS and (θ, ϕ) are angles seen by the observer.

- (a) Find an expression for $a_{\ell m}(\mathbf{r}')$ by using the expansion

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi \sum_{\ell m} i^{\ell} j_{\ell}(kr) Y_{\ell m}(\boldsymbol{\theta}_{\mathbf{r}}) Y_{\ell m}^*(\boldsymbol{\theta}_{\mathbf{k}}), \quad (7.154)$$

where $\boldsymbol{\theta}_{\mathbf{f}}$ and $\boldsymbol{\theta}_{\mathbf{k}}$ refer respectively to the directions of \mathbf{r} and \mathbf{k} and where j_{ℓ} is a spherical Bessel function.

- (b) Assuming scale-invariant potential fluctuations, prove the relation (7.77). It will be necessary to use the relation

$$\int_0^\infty \frac{dx}{x} |j_{\ell}(x)|^2 = \frac{1}{2\ell(\ell+1)}. \quad (7.155)$$

7.6 Prove (7.62) by calculating $a_{\text{exit}}(k)$ for modes that enter the Hubble radius during the radiation epoch.

7.7 The comparison of measured $P(k)$ and C_{ℓ} with theoretical predictions is limited by “cosmic variance.” Inflation predicts that the $\delta_{\mathbf{k}}$ in (7.24) are random numbers distributed Gaussianly about zero. A measurement of $P(k)$ over an interval Δk consists of an average over the $N_{\Delta k}$ modes in that interval given by

$$N_{\Delta k} = \frac{V k^3}{4\pi^2} \frac{\Delta k}{k} \quad (7.156)$$

where V is the volume of the survey. (In fact only half of these modes are independent because of the reality condition on the $\delta_{\mathbf{k}}$.) A measurement of $P(k)$ will then differ from the theoretical prediction by a random number of variance $P(k)/\sqrt{N_{\Delta k}}$. This is called the cosmic variance.

- (a) For the low k points of the SDSS power spectrum in Fig. 7.6, estimate the cosmic variance and compare with the error bars. (SDSS uses galaxies out to $z \sim 0.3$ over $\sim 1/3$ of the sky.)
- (b) The C_ℓ in the temperature anisotropy spectrum also suffer from cosmic variance and this causes the large error bars at low multipoles in Fig. 7.11. Estimate the uncertainty in C_2 .

7.8 Use the $\Delta\theta_s$ from Fig. 5.8 and the $\Delta\ell$ from (7.83) to verify the $(\Omega_M, \Omega_\Lambda)$ assignments of the theoretical spectra in Fig. 7.15.

7.9 Consider the mass $M_R(\mathbf{R})$ contained in a sphere of radius R centered at \mathbf{R} . The local fluctuations of mass can be estimated by comparing it to the mean mass around the sphere, e.g.

$$\delta M_R = M_R(\mathbf{R}) - (1/8)M_{2R}(\mathbf{R}) . \quad (7.157)$$

Calculate the variance $\langle \delta M_R^2 \rangle$ and show that it depends only on the function $\Delta(k)$ for values of $\ln k$ near $\ln(1/R)$.

7.10 What is the present length of the largest mode that was never outside the Hubble radius during the radiation epoch? This is the largest mode that would not be expected to have a roughly scale-invariant spectrum.

7.11 Draw qualitatively a diagram showing our LastSS, $\chi(z_{\text{rec}})$, and our “reionization surface,” $\chi(z_{\text{reion}})$, defined by the redshift of reionization, $z_{\text{reion}} \sim 10$. The photons that scatter at a point, \mathbf{r} , on our reionization surface that are latter observed by us at t_0 must have originated on the LastSS of the point \mathbf{r} *at the time of reionization*. Draw that LastSS and calculate its radius. What is the angle subtended by \mathbf{r} ’s LastSS as seen by us. This is the correlation angle for polarization due to reionization. What is the associated multipole, ℓ_{reion} ?

7.12 Consider a reference frame in which the CMB has an isotropic Planckian phase space distribution:

$$F(p_x, p_y, p_z) = F(E) = \frac{1}{(2\pi)^3} \frac{1}{\exp(E/T) - 1} . \quad (7.158)$$

Consider a second (primed) frame that is related to the first by a Lorentz boost along the z direction with $\beta \ll 1$. Consider the dN photons that in the original frame are, at $t = 0$, inside a cube at $\mathbf{R} = (0, 0, 0)$ of volume $dx dy dz$ and in a momentum space cube at $\mathbf{p} = (p_x, p_y, p_z)$ of volume $dp_x dp_y dp_z$. By performing the Lorentz transformations to order β , show that at $t' = 0$ the dN photons are in the volume

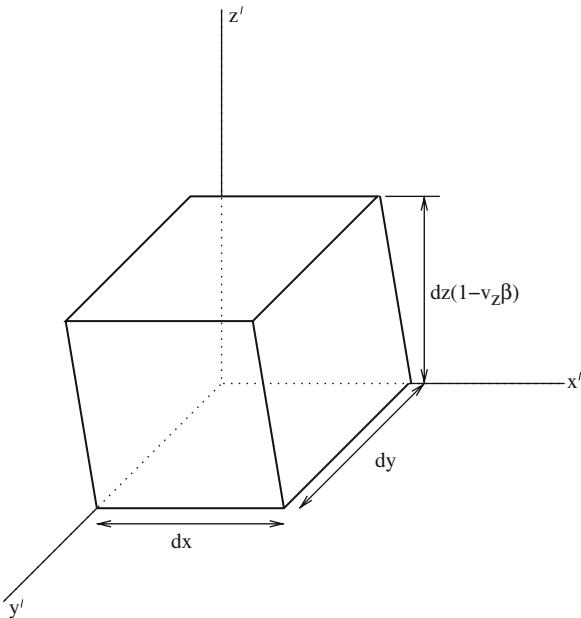


Fig. 7.22 Photons that are inside a cube of volume $dxdydz$ at $x = y = z = t = 0$ are, in the primed system, inside the parallelepiped shown in the figure at $t' = 0$

shown in Fig. 7.22. Show that in momentum space the particles are in a space of volume $(1 + v_z\beta)dp_x dp_y dp_z$. Conclude that the phase-space densities in the two frames are equal with

$$F' (p'_x, p'_y, p'_z) = F(E'(1 - \beta \cos \theta')) , \quad (7.159)$$

where $\theta' = \cos^{-1}(p'_z/p')$ is the angle of the photon seen in the primed frame. This distribution corresponds to a Planckian distribution with an angle-dependent temperature:

$$T'(\theta') = T(1 - \beta \cos \theta') . \quad (7.160)$$

7.13 Consider an isotropic universe with a coordinates (t, r, θ, ϕ) and a metric of the form (7.112).

(a) Show that the field equations are [160]

$$\dot{\beta}^2 + 2\dot{\alpha}\dot{\beta} + e^{-2\beta} - e^{-2\alpha}[2\beta'' + 3(\beta')^2 - 2\alpha'\beta'] = 8\pi G\rho , \quad (7.161)$$

$$2\ddot{\beta} + 3\dot{\beta}^2 + e^{-2\beta} - e^{-2\alpha}(\beta')^2 = 0 , \quad (7.162)$$

$$2\dot{\beta}' + 2\dot{\beta}\beta' - 2\dot{\alpha}\beta' = 0 , \quad (7.163)$$

where the dots and primes signify derivatives with respect to t and r and where

$$A(r, t) = e^{2\alpha} \quad B(r, t) = e^{2\beta}. \quad (7.164)$$

(b) Show that the solution of (7.163) is

$$\beta' = \frac{e^{\alpha-\beta}}{g(r)} \Rightarrow e^\alpha = g(r) e^\beta \beta', \quad (7.165)$$

where $g(r)$ is an arbitrary function of r .

(c) By defining

$$e^\beta \equiv r b(r, t) \quad g(r) \equiv \frac{1}{(1 - \kappa(r)r^2)^{1/2}} \quad (7.166)$$

show that the metric takes the form (7.114).

(d) With the notation given by (7.166) show that (7.161) and (7.162) become

$$\frac{\partial}{\partial r} [\dot{b}^2 b r^3 + b r^3 \kappa(r)] = [(8/3)\pi G\rho] \frac{\partial}{\partial r} (b r^3), \quad (7.167)$$

$$2\ddot{\frac{b}{b}} + \frac{\dot{b}^2}{b^2} + \frac{\kappa}{b^2} = 0. \quad (7.168)$$

(e) Show that the solution of (7.168) is

$$\dot{b}^2 b + \kappa b = F(r), \quad (7.169)$$

where $F(r)$ is an arbitrary function of R . Now redefine the coordinate r and the functions b and κ as follows

$$\tilde{r} = r \left[\frac{F(r)}{A} \right]^{1/3} \quad \tilde{b} = a \left[\frac{A}{F(r)} \right]^{1/3} \quad \tilde{\kappa} = \kappa \left[\frac{A}{F(r)} \right]^{2/3},$$

where A is a constant. Show that with this transformation, (7.167) and (7.168) become (7.115) and (7.116).

Appendix A

Lorentz Vectors and Tensors

Consider a coordinate system \tilde{x}^μ , $\mu = 0, 1, 2, 3$ and another coordinate system $x(\tilde{x})$. The differential dx^μ is given by the chain rule

$$dx^\mu = \frac{\partial x^\mu}{\partial \tilde{x}^\alpha} d\tilde{x}^\alpha. \quad (\text{A.1})$$

Any object that transforms like dx^μ is called a “contravariant 4-vector,” i.e.

$$V^\mu = \frac{\partial x^\mu}{\partial \tilde{x}^\alpha} \tilde{V}^\alpha, \quad (\text{A.2})$$

where \tilde{V} is the vector in the \tilde{x} system and V is the vector in the x system. The best examples of contravariant vectors are the 4-velocity, $v^\mu = dx^\mu/d\tau$, and the 4-momentum, $p^\mu = mv^\mu$. These two are obviously contravariant 4-vectors because dx^μ is a contravariant 4-vector and $d\tau$ is an invariant. (It is defined by the clock reading in the clock’s rest-frame.)

A second kind of vector arises naturally when one considers the 4-gradient of a scalar (Lorentz invariant) function ϕ . By the chain rule again, this object transforms as

$$\frac{\partial \phi}{\partial x^\mu} = \frac{\partial \tilde{x}^\alpha}{\partial x^\mu} \frac{\partial \phi}{\partial \tilde{x}^\alpha}. \quad (\text{A.3})$$

Any object that transforms like a 4-gradient is called a “covariant 4-vector,” i.e.

$$U_\mu = \frac{\partial \tilde{x}^\alpha}{\partial x^\mu} \tilde{U}_\alpha. \quad (\text{A.4})$$

Tensors are just generalizations of vectors. For example, a contravariant rank-2 tensor, $T^{\mu\nu}$, transforms as

$$T^{\mu\nu} = \frac{\partial x^\mu}{\partial \tilde{x}^\alpha} \frac{\partial x^\nu}{\partial \tilde{x}^\beta} \tilde{T}^{\alpha\beta}, \quad (\text{A.5})$$

while a covariant rank-2 tensor, $T_{\mu\nu}$, transforms as

$$T_{\mu\nu} = \frac{\partial \tilde{x}^\alpha}{\partial x^\mu} \frac{\partial \tilde{x}^\beta}{\partial x^\nu} \tilde{U}_{\alpha\beta}. \quad (\text{A.6})$$

The energy-momentum tensor of a system of particles (4.35) is an example of a contravariant tensor. On the other hand, the energy-momentum tensor of a scalar field is more naturally written as a covariant tensor (4.77). Since we will want to put the two in the same Friedmann equation, we will need a way to change covariant tensors into contravariant tensors. This will be done soon.

In this book we are mostly concerned with Lorentz transformations and we will now concentrate on this simple class of transformations. As shown in Exercise 3.2, the transformations coefficients $\partial \tilde{x}^\alpha / \partial x^\mu$ of Lorentz boosts and rotations must be constants satisfying

$$\eta_{\kappa\lambda} \frac{\partial x^\kappa}{\partial \tilde{x}^\alpha} \frac{\partial x^\lambda}{\partial \tilde{x}^\beta} = \eta_{\alpha\beta}. \quad (\text{A.7})$$

Comparison with (A.6) shows that $\eta_{\mu\nu}$ is an invariant rank-2 covariant tensor. It is easy to show that $\eta^{\mu\nu} = \eta_{\mu\nu}$ is an invariant rank-2 contravariant tensor:

$$\eta^{\alpha\beta} \frac{\partial x^\kappa}{\partial \tilde{x}^\alpha} \frac{\partial x^\lambda}{\partial \tilde{x}^\beta} = \eta^{\kappa\lambda}. \quad (\text{A.8})$$

The rank of a tensor can be changed by contraction. For example, it is easy to show that the 4-divergence of a Lorentz vector, $\partial V^\mu / \partial x^\mu$, is a Lorentz invariant. Similarly, the 4-divergence of a Lorentz tensor, $\partial T^{\mu\nu} / \partial x^\mu = V^\nu$, is a Lorentz vector.

A quantity with vanishing 4-divergence gives a local Lorentz invariant conservation law, e.g.

$$\frac{\partial j^\mu}{\partial x^\mu} = \frac{\partial j^0}{\partial t} - \nabla \cdot \mathbf{j} = 0, \quad (\text{A.9})$$

meaning that the time derivative of the “density” (j^0) is equal to the divergence of the “current” (\mathbf{j}). If there is a closed surface over which the current vanishes, the total “charge” within the surface is conserved:

$$\frac{d}{dt} \int j^0 dV = 0. \quad (\text{A.10})$$

We will now show that it is possible to change a contravariant 4-vector into a covariant 4-vector by the simple operation

$$\tilde{V}_\mu = \eta_{\mu\nu} \tilde{V}^\nu \quad V_\mu = \eta_{\mu\nu} V^\nu. \quad (\text{A.11})$$

To verify that the objects obtained by this operation are indeed contravectors we transform from the \tilde{x} system:

$$\frac{\partial \tilde{x}^\alpha}{\partial x^\nu} \tilde{V}_\alpha = \frac{\partial \tilde{x}^\alpha}{\partial x^\nu} \eta_{\alpha\beta} \tilde{V}^\beta = \frac{\partial \tilde{x}^\alpha}{\partial x^\nu} \eta_{\kappa\lambda} \frac{\partial x^\kappa}{\partial \tilde{x}^\alpha} \frac{\partial x^\lambda}{\partial \tilde{x}^\beta} \tilde{V}^\beta, \quad (\text{A.12})$$

where in the second form we have used (A.7). The first and third factors on the right combine to give a Kronecker δ

$$\frac{\partial \tilde{x}^\alpha}{\partial x^\nu} \frac{\partial x^\kappa}{\partial \tilde{x}^\alpha} = \frac{\partial x^\kappa}{\partial x^\nu} = \delta_\nu^\kappa, \quad (\text{A.13})$$

so we find

$$\frac{\partial \tilde{x}^\alpha}{\partial x^\nu} \tilde{V}_\alpha = \eta_{\nu\lambda} \frac{\partial x^\lambda}{\partial \tilde{x}^\beta} \tilde{V}^\beta = \eta_{\nu\lambda} V^\lambda = V_\nu. \quad (\text{A.14})$$

Comparing the extreme right with the extreme left, we see that V_ν as defined by (A.11) is indeed a covariant 4-vector.

Equation (A.11) can be generalized to tensors in the obvious way, e.g.

$$V_{\mu\nu} = \eta_{\mu\alpha} \eta_{\nu\beta} V^{\alpha\beta}. \quad (\text{A.15})$$

Appendix B

Natural Units

In this book we often use “natural units” meaning that in formulas we omit factors of \hbar , c , and k . This is basically an ink-saving convention because there is no information contained in such factors. Consider two quantities, A and B , which have the dimensions

$$[A] = \text{length}^l \text{ mass}^m \text{ time}^n . \quad (\text{B.1})$$

$$[B] = \text{length}^\lambda \text{ mass}^\mu \text{ time}^\nu . \quad (\text{B.2})$$

We can generally find unique exponents x , y , and z such that the following equation is dimensionally correct

$$A = c^x \hbar^y B^z , \quad (\text{B.3})$$

where x , y , and z are the solutions of the simultaneous equations:

$$l = x + 2y + \lambda z \quad m = y + \mu z \quad n = -x - y + \nu z . \quad (\text{B.4})$$

Since the exponents are unique, we can drop the factors of c and \hbar and write (B.3) as

$$A = B^z . \quad (\text{B.5})$$

If we have used the correct z in this formula, x and y can be found by dimensional analysis and the “dimensionally correct” (B.3) recovered.

We can obviously generalize this procedure for formulas involving sums of terms. The dimension of temperature can also be added since temperatures can be converted to energies by multiplying by k .

The most common use of natural units in this book involves relativistic expressions where the factors of c are omitted, e.g.

$$d\tau^2 = dt^2 - dx^2 \quad \rightarrow \quad d\tau^2 = dt^2 - dx^2 c^{-2} , \quad (\text{B.6})$$

$$E^2 = p^2 + m^2 \quad \rightarrow \quad E^2 = p^2 c^2 + m^2 c^4 \quad (\text{B.7})$$

and

$$d_H = H_0^{-1} \quad \rightarrow \quad d_H = c H_0^{-1}. \quad (\text{B.8})$$

Of special interest are “dimensionless” quantities, e.g. velocities

$$v \sim 10^{-3} \quad \rightarrow \quad \frac{v}{c} \sim 10^{-3} \quad (\text{B.9})$$

and gravitational potentials

$$\Phi_g \sim 10^{-6} \quad \rightarrow \quad \frac{\Phi_g}{c^2} \sim 10^{-6}. \quad (\text{B.10})$$

In quantum formulas, we usually set quantities equal to some power of an energy (or temperature). Energies, masses, and momenta clearly have $z = 1$ as in (B.7). Of more interest is the fact that lengths and times have $z = -1$, e.g.

$$\lambda = \frac{2\pi}{E\gamma} \quad \rightarrow \quad \lambda = \frac{2\pi\hbar c}{E\gamma} \quad (\text{B.11})$$

$$\omega = E\gamma \quad \rightarrow \quad \omega = E\gamma/\hbar. \quad (\text{B.12})$$

The fact that lengths have $z = -1$ means that cross-sections have $z = -2$

$$\sigma_T = \frac{8\pi}{3} \alpha^2 m_e^{-2} \quad \rightarrow \quad \sigma_T = \frac{8\pi}{3} \alpha^2 (m_e c^2)^{-2} (\hbar c)^2. \quad (\text{B.13})$$

Similarly, particle densities have $z = 3$:

$$n = \frac{2.4}{\pi^2} T^3 \quad \rightarrow \quad n = \frac{2.4}{\pi^2} (kT)^3 \frac{1}{(\hbar c)^3} \quad (\text{B.14})$$

$$\begin{aligned} n &= 2 \left(\frac{mT}{2\pi} \right)^{3/2} \exp(-m/T) \\ &\rightarrow \quad n = 2 \left(\frac{mc^2 kT}{2\pi} \right)^{3/2} \exp(-mc^2/kT) \frac{1}{(\hbar c)^3}. \end{aligned} \quad (\text{B.15})$$

Note that in this last non-relativistic formula, the factors of c cancel as expected.

Energy densities have $z = 4$

$$\rho = \frac{2\pi^2}{30} T^4 \quad \rightarrow \quad \rho = \frac{2\pi^2}{30} (kT)^4 \frac{1}{(\hbar c)^3}. \quad (\text{B.16})$$

Finally, the Friedmann equation has $z = 2$

$$H^2 = \frac{8\pi^3}{45} \frac{T^4}{m_{\text{pl}}^2} \quad \rightarrow \quad H^2 = \frac{8\pi G\rho}{3}, \quad (\text{B.17})$$

where the Planck mass m_{pl} is defined by $m_{\text{pl}} = (\hbar c/G)^{1/2}$.

$$G = \frac{1}{m_{\text{pl}}^2} \quad \rightarrow \quad G = \frac{\hbar c}{m_{\text{pl}}^2}. \quad (\text{B.18})$$

Appendix C

Standard Particles and Beyond

The known particles listed in Table 6.2 can scatter elastically and inelastically with each other. The rules that tell us how to calculate the reaction rates constitute the “standard model of particle physics” [162–165, 18]. In this model, most of the allowed reactions can be assembled from primitive three-particle reactions of the form

$$\text{fermion} + \text{boson} \leftrightarrow \text{fermion}' \quad (\text{C.1})$$

and

$$\text{fermion} + \text{antifermion} \leftrightarrow \text{boson}. \quad (\text{C.2})$$

(There are also three-body reactions of the form $(\text{boson} + \text{boson} \leftrightarrow \text{boson})$ but they are not important for the reactions considered here.) The only allowed three-particle reactions are those that conserve electric charge and, in the case of quark-gluon interactions, “color.” Examples of some assembled reactions are shown in Figs. C.1 and C.2.

The Standard Model is a quantum field theory that provides the rules for turning each picture of Figs. C.1 and C.2 into amplitudes. Rates for a given reaction are found by summing over all amplitudes for the same initial and final particles and then squaring the sum. When the as-yet undiscovered “Higgs boson” is added to the known particles, the standard model gives well-defined rates for all physical processes, i.e. the theory is renormalizable.

Amongst other things, each “vertex” of a fundamental three-particle reaction is associated with a coupling constant that gives the strength of the interaction. In the standard model, all coupling constants are within an order of magnitude of the electric charge. A diagram with n vertices therefore gives an amplitude of order $\alpha^{n/2}$.

When cross-sections are calculated by summing amplitudes, the effective coupling constants appear to evolve with energy. All effective constants are expected to become equal to each other at the “Grand Unification” energy scale, estimated to be of order 10^{15} to 10^{16} GeV. A more complete theory containing heavy particles

Fig. C.1 Diagrams for the reactions $e^+e^- \rightarrow \gamma\gamma$, $\nu_e\bar{\nu}_e \rightarrow e^+e^-$, and $\nu_\mu\bar{\nu}_\mu \rightarrow e^+e^-$

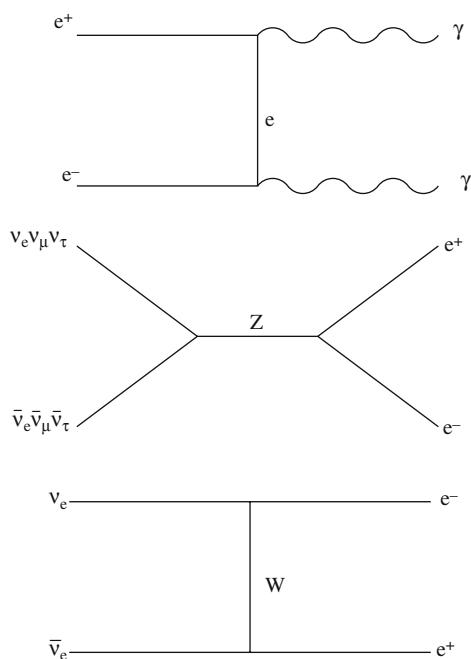
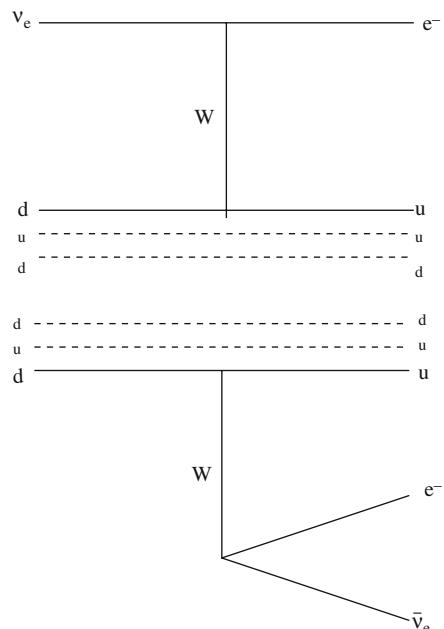


Fig. C.2 Diagrams for the reactions $\nu_en \rightarrow e^-p$ and $n \rightarrow pe^-\bar{\nu}_e$



is expected to become manifest at these energies. Quantum gravity effects should become important at the Planck scale $\sim 10^{19}$ GeV.

Below the deconfinement temperature, ~ 400 MeV, quarks and gluons interact in a way that confines them in hadrons. Baryons, e.g. protons and neutrons, are bound states of three quarks and associated gluons while mesons, e.g. π^\pm, π^0 , are bound states of quarks and antiquarks. Above the deconfinement temperature, quarks and gluons are believed to be liberated and behave as normal particles.

Most cosmologically interesting reactions involve “electroweak” interactions where the bosons in the fundamental reactions are γ, Z^0 , or W^\pm . The fundamental three-particle interactions are of two types, “neutral current interactions” where the boson is γ or Z^0 and “charged current interactions” where the boson is W^\pm . Reactions due to the exchange of W or Z bosons are said to be due to “weak” interactions because in the low-energy limit, the amplitudes are inversely proportional to the square of the heavy boson mass and are thus “small.”

Neutral current reactions mostly appear in elastic scattering and particle–antiparticle annihilation. In the standard model, they all obey the rule that the two fermions must be identical, i.e.

$$e^- Z^0 \leftrightarrow e^- \quad \text{but not} \quad e^- Z^0 \leftrightarrow \mu^-, \quad (C.3)$$

$$e^+ e^- \leftrightarrow Z^0 \quad \text{but not} \quad e^+ \mu^- \leftrightarrow Z^0. \quad (C.4)$$

Charged current reactions change fermions into other types of fermions e.g.:

$$\nu_e W^- \leftrightarrow e^- \quad \text{and} \quad u W^- \leftrightarrow d. \quad (C.5)$$

Charged current interactions can also annihilate fermions with antifermions:

$$\nu_e e^+ \leftrightarrow W^+ \quad \text{and} \quad u \bar{d} \leftrightarrow W^+. \quad (C.6)$$

Because W^\pm have integer charge, there are no reactions changing leptons into quarks.

Diagrams for the annihilation of electron–positron pairs and for neutrinos are shown in Fig. C.1.

Reactions for neutrons and protons can be treated by including “spectator quarks” in the diagram. Examples are shown in Fig. C.2.

The rules for turning the diagrams into amplitudes are given in the standard texts [162–165]. Besides the coupling constants for each vertex, an amplitude contains kinematic factors as dictated by quantum field theory. For example, in the low-energy neutrino interactions important for cosmology at $T \sim$ MeV, the presence of the W^\pm and Z^0 give factors of $m_W^{-2} \sim m_Z^{-2}$ in the amplitudes. Combined with the coupling constants, this gives squared amplitudes of order $\alpha^2/m_W^4 \sim G_F^2$ where G_F is the Fermi coupling constant. Cross-sections and decay rates are obtained by multiplying by factors that are related to the total phase space for the final state particles. The forms of these factors can often be guessed by dimensional analysis.

Cross-sections have dimensions of energy⁻² so G_F^2 must be multiplied by the square of an energy. For example the cross-section for $\nu_e \bar{\nu}_e$ annihilation is

$$\sigma = \frac{G_F^2 E_{\text{cm}}^2}{12\pi} \left[\frac{1}{2} + 2 \sin^2 \theta_W + 4 \sin^4 \theta_W \right], \quad (\text{C.7})$$

while that for $\nu_\mu \bar{\nu}_\mu$ and $\nu_\tau \bar{\nu}_\tau$ annihilation is

$$\sigma = \frac{G_F^2 E_{\text{cm}}^2}{12\pi} \left[\frac{1}{2} - 2 \sin^2 \theta_W + 4 \sin^4 \theta_W \right]. \quad (\text{C.8})$$

In these two expressions, $\sin^2 \theta_W \sim 0.23$ is one of the parameters of the standard model.

In the range $m_e \ll E_\nu \ll m_p$ the cross-section for $\nu_e n \rightarrow e^- p$ is

$$\sigma = \frac{G_F^2 E_\nu^2}{\pi} [1 + 3g_A^2] \cos^2 \theta_c \sim 9 \times 10^{41} \left(\frac{E_\nu}{10 \text{ MeV}} \right)^2 \text{ cm}^2, \quad (\text{C.9})$$

where $\cos^2 \theta_c \sim 0.98$ and $g_A \sim 1.2$.

Decay rates have dimension of energy so G_F^2 must be multiplied by the fifth power of an energy. For example, the neutron decay rate is

$$\tau_n^{-1} \sim G_F^2 Q^5, \quad (\text{C.10})$$

where $Q = m_n - m_p - m_e$ is the energy release per decay.

Supersymmetric extensions of the standard model are cosmologically interesting because they provide a well-motivated nonbaryonic dark matter candidate. In these theories, each fermion (boson) in Table 6.2 is paired with a heavy, as-yet undiscovered, partner that is a boson (fermion). For example the spin 1 photon, γ is paired with a heavy spin 1/2 “photino” $\tilde{\gamma}$. Each spin 1/2 quark, q is paired with a heavy spin 0 “squark” \tilde{q} .

The interactions of the supersymmetric particles can be found by taking the fundamental three-particle interactions of the normal particles and turning two of the particles into their supersymmetric partners. For example:

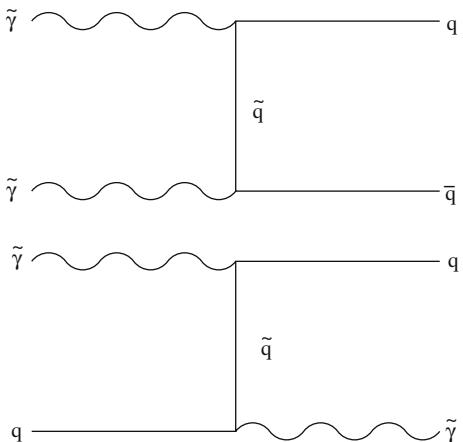
$$e^- \gamma \leftrightarrow e^- \rightarrow e^- \tilde{\gamma} \leftrightarrow \tilde{e}^-, \quad (\text{C.11})$$

where \tilde{e} is the “selectron.”

The diagrams for photino annihilation to quarks and photino scattering on quarks are shown in Fig. C.3.

The lightest supersymmetric particle (LSP), called generically χ , is expected to be a linear combination of the photinos, the Zino (partner of Z^0) and the higgsinos (partners of the Higgs bosons). The relic abundance of this particle is determined by its annihilation cross-section (Sect. 6.6). As with neutrinos, the annihilation

Fig. C.3 Diagrams for the reactions $\tilde{\gamma}\tilde{\gamma} \rightarrow q\bar{q}$ and $\tilde{\gamma}q \rightarrow \tilde{\gamma}q$. The first would contribute to the photino annihilation diagram, determining the photino relic abundance. The second determines the photino elastic scattering cross-section on nuclei



cross-section will contain a factor α^2 , a factor m^{-4} for the exchanged particle, and a kinematic factor that turns out to be m_χ^2 :

$$\sigma(v/c) \sim \frac{\alpha^2}{m_{\tilde{e}}^4} m_\chi^2. \quad (\text{C.12})$$

For a given m_χ it is then possible to find an appropriate $m_{\tilde{e}}$ such that the annihilation cross-section is of the correct magnitude ($\sigma v/c \sim 10^{-37} \text{ cm}^2$, see Sect. 6.6) to provide the cosmological dark matter.

Appendix D

Magnitudes

Because of the wide range of stellar and galactic luminosities, astronomers generally give luminosities on a logarithmic scale of “absolute magnitude.” For an object of total luminosity L , the absolute bolometric luminosity is defined as

$$M_{\text{bol}} \equiv -2.5 \log(L/L_{\odot}) + 4.76, \quad (\text{D.1})$$

where L_{\odot} is the solar luminosity and 4.76 is the solar magnitude. We note that because of the minus sign bright stars have small magnitudes. The choice of a factor 2.5 may seem strange but it has the characteristic of giving simple expressions for both large luminosity differences (5 magnitudes is a factor 100 in luminosity) and small luminosity differences ($\Delta L/L \sim \Delta M$).

It is very difficult to measure bolometric magnitudes and astronomers generally observe with a filter that makes them sensitive to a particular spectral band. The most widely used filters are described in Table D.1. The band passes for the traditional Johnson-Cousins *UBVRI* bands [38] are superimposed on the spectrum of the A0 star in Fig. 2.1 while those of the SDSS *ugriz* bands [39] are superimposed on the spectrum of the G0 star.

For an object of luminosity L_C in the band C , the absolute magnitude M_C is defined by

$$M_C \equiv -2.5 \log(L_C/L_{C\odot}) + M_{C\odot} \quad C = U, B, V, R, I\dots \quad (\text{D.2})$$

The solar magnitude $M_{C\odot}$ is set by convention. Two conventions are in wide use. The conventional “Vega-based” magnitudes choose the $M_{C\odot}$ so that all the M_C for the A0 star Vega (α Lyrae) are near zero. Since the flux in the U band of an A0 star is small compared to the B band (Fig. 2.1), this leads to a M_U that does not reflect the true flux compared to M_B . Because of this, the use of AB magnitudes are becoming increasingly popular, especially for *ugriz* magnitudes. For AB magnitudes, the $M_{C\odot}$ are chosen so that all the M_C are equal for a star with an energy spectrum $dE/d\lambda \propto 1/\lambda^2$ (corresponding to $dE/d\nu = \text{constant}$).

The difference between two magnitudes, e.g. $M_B - M_V$, is a “color index” of an object. By convention, a color index is always the shorter wavelength magnitude

Table D.1 Standard filters in the Johnson–Cousins *UBVRI* and SDSS *ugriz* systems used for ground-based observations [36]. The *UBVRI* and bolometric solar magnitudes are Vega-based while the *ugriz* magnitudes are approximate AB magnitudes. The last column gives the band’s Galactic interstellar absorption (extinction coefficient) relative to that of the V band. Except for certain infrared bands, wavelengths between radio ($\lambda \sim 1$ cm) and near-infrared ($\lambda \sim 1000$ nm) are strongly absorbed by the atmosphere and observations are done from aircraft, balloons, or satellites. Satellite observations are also required for short wavelengths between $\lambda \sim 300$ nm and those of TeV photons

Filter	$\langle \lambda \rangle$ (nm)	$\Delta\lambda$ (FWHM)	$M_{\lambda\odot}$ (Sun)	A_C/A_V
Bolometric		∞	4.76	
<i>U</i>	365	66	5.61	1.531
<i>B</i> (“blue”)	445	94	5.48	1.324
<i>V</i> (“visible”)	551	88	4.83	1.0
<i>R</i>	658	138	4.42	0.748
<i>I</i>	806	149	4.08	0.482
<i>J</i>	1200	213	3.64	0.282
<i>K</i>	2190	390	3.28	0.112
<i>u'</i>	352	63	6.55	
<i>g'</i>	480	141	5.12	
<i>r'</i>	625	139	4.68	
<i>i'</i>	769	154	4.57	
<i>z'</i>	911	141	4.54	

minus the longer wavelength magnitude. Because of the inverse log scale, a blue (hot) star has small color indices.

As with luminosities, astronomers generally give fluxes on a logarithmic scale of “apparent magnitude.” For a star with flux ϕ_C (outside the Earth’s atmosphere) in the band C , the apparent magnitude m_C is defined as

$$m_C \equiv C = -2.5 \log \phi_C + \alpha_C \quad C = U, B, V, R, I, \dots$$

where α_C is a constant. If there is no interstellar absorption, we have $\phi_C = L_C/(4\pi R^2)$ where R is the distance to the star. This implies that $m = M + 5 \log R + \text{constant}$. The constant α_C in the definition of m_C is chosen so that, in the absence of absorption, the apparent magnitude is equal to the absolute magnitude if $R = 10$ pc:

$$m_C = M_C + 5 \log(R/10 \text{ pc}) + A_C \quad C = U, B, V, R, I, \dots$$

The “extinction coefficient,” A_C , takes into account interstellar or intergalactic absorption. The absorption length, corresponding to $A_V \sim 1$, is typically ~ 1 kpc in the galactic plane, but varies greatly according to the line-of-sight because of discrete absorbing clouds. Perpendicular to the galactic plane, the absorption of an extragalactic source is typically 10% ($A_V \sim 0.1$).

We note that in the absence of a wavelength-dependent absorption, the color index of an object is equal to its apparent color index, e.g. $m_B - m_V = B - V = M_B - M_V$. In reality, absorption is stronger at short wavelengths than long so in the presence of absorption, an object is “reddened.” Table D.1 gives typical relative absorption for the interstellar medium of the Milky Way.

Appendix E

Solutions and Hints for Selected Exercises

Chapter 1

1.3 The universe is expanding today because it was expanding yesterday (see (1.58)). It was expanding yesterday because.....

It will be difficult to get an ultimate explanation since it will require knowledge of the physics that was in charge of things at the Beginning.

Chapter 2

2.1 $\phi \sim 100 \text{ m}^{-2} \text{s}^{-1} / z^2$. $\phi_{\text{galaxy}} \sim 10^{-2} \phi_{\text{star}}$.

2.2 $n_{\text{starlight}} \sim 5 \times 10^{-7} n_{\text{CMB}}$. $n_{\text{p} \rightarrow {}^4\text{He}} \sim 5 \times 10^{-3} n_{\text{p}}$.

2.3 Compton scattering dominates with a mean free path of order $10^3 d_{\text{H}}$.

2.4 It is possible to count the number of galaxies with a redshift less than z . The volume of the corresponding space is $V = (4\pi/3)z^3 d_{\text{H}}^3 \propto h_{70}^{-3}$ so the measured number density is $\propto h_{70}^3$.

2.5 You should get a reasonable value of H_0 with this method.

2.7 NGC 5033 is a typical spiral galaxy so you should get numbers comparable to those in the text for typical spiral galaxies.

Chapter 3

3.4 The rocket is in free-fall after lift-off and therefore follows a geodesic. The rocket's clock must, therefore, measure a longer elapsed time than the ground clock which is not in free-fall. On the other hand, an airplane is not in free-fall so the answer depends on its velocity and altitude.

3.9 To order z^2 , we find $d_1 = d_A < d_0 < d_L$. The distances differ in the coefficient of the z^2 term so they differ by $\sim 10\%$ at $z \sim 0.1$ or $R \sim 430 h_{70}^{-1} \text{ Mpc}$.

3.10 If a comoving observer sees the explorer leave with velocity v , then

$$\frac{dR}{dt} = a \frac{d\chi}{dt} = v. \quad (\text{E.1})$$

The derivative is with respect to t since this is the time measured by the comoving clock. Using the metric, we then find

$$\frac{d\chi}{d\tau} = \frac{v}{a_0 \sqrt{1 - v^2}} \quad (\text{E.2})$$

$$\frac{dt}{d\tau} = \frac{1}{\sqrt{1 - v^2}}. \quad (\text{E.3})$$

For an empty universe ($a \propto t$), the most distant galaxy that can be reached has a redshift of $z = v/c$ (not too surprising). For a critical universe ($a \propto t^{2/3}$), the most distant galaxy that can be reached has a redshift of $z = 2v/c$. A simple Newtonian argument can explain the extra distance in a critical universe.

3.13 B measures the length of the rod by radar, emitting two photons at $t = -L/2$ and receiving the echo at $t = L/2$. B sees a Lorentz contraction because B's clock measures $t' = \pm 0.5L\sqrt{1 - \beta^2}$ and therefore concludes that the bar has a length $L' = L\sqrt{1 - \beta^2}$.

Chapter 4

4.1 For $\ddot{a} = 0$ and for $dr = 0$ (for a comoving clock), we have

$$dT = dt + (1/2)(\dot{a}/a)^2(ar)^2dt = dt + (1/2)v^2dt, \quad (\text{E.4})$$

where $v = HR$ is the Hubble velocity of the comoving clock. This gives

$$\frac{dt}{1 + v^2/2} \sim dT(1 - v^2/2) \quad (\text{E.5})$$

This is as predicted by special relativity where we expect the time dt measured by the moving clock to be smaller than the time dT measured by the stationary clock by a factor $\sqrt{1 - v^2} \sim 1 - v^2/2$.

To second order, the trajectory of the photon in Fig. 4.1 between t_a and t is governed by

$$a(t)\chi \sim \int_{t_a}^t dt' (1 - H(t)(t' - t)) = (t - t_a) + \frac{H(t)}{2}(t - t_a)^2. \quad (\text{E.6})$$

For the trajectory between t and t_b we have

$$a(t)\chi \sim (t_b - t) - \frac{H(t)}{2}(t_b - t)^2 \quad (\text{E.7})$$

Taking the sum we find

$$a(t)\chi \sim \frac{(t_b - t_1)}{2} + \frac{H(t)}{4} [(t - t_a)^2 - (t_b - t)^2]. \quad (\text{E.8})$$

Since, to first order, both $(t - t_a)$ and $(t_b - t)$ are equal to $a(t)\chi$, the second term vanishes to order χ^2 so we have

$$a(t)r \sim \frac{(t_b - t_1)}{2} + O(r^3) \quad (\text{E.9})$$

where we use the fact that $\chi = r + O(r^3)$.

Taking the difference between the two trajectories we have

$$0 = t_b + t_a - 2t - \frac{H(t)}{2} [(t_b - t)^2 + (t - t_a)^2]. \quad (\text{E.10})$$

Using $(t - t_a) \sim (t_b - t) \sim a(t)r$, we find

$$\frac{t_b + t_a}{2} = t + (1/2)a\dot{r}r^2 + O(r^3) \quad (\text{E.11})$$

which is equivalent to (4.11).

4.3 For a non-relativistic ideal gas, $E \sim m$ and $\langle p^2/2m \rangle = (3/2)T$ from which it follows that $p = nT \ll mT = \rho$. For a relativistic ideal gas, $p = E$ from which it follows that $p = \rho/3$.

Chapter 5

5.1 For $\Omega_T = \Omega_M = 0$, we find

$$a(t) \propto t \quad q_0 = 0 \quad a_0\chi_1(z) = H_0^{-1} \ln(1+z). \quad (\text{E.12})$$

Using $a_0 = H_0^{-1}$ for an empty universe, we find

$$r_1(z) = \frac{1 + z + (1+z)^{-1}}{2}. \quad (\text{E.13})$$

For $\Omega_T = \Omega_M = 1$, the minimum angular size occurs for $z = 1.25$.

5.2 The equation should be a good approximation as long as there are no other relativistic species, i.e. for $T \ll m_e$. After the neutrinos start to become non-relativistic, the equation is still a good approximation because the Ω_M term dominates in any case.

For t_{rec} we can make the approximation that the universe is matter-dominated in which case the age is just (2/3) the Hubble time at that epoch:

$$t_{\text{rec}} = (2/3)H_0^{-1}\Omega_M^{-1/2}\hat{a}_{\text{rec}}^{3/2} \sim \frac{2.6 h_{70}^{-1} \times 10^5 \text{ yr}}{\sqrt{\Omega_M}}. \quad (\text{E.14})$$

For t_{eq} , we can neglect neither the radiation nor the matter (at t_{eq}) so it is necessary to do a non-trivial integral:

$$t_{\text{eq}} = H_0^{-1} \frac{(1.68\Omega\gamma)^{3/2}}{\Omega_M^2} \int_0^1 \frac{x dx}{\sqrt{1+x}} \sim 5 \times 10^4 \text{ yr} \left(\frac{0.3}{\Omega_M h_{70}^2} \right)^2. \quad (\text{E.15})$$

Neglecting the matter would have given the correct order of magnitude and the correct dependence on $\Omega_M h_{70}^2$.

5.4 $t_0 \sim (2/3)H_0^{-1}[1 + \Omega_\Lambda/3 + \dots]$

5.12 Absorption by interstellar matter would not resolve Olbers' paradox because the matter would heat up until it reached a temperature at which it would radiate blackbody photons at the same rate as it absorbed starlight.

The modern calculation gives a flux per unit solid angle of

$$\begin{aligned} \frac{d\phi}{d\Omega} &= n_0 a_0^3 L \int \frac{r_1^2 d\chi_1}{4\pi d_L^2} = \frac{n_0 L}{4\pi} \int_0^{t_0} dt_1 \frac{a(t_1)}{a_0} \\ &= \frac{n_0 L}{4\pi} \int_0^{a_0} \frac{da}{a_0(\dot{a}/a)}, \end{aligned} \quad (\text{E.16})$$

where we have used $d_L = a_0 r_1 (1+z)$ and $d\chi = dt/a$. Using the Friedmann equation to evaluate \dot{a}/a we find

$$\frac{d\phi}{d\Omega} = (3/5) \frac{n_0 L t_0}{4\pi} \quad t_0 = (2/3)H_0^{-1} \quad \Omega_M = \Omega_T = 1 \quad (\text{E.17})$$

$$\frac{d\phi}{d\Omega} = (1/2) \frac{n_0 L t_0}{4\pi} \quad t_0 = H_0^{-1} \quad \Omega_M = \Omega_T = 0. \quad (\text{E.18})$$

In both cases the correct calculation adds only a numerical factor to the naive answer calculated assuming that the stars have been burning for a time t_0 . The fact that the factors are less than unity is due to the redshift.

In the inflationary model, we find

$$\rho(a_0) = \rho_V \left(\frac{a_{\text{inf}}}{a_0} \right)^4. \quad (\text{E.19})$$

We see that the energy density falls as a^{-4} as expected.

Chapter 6

6.1 It is important to factor out the physical parameters, e.g.:

$$\rho(T, \mu = 0) = \frac{g}{(2\pi)^3} \int \frac{4\pi p^3 dp}{\exp(p/T) \pm 1} = \frac{g T^4}{2\pi^2} \int_0^\infty \frac{x^3 dx}{\exp(x) \pm 1}. \quad (\text{E.20})$$

The integrals give only the numerical factors in Table 6.1.

For a neutral relativistic gas of electrons and positrons, the potential energy per particle is of order $\alpha \sim 1/137$ times the kinetic energy per particle. The gas should be ideal to a good approximation.

The particle–antiparticle asymmetry for a relativistic gas in thermal equilibrium is of order $(n - \bar{n})/n \sim \mu/T$.

6.4 The number of remaining interactions is

$$\int_{t_1}^\infty \Gamma(t) dt = \int_{a_1}^\infty \Gamma \frac{da}{a(\dot{a}/a)}. \quad (\text{E.21})$$

To quickly get the answer, it is a good idea to evaluate \dot{a}/a with a Friedmann equation normalized at a_1 , e.g.

$$\frac{\dot{a}}{a} = H_1 \left[\Omega_R(a_1) \left(\frac{a}{a_1} \right)^{-4} + \dots \right]^{1/2}. \quad (\text{E.22})$$

Substituting this into the integral, we find that the number of interactions is just $\Gamma(t_1)H_1^{-1}$ times a numerical factor of order unity. Since $\Gamma(t_1)H_1^{-1} \ll 1$, this proves the conjecture.

6.5 Numerically, the photon scattering rate (always dominated by Compton scattering) is equal to the expansion rate at $T \sim 0.236$ eV. The recombination rate is equal to the expansion rate at $T \sim 0.215$ eV. A fraction $\sim 3 \times 10^{-5}$ of the electrons remain free.

More realistic calculations including all the states of the hydrogen atom give a recombination time of $T \sim 0.26$ eV.

6.6 The last annihilations take place at $T \sim 10\text{ keV}$. A photon of $E = 510\text{ keV}$ needs about 10 collisions to reach a reasonably thermal energy of $E \sim 30\text{ keV}$. The time to do this, $\sim 10(n_e\sigma_T c)^{-1}$, is much less than the Hubble time at this epoch.

Chapter 7

6.8 The minimum mass is of order $(m_W/\alpha)\sqrt{m_W/m_{\text{pl}}}$.

Appendix F

Useful Formulas and Numbers

- Friedmann equation for the scale factor $a(t)$:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} + H_0^2(1 - \Omega_T)\hat{a}^{-2}, \quad (\text{F.1})$$

where Ω_T is the present-day total density divided by the critical density

$$\Omega_T = \frac{\rho(a_0)}{3H_0^2/8\pi G}, \quad (\text{F.2})$$

and where

$$a_0 = \frac{H_0^{-1}}{\sqrt{|1 - \Omega_T|}} \quad \hat{a}(t) = \frac{a(t)}{a_0}. \quad (\text{F.3})$$

- Friedmann equation for the present epoch:

$$\frac{\dot{a}}{a} = H_0 \left(\Omega_R \hat{a}^{-4} + \Omega_M \hat{a}^{-3} + \Omega_\Lambda + (1 - \Omega_T) \hat{a}^{-2} \right)^{1/2}, \quad (\text{F.4})$$

where Ω_R , Ω_M , and Ω_Λ are the present-day contributions of relativistic matter, non-relativistic matter, and vacuum energy:

$$\Omega_T = \Omega_R + \Omega_M + \Omega_\Lambda. \quad (\text{F.5})$$

- Friedmann equation during the radiation epoch:

$$\frac{\dot{a}}{a} = \left(\frac{8\pi G}{3} g_E(T) \frac{\pi^2}{30} T^4 \right)^{1/2} \sim 0.65 \text{ s}^{-1} \left(\frac{T}{1 \text{ MeV}} \right)^2 \left(\frac{g_E}{10} \right)^{1/2} \quad (\text{F.6})$$

where g_E is the effective number of spin states (Fig. 6.1).

- The radial coordinate χ of an object of redshift z :

$$\begin{aligned}\chi(z) &= \int_{a_0/(1+z)}^{a_0} \frac{da}{a^2 (\dot{a}/a)} \\ &= \frac{H_0^{-1}}{a_0} \int_{(1+z)^{-1}}^1 \frac{d\hat{a}}{\hat{a}^2 (\Omega_R \hat{a}^{-4} + \Omega_M \hat{a}^{-3} + \Omega_A + (1 - \Omega_T) \hat{a}^{-2})^{1/2}}.\end{aligned}\quad (\text{F.7})$$

- $\chi(z)$ for $z \ll 1$:

$$a_0 \chi_1(z) = a_0 r_1(z) = H_0^{-1} z \left[1 - \frac{1+q_0}{2} z + \dots \right]. \quad (\text{F.8})$$

- Luminosity and angular distances:

$$d_L(z) = a_0 r(z)(1+z) \quad (\text{F.9})$$

$$d_A(z) = a_0 r(z)/(1+z), \quad (\text{F.10})$$

where the radial coordinate r is

$$\begin{array}{lll} \sin \chi = \chi - \chi^3/6 + \dots & \Omega_T > 1 \\ r = \chi & \Omega_T = 1 \\ \sinh \chi = \chi + \chi^3/6 + \dots & \Omega_T < 1 \end{array} \quad (\text{F.11})$$

Table F.1 Selected physical constants, adapted from [17]

Quantity	Symbol	Value
Speed of light in vacuum	c	$2.99\,792\,458 \times 10^8 \text{ m s}^{-1}$
Planck constant	\hbar	$1.054\,571\,596(82) \times 10^{-34} \text{ J s}$
Conversion constant	$\hbar c$	$1.97\,326\,960\,2(77) \times 10^{-7} \text{ eV m}$
Conversion constant	$2\pi\hbar c$	$1.24 \times 10^3 \text{ eV nm}$
Fine structure constant	α	$1/137.035\,999\,76(50)$
Thomson cross-section	σ_T	$0.665\,245\,854(15) \times 10^{-28} \text{ m}^2$
Gravitational constant	$G_N (= G)$	$6.673(10) \times 10^{-11} \text{ m}^3 \text{kg}^{-1} \text{s}^{-2}$
Planck mass	$m_{\text{pl}} = \sqrt{\hbar c/G}$	$1.221\,0(9) \times 10^{19} \text{ GeV}/c^2$
Fermi coupling constant	$G_F/(\hbar c)^3$	$1.166\,39(1) \times 10^{-5} \text{ GeV}^{-2}$
Electron mass	m_e	$0.510\,998\,902(21) \text{ MeV}/c^2$
Proton mass	m_p	$938.271\,998(38) \text{ MeV}/c^2$
Neutron–proton Δm	$m_n - m_p$	$1.293\,318(9) \text{ MeV}/c^2$
Deuteron mass	m_d	$1875.612\,762(75) \text{ MeV}/c^2$
Boltzmann constant	k	$1.380\,650\,3(24) \times 10^{-23} \text{ J K}^{-1}$ $8.617\,342(15) \times 10^{-5} \text{ eV K}^{-1}$

Table F.2 Selected astrophysical and cosmological quantities, adapted from [17]. H_0 and all densities refer to the present epoch

Quantity	Symbol	Value
Astronomical unit	AU	$1.495\,978\,706\,60(20) \times 10^{11}$ m
Parsec	pc	$3.085\,677\,580\,7(4) \times 10^{16}$ m = 3.26... ly
Solar mass	M_\odot	$1.988\,9(30) \times 10^{30}$ kg = $1.189 \times 10^{57} m_p$
Solar luminosity	L_\odot	$3.846(8) \times 10^{26}$ W s $^{-1}$ 2.40×10^{45} eV s $^{-1}$
Solar equatorial radius	R_\odot	6.961×10^8 m
Hubble expansion rate	H_0	$70 h_{70}$ km s $^{-1}$ Mpc $^{-1}$ $100 h$ km s $^{-1}$ Mpc $^{-1}$ $h_{70} = 1.0 \pm 0.15$ $h = 0.7 \pm 0.1$
Hubble time	$t_H = H_0^{-1}$	$1.40 h_{70}^{-1} \times 10^{10}$ yr $4.41 h_{70}^{-1} \times 10^{17}$ s
Hubble distance	$d_H = c H_0^{-1}$	$4280 h_{70}^{-1}$ Mpc $1.32 h_{70}^{-1} \times 10^{26}$ m
Critical density	$\rho_c = 3H_0^2/8\pi G$	$0.92 h_{70}^2 10^{-26}$ kg m $^{-3}$ $5.16 h_{70}^2 10^9$ eV/c 2 m $^{-3}$ $1.36 h_{70}^2 10^{11} M_\odot$ Mpc $^{-3}$
CMB temperature	T_γ	2.725 ± 0.001 K
	kT_γ	$(2.348 \pm 0.002) \times 10^{-4}$ eV
CMB energy density	ρ_γ	$0.26038 (T_\gamma/2.725)^4$ eV cm $^{-3}$
	Ω_γ	$5.06 h_{70}^{-2} \times 10^{-5}$
CMB number density	n_γ	$410.50 (T_\gamma/2.725)^3$ cm $^{-3}$
Neutrinos (+antineutrinos) number density per species	$n_\nu = (3/11)n_\gamma$	$111.95 (T_\gamma/2.725)^3$ cm $^{-3}$
Matter-radiation equality		
Scale factor	$\hat{a}_{\text{eq}} = 1.68\Omega_\gamma/\Omega_M$	$0.85 \times 10^{-4}/(\Omega_M h_{70}^2)$
CMB temperature	kT_{eq}	$2.8 \Omega_M h_{70}^2$ eV
Recombination		
Redshift	$\hat{a}_{\text{rec}}^{-1} = z_{\text{rec}}$	1090.51 ± 0.95
CMB temperature	kT_{rec}	0.26 eV

References

1. E. Komatsu et al. (WMAP): *Astrophys. J. Supp.* **180**, 330 (2009)
2. P. Astier et al. (SNLS) *Astron. Astrophys.* **447**, 31–48 (2006)
3. W.M. Wood-Vasey et al. (ESSENCE): *Astrophys. J.* **666**, 694 (2007); R.A. Knopp et al.: *Astrophys. J.* **598**, 102 (2003); A.G. Riess et al. *Astrophys. J.* **659**, 98 (2007); and references therein
4. W.J. Percival et al.: *Mon. Not. R. Astron. Soc.* **381**, 1053–1066 (2007)
5. J. Mather et al. (COBE): *Astrophys. J.* **512**, 511 (1999)
6. Y. Fukuda et al. (KAMIOKA): *Phys. Rev. Lett.* **81**, 1562 (1998)
7. <http://www.eso.org>
8. <http://space.gsfc.nasa.gov/astro/cobe>
9. W.L. Freedman et al.: *Astrophys. J.* **553**, 47–72 (2001)
10. D. Tytler et al.: *Physica Scripta* **T85**, 12
11. R. Cen and J.P. Ostriker: *Astrophys. J.* **514**, 1 (1999)
12. F. Combes, and D. Pfenniger: *Astron. Astrophys. V*, 453 (1997)
13. de Paolis et al. *Astrophys. J.* **510L**, 103 (1999); de Paolis et al. *Phys. Rev. Lett.* **74**, 14 (1995)
14. A.G. Cohen, A. De Rujula and S.L. Glashow: *Astrophys. J.* **495**, 539 (1998)
15. K. Griest and J. Silk: *Nature* **343**, 26 (1990)
16. B.J. Carr: arXiv:astro-ph/0511743 (2005)
17. Particle Data Group: *Eur. Phys. J. C* **15**, 1 (2000)
18. P. Smith and J. Lewin: *Phys. Rep.* **187**, 203 (1990)
19. L.R. Wang, R.R. Caldwell, J.P. Ostriker and P.J. Steinhardt: *Astrophys. J.* **530**, 17 (2000)
20. R. Srianand, P. Petitjean and C. Ledoux: *Nature* **408**, 931 (2000)
21. J. R. Primack: arXiv:0902.2506 (2009)
22. M. Milgrom: *Astrophys. J.* **270**, 365 (1983)
23. R. Durrer and R. Maartens: arXiv:0811.4132 (2008)
24. T. Buchert: *Gen. Rel. Grav.* **40**, 467–527 (2008)
25. L. Guzzo et al.: *Nature* **451**, 541 (2008)
26. V. Acquaviva et al.: *Phys. Rev. D* **78**, 043514 (2008); see however E. Bertschinger and P. Zukin: *Phys. Rev. D* **78** 024015 (2008)
27. N.J. Cornish, D.N. Spergel and C.L. Bennett: arXiv:astro-ph/0202001
28. T. Dent, S. Stern and C. Wetterich: *Phys. Rev. D* **79**, 083533 (2009)
29. M. Tegmark, A. Aguirre, M.J. Rees and F. Wilczek: *Phys. Rev.D* **73**, 023505 (2006)
30. L. McAllister and E. Silverstein: *Gen. Rel. Grav.* **40**, 565 (2008)
31. H. Martel, P.R. Shapiro and S. Weinberg: *Astrophys. J.* **492**, 29 (1998)
32. N. Arkani-Hamed et al.: *Phys. Rev. Lett.* **85**, 4434 (2000)
33. J.D. Barrow and F.J. Tipler: *The Anthropic Cosmological Principle* (Oxford University Press, Oxford, 1988)
34. C.J. Hogan: *Rev. Mod. Phys.* **72**, 1149 (2000)
35. H. Oberhummer, A. Csótó and H. Schlattl: *Science* **289** (2000) 88–90.

36. J. Binney and M. Merrifield: *Galactic Astronomy* (Princeton University Press, Princeton, 1998)
37. A.J. Pickles: Publ. Astron. Soc. Pac. **110**, 863 (1998).
38. M.S. Bessel : Publ. Astron. Soc. Pac. **102**, 1181 (1990)
39. M. Fukugita et al.: Astron. J.**111**, 1748–1756 (1996)
40. <http://astro.estec.esa.nl/Hipparcos>
41. J. Binney and M. Merrifield: *Galactic Astronomy* (Princeton University Press, Princeton, 1998) p. 137
42. J.A. Holtzman et al.: Astron. J. **118**, 2262 (1999); <http://astronomy.nmsu.edu/holtz/archival/>
43. L. Girardi, M.A.T. Groenewegen, A. Weiss and M. Salaris: Mon. Not. R. Astron. Soc. **V**, 149 (1998); L. Girardi and M. Salaris: astro-ph/0007343
44. F. Pont: astro-ph/9812074
45. A. Udalski et al.: Acta Astronomica **49**, 201 (1999)
46. S. Perlmutter, personal communication
47. N. Suntzeff et al.: Astron. J. **117**, 1175 (1999)
48. A. Pastorello et al.: Mon. Not. R. Astron. Soc. **376** 1301 (2007)
49. B.K. Gibson et al.: Astrophys. J. **529**, 723–744 (2000)
50. L. Ferrarese and H. Ford: Space Science Reviews **116**, 523 (2005)
51. J. Binney and S. Tremaine: *Galactic Dynamics* (Princeton University Press, Princeton, 1987)
52. A.L. Kinney et al.: Astrophys. J. **467**, 38–60 (1996)
53. B.F. Madore et al.: Astrophys. J. **515**, 29–41 (1999)
54. M.R. Blanton et al.: Astrophys. J. **592**, 819–838 (2003)
55. R. Mandelbaum et al.: Mon. Not. R. Astron. Soc. **368**, 715–731 (2006)
56. M.A. Zwaan, M.J. Meyer, L. Staveley-Smith and R.L. Webster: Mon. Not. R. Astron. Soc. **359**, L30 (2005)
57. J.F. Navarro, C.S. Frenk, S.D.M. White: Astrophys. J. **462**, 563–575 (1996)
58. K. Begeman: Astron. Astrophys. **223**, 47 (1989)
59. S. Schindler: astro-ph/9811470; S. Schindler et al.: Astron. Astrophys. **317**, 646 (1997)
60. D. Clowe et al.: Astrophys. J. **648**, L109–L113 (2006)
61. <http://www.sdss.org/>
62. D. Hooper and E.A. Baltz: Annu. Rev. Nucl. Part. Sci. **58**, 293–314 (2008)
63. M.W. Goodman and E. Witten: Phys. Rev. D **31**, 3059 (1985).
64. K. Freese, J. Freedman and A. Gould: Phys. Rev. D **37**, 3388 (1988)
65. Z. Ahmed et al. (CDMS): Phys. Rev. Lett. **102**, 011301 (2009)
66. J. Angle et al. (XENON Coll.): Phys. Rev. Lett. **100**, 021303 (2008)
67. R. Bernabei et al. (DAMA/LIBRA) Eur. Phys. J. C **56**, 333–355 (2008)
68. J. Chang et al.: Nature **456**, 362–365 (2008)
69. M. Mori et al.: Phys. Lett. B **270**, 89 (1991)
70. M. Mori, et al.: Phys. Lett. B **289**, 463 (1992)
71. F. Halzen, T. Stelzer and M. Kamionkowski: Phys. Rev. **D45** 4439 (1992)
72. M. Spurio et al. (ANTARES coll.): arXiv:0904.3836 (2009)
73. R. Abbasi et al. (IceCube coll.): arXiv:0902.2460 (2009)
74. H. Cheng: Phys. Rep. **158**, 1 (1988)
75. P. Sikivie: Phys. Rev. Lett. **51**, 1415 (1983)
76. W. Wuensch et al.: Phys. Rev. D **40**, 3153 (1988)
77. C. Hagmann et al.: Phys. Rev. D **42**, 1297 (1990)
78. C. Hagmann, et al.: Phys. Rev. Lett. **80**, 2043 (1998)
79. B. Carr: Comments Astrophys. **14**, 257 (1990)
80. B.D. Fields, K. Freese and D. S. Graff: Astrophys. J. **534**, 265 (2000)
81. B. Paczyński: Astrophys. J. **304**, 1 (1986)
82. K. Griest: Astrophys. J. **366**, 412 (1991)
83. C. Hamadache et al. (EROS-2): Astron. Astrophys. **454**, 185 (2006)
84. C. Alcock et al. (MACHO): Astrophys. J. Lett. **499**, L9 (1998)

85. C. Alcock et al. (MACHO): *Astrophys. J.* **542**, 281 (2000)
86. P. Tisserand et al. (EROS-2): *Astron. Astrophys.* **469**, 387 (2007); see also L. Wyrzykowski et al. (OGLE), *Mon. Not. R. Astron. Soc.* **397**, 1228 (2009)
87. C. Alcock, et al. (MACHO coll.) *Astrophys. J.* **550**, L169 (2001)
88. J. Yoo, J. Chanamé and A. Gould: *Astrophys. J.* **601**, 311 (2004); see however D.P. Quinn et al.: *Mon. Not. R. Astron. Soc.* **396**, L11 (2009)
89. P. Arras and I. Wasserman: *Mon. Not. R. Astron. Soc.* **306**, 257 (1999).
90. G.C. Lacev and J.P. Ostriker: *Astrophys. J.* **299**, 633 (1985)
91. D. P. Bennett: *Astrophys. J.* **633**, 906 (2005)
92. S. Calchi Novati et al. (AGAPE coll.) *Astron. Astrophys.* **443**, 911 (2005); J.T. de Jong et al. (MEGA coll.): *Astron. Astrophys.* **446**, 855 (2006)
93. A. Vidal-Madjar et al.: *Astrophys. J. Lett.* **538**, L77 (2000)
94. P. Salati et al.: *Dark Matter in Astro- and Particle Physics: (DARK '96)* editors, H.V. Klapdor-Kleingrothaus et al. (World Scientific, 1997)
95. R.R. Rafikov and B.T. Draine:astro-ph/0006320
96. H. Tadros, S.J. Warren and P.C. Hewett: astro-ph/0003422
97. B. Schmidt et al: *Astrophys. J.* **432** 42 (1994); P. Nugent et al.: *Astrophys. J.* **645**, 841 (2006)
98. G.W. Pratt et al.: *Astron. Astrophys.* **498**, 361 (2009)
99. S.W. Allen et al.: *MNRAS*, 383, 879 (2008)
100. A. Vikhlinin et al. *Astrophys. J.* **692**, 1060–1074 (2009)
101. F.C. Durret et al.: *Astron. Astrophys.* **356**, 815 (2000)
102. A. Bosma: *Astron. J.* **86**, 12 (1981)
103. E.F. Guinan et al.: *Astrophys. J. Lett.* **509**, L21 (1998)
104. M. Nauenberg and V. Weisskopf: *Am. J. Phys.* **46**(1), 23 (1978).
105. M. Arnaud et al.: *Astron. Astrophys.* **365**, L67–L73 (2001)
106. J.D. Jackson: *Classical Electrodynamics* (Wiley, New York, 1962)
107. J.W. Bailey et al.: *Phys. Lett.* **28B**, 287 (1968)
108. R.F.C. Vessot et al.: *Phys. Rev. Lett.* **45**, 2081–2084 (1980)
109. A. Riess et al.: *Astron. J.* **114**, 722 (1997)
110. P. Schneider, J. Ehlers, and E.E. Falco: *Gravitational Lenses* (Springer-Verlag, 1992)
111. R. Gavazzi et al.: *Astrophys. J.* **667**, 176–190 (2007)
112. J.E. Ovaldsen et al.: *Astron. Astrophys.* **402**, 891–904 (2003)
113. L. Fu et al.: *Astron. Astrophys.* **479**, 9–25 (2008)
114. B. Chaboyer: *Phys. Rep.* **307**, 23 (1998)
115. S. Cassisi, V. Castellani, S. degl'Innocenti, M. Salaris and A. Weiss: *Astron. Astrophys. Suppl.* **134**, 103 (1999)
116. L.M. Krauss and B. Chaboyer: *Science* **479**, 9–25 (2003)
117. B.M.S. Hansen et al.: *Astrophys. J.* **574**, L155–L158 (2002)
118. S. Perlmutter et al.: *Astrophys. J.* **517**, 565 (1999)
119. B. Schmidt et al.: *Astrophys. J.* **507**, 46 (1998)
120. D.J. Eisenstein et al.: *Astrophys. J.* **633**, 560–574 (2005)
121. T.J. Bronder et al.: *Astron. Astrophys.* **477**, 717–734 (2008)
122. Aguirre, A.: *Astrophys. J.* **525**, 583 (1999)
123. D.J. Eisenstein, H.-J. Seo and M. White: *Astrophys. J.* **664**, 660 (2007)
124. P. de Bernardis et al.: *Nature* **404**, 995 (2000)
125. S. Hanany et al.: *Astrophys. J. Lett.* **545**, L5 (2000)
126. A.R. Liddle and D.H. Lyth: *Cosmological Inflation and Large-Scale Structure* (Cambridge University Press, Cambridge, 2000)
127. A.H. Guth: *Phys. Rev. D* **23**, 347 (1981)
128. A.D. Linde: *Phys. Lett.* **108B**, 389 (1982); A. Albrecht and P.J. Steinhardt: *Phys. Rev. Lett.* **48**, 1220 (1982)
129. J.J. Sakurai: *Advanced Quantum Mechanics* (Adison-Wesley, Reading, 1967)
130. G. Steigman: *Annu. Rev. Nucl. Part. Sci.* **57**, 463–491 (2007)

131. E. Browne and R.B. Firestone: *Table of Radioactive Isotopes* (Wiley, New York, 1986)
132. E. Vangioni-Flam, A. Coc and M. Casse: *Astron. Astrophys.* **360**, 15 (2000)
133. S. Burles and D. Tytler: *Stellar Evolution, Stellar Explosions and Galactic Chemical Evolution, Second Oak Ridge Symposium on Atomic and Nuclear Astrophysics* (IOP, Bristol, 1998) and astro-ph/9803071
134. M. Trodden: *Rev. Mod. Phys.* **71** 1463 (1999)
135. E.W. Kolb and M.S. Turner: *The Early Universe* (Addison-Wesley, Redwood City, 1990) p. 160
136. <http://www.mpa-garching.mpg.de/galform/virgo/millennium/>
137. <http://www.projet-horizon.fr>
138. P. Coles and F. Lucchin: *Cosmology: The Origin and Evolution of Cosmic Structure* (Wiley, Chichester 2002)
139. E.F. Bunn and M. White: *Astron. J.* **480**, 6–21 (1997); A.R. Liddle et al.: *Phys. Rev. D* **74**, 083512 (2006)
140. D. Eisenstein and W. Hu: *Astrophys. J.* **496**, 605–614 (1998)
141. W.J. Percival et al.: *Astrophys. J.* **657** 645–663 (2007)
142. U. Seljak et al.: *Phys. Rev. D* **71** 043511 (2005)
143. R.J. Croft et al.: *Astrophys. J.* **581**, 20–52 (2002)
144. S.M. Carroll, W.H. Press, E.L. Turner: *Ann. Rev. Astr. Astrphys.* **30**, 499 (1992)
145. S. Colombi, M.J. Chodorowski and R. Teyssier: *Mon. Not. R. Astron. Soc.* **375**, 348–370 (2007)
146. J.M. Bardeen, P.J. Steinhardt and M.S. Turner: *Phys. Rev.* **28**, 679 (1983); A.H. Guth and S.-Y. Pi: *Phys. Rev. Lett.* **49**, 1110 (1982); A.A. Starobinskii: *Phys. Lett.* **117B**, 175 (1982); S.W. Hawking: *Phys. Lett.* **115B**, 295 (1982)
147. M. Tegmark and M.J. Rees: *Astrophys. J.* **499**, 526 (1998)
148. C. Bennett et al. (COBE): *Astrophys. J. Lett.* **464**, L1 (1996)
149. C.H. Lineweaver et al. (COBE): *Astrophys. J.* **470**, 38L (1996)
150. M.R. Nolta et al. (WMAP): *Astrophys. J. Suppl.* **180**, 296, 305 (2009)
151. C.L. Reichardt et al. (ACBAR): *Astrophys. J.* **694**, 1200–1219 (2009)
152. U. Seljak and M. Zaldarriaga (CMBFAST): *Astrophys. J.* **469**, 437–444 (1996); A. Lewis, A. Challinor and A. Lasenby (CAMB): *Astrophys. J.* **538**, 473–476 (2000) [<http://camb.info/>]
153. W. Hu and S. Dodelson: *Ann. Rev. Astr. Astrphys.* **2002**, 1–51 (2002)
154. W. Hu, N. Sugiyama and J. Silk: *Nature* **386**, 37 (1997)
155. T. Giannantonio et al.: *Phys. Rev. D* **77**, 123520 (2008)
156. W. Hu et al.: *Astrophys. J.* **549**, 669–680 (2001)
157. C. Pryke et al. *Astrophys. J.* **692**, 1247–1270 (2009)
158. M. Kamionkowski, A. Kosowsky and A. Stebbins: *Phys. Rev. D* **55**, 7368 (1997); M. Zaldarriaga and U. Seljak, *Phys. Rev. D* **55**, 1830 (1997)
159. A. Riazuelo, N. Deruelle and P. Peter: *Phys. Rev. D* **61**, 123504 (2000)
160. P.J.E. Peebles: *Principles of Physical Cosmology* (Princeton University Press, Princeton, 1993) pp. 290–292
161. M. White and W. Hu: *Astron. Astrophys.* **321**, 8 (1997)
162. F. Halzen and A.D. Martin: *Quarks and Leptons* (Wiley, New York, 1984)
163. W. Greiner and B. Müller: *Gauge Theory of Weak Interactions* (Springer, Berlin, 2000)
164. U. Mosel: *Fields, Symmetries, and Quarks* (Springer, Berlin, 1999)
165. Q. Ho-Kim and X.-Y. Pham: *Elementary Particles and Their Interactions* (Springer, Berlin, 1998)

Index

A

Absorption (interstellar), 45, 310

ACBAR, 266

Acoustic oscillations, 280

Adiabatic density fluctuations, 278

Adiabatic fluctuations, 260

Age of the universe, 167

Angular distance, 115

Anthropic selection, 39, 190, 264

Antimatter, 11, 228

Axion, 38, 68

B

BAO, 2, 175, 256, 273

Baryogenesis, 227

Baryon-to-photon ratio, η , 2, 78

and baryogenesis, 227, 236

and nucleosynthesis, 217, 222, 234, 237

Baryonic dark matter, 68, 278

Bias parameter, 257

Big bang, 8

Black holes, 12, 13, 52, 68

Boltzmann equation, 26, 204

Boomerang, 180

Bose–Einstein statistics, 200, 208

Bottom-up structure formation, 247, 280

Bremssstrahlung, 61, 77, 87, 88, 208

Bullet cluster, 62, 63

C

C and CP violation, 228

Cepheid variable stars, 48, 52, 56, 74, 81, 83

Chandra, 60

Chemical equilibrium, 197, 209, 210, 215, 219

Chemical potential, 200, 208, 227, 230, 232

Classical horizon, 186

Clump giants, 45, 47

Clusters of galaxies, 7, 60, 72, 82

and filaments, 243

and hot dark matter, 280

baryon fraction, 62, 78

formation, 242

gravitational lensing by, 76, 122

number density evolution, 60, 79

SZ effect, 77

X-ray emission, 87

Co-moving coordinates, 93

COBE, 4, 13, 78, 247

Cold dark matter (CDM), 11, 64, 225

and CMB anisotropies, 268

evolution of fluctuations, 242

Color-magnitude diagram, 45–47, 169

Comoving coordinates, 100, 101

Correlation function, 258, 288

Cosmic background radiation (CMB), 13, 78,

80, 196

and causality, 181, 183

and SZ effect, 77

anisotropies, 174, 264, 266, 267, 269, 270,
272, 274, 288

polarization, 274

temperature evolution, 24, 198

Cosmic variance, 290

Cosmological constant, 16

Cosmological principle, 9

Critical density, 9

Curvature epoch, 30

Curved space versus curved space–time, 142

D

Dark matter

in clusters of galaxies, 61

in galaxies, 3, 5, 53, 57

Deceleration parameter q_0 , 29, 112, 151,

180

and expanding photosphere, 194

and luminosity distance, 114

and number counts, 194

and SZ effect, 194
 and type Ia supernovae, 170
 correlation with distance, 111, 113, 171
 Deuterium, 221, 224, 237
 Distance ladder, 47, 74

E

Effective number of spin states, 202, 203
 Einstein equations, 157, 158, 163
 Einstein tensor, 157
 Electroweak interactions, 305
 Empty universe, 142
 Energy conservation, 137, 146–148
 Energy-momentum tensor, 143, 162
 Entropy, 198, 201, 203, 213, 229, 233
 Equation of state, 145, 165
 Event horizon, 186, 195

F

Faber–Jackson relation, 54
 Fermi constant, 305
 Fermi–Dirac statistics, 200, 208
 2dFGRS, 178, 258
 Filter systems, 310
 Fine-tuning, 39, 187, 188
 Free-streaming, 279
 Freely falling coordinates, 92, 98, 138, 141
 Freeze-out, 198, 232, 233
 and entropy production, 229, 230
 and free energy, 198
 electron–positron, 211, 212
 neutrino, 215, 238
 neutron, 219
 nuclear, 221
 positron, 214, 235
 WIMPs, 226
 Friedmann equation, 20–22, 29, 148, 150, 158, 161
 Fundamental Plane, 54

G

Galactic halos, 3, 53, 57, 167
 Galactic rotation curves, 3, 53, 57, 83
 Galaxy number counts, 194
 Geodesics, 98, 110, 116, 119
 Globular clusters, 168, 169
 Grand unified theories (GUTs), 35, 187, 188, 228, 235
 Gravitational collapse
 dissipationless, 242
 dissipative, 242
 Gravitational lensing, 61, 119, 122, 124–128
 galaxy-galaxy, 58
 microlensing, 69

on large-scale structure, 126, 257, 278
 time delay, 76, 126
 weak, 57
 Gravitational redshift, 120
 Gravitational waves, 260, 274

H

Helium, 11, 41, 42, 80, 219–221
 High-z Supernova Search, 171
 Hipparcos, 43, 46
 Horizon, 181, 195
 sound, 175, 273
 Horizon problem, 181
 Hot dark matter, 278
 Hubble constant H_0 , 6, 18, 73, 273
 Hubble diagram, 7, 173, 179
 Hubble distance, 8
 Hubble Key project, 7, 74
 Hubble law, 6, 8, 112, 113
 Hubble radius exit/entry, 260
 Hubble Space Telescope, HST, 49, 74, 124
 Hubble time, 9, 167, 170

I

Inertial frames, 95, 99
 Inflation, 170, 183, 188, 189
 and \mathcal{Q}_T , 186, 187
 and causality, 182, 184
 and CMB anisotropies, 278
 and gravitational waves, 277
 and homogeneity, 186
 and particle physics, 190
 and primordial fluctuations, 260, 264
 beginning of, 185, 199
 duration of, 184
 gravitational waves, 36, 260
 reheating, 189
 Integrated Sachs–Wolfe (ISW) effect
 early time, 271, 280
 late time, 271
 Isocurvature fluctuations, 260, 278, 279

K

Kinetic equilibrium, 197, 235

L

Large-scale structure, 62
 Last-scattering surface, LastSS, 266, 267
 Lepton number asymmetry, 228
 Liouville equation, 26, 204, 233
 Lorentz transformations, 95, 131, 134, 295
 Luminosity distance, 113, 170
 Lyman- α forest, 52, 223, 257

M

MACHOs, 11, 68
 Magellanic Clouds, 45, 47, 69, 72, 74, 83
 Magnetic monopoles, 35
 Magnitudes, 307
 Main sequence, 42, 45, 167
 Mass-to-light ratio, 4, 54, 58, 61, 82
 Matter epoch, 30
 Matter-radiation equality, a_{eq} , 25
 Maxima, 180
 Metric, 91
 Molecular gas as dark matter, 71
 Multiverse, 38

N

Natural units, 8, 297
 Neutrino oscillations, 16
 Neutrinos, 12, 14, 50, 67, 78, 215
 and a_{eq} , 25
 as dark matter, 12, 279
 density, 24
 masses, 15, 280
 number of species, 238
 wrong helicity, 235
 Neutron stars, 50, 68
 Neutrons, 219, 237
 NFW profile, 57

O

Olbers' paradox, 195
 Ω_b , 2, 10, 78, 273
 Ω_{CDM} , 2, 12, 273
 Ω_A , 2, 16, 39, 79, 170, 273
 from age of universe, 167
 Ω_M , 13, 39, 79, 170
 and structure formation, 32, 79
 from age of universe, 167
 from cluster M/L , 82
 from cluster baryon fraction, 78, 82
 Ω problem, 186
 Ω_R , 25
 Ω_T , 2, 10, 39, 79, 273
 evolution, 31, 186
 Ω_{vis} , 10

P

Parallax, 43, 45
 Particle-antiparticle asymmetry, 213, 215, 227
 Peculiar velocity, 6, 259
 Phase transitions, 17, 188, 203
 Phase-space distribution, 26–28, 200, 204, 233, 291
 Planck mass, 17, 301, 305
 Polarization, 289, 291

Power spectrum, 248

Pressure, 145, 162, 201
 negative, 146
 Primordial nucleosynthesis, 11, 78, 216
 Proper distance, 155
 Proper time, 91

Q

QUAD, 275
 Quasi-stellar objects (QSOs), quasars, 51, 76, 128, 221, 223
 Quintessence, 17

R

Radiation epoch, 30
 Reaction rate, 197, 205
 Recombination, 14, 33, 180, 233, 267
 Redshift, 7, 24, 26
 Redshift distortions, 36, 258
 Reduced scale factor $\hat{a}(t)$, 19
 Reionization, 191, 273, 291
 Relic particle densities, 198, 213, 225
 Ricci tensor, 163
 Riemann tensor, 154, 163
 Robertson-Walker metric, 105, 109

S

Sachs–Wolfe effect, 254, 256, 262, 268, 272, 287, 290
 Saha equation, 221, 234, 237
 Scalar field, 17, 152, 188
 Scale factor $a(t)$, 18, 22, 23, 30
 Scale-invariant fluctuation spectrum, 246, 260
 Schwarzschild metric, 119
 SDSS, 56, 59, 64, 124, 174, 178, 179, 256, 258, 271
 σ_8 , 257
 Silk damping, 279
 SNLS, 52, 173, 174
 Sound horizon, 175, 273
 Sound speed, 269, 281
 Spherical collapse, 32, 242
 Standard candles, 47, 50, 51, 74, 170
 Standard model of particle physics, 301
 Steady-state universe, 170
 Stellar evolution, 41, 167
 Sunyaev–Zel'dovich (SZ) effect, 77, 194
 Supernova Cosmology Project, 49, 171
 Supernovae, 2, 49, 74, 273
 expanding photosphere, 76, 194
 Superstring landscape, 38, 185
 Supersymmetry, 12, 37, 64, 67, 225, 306

T

- t_0 , 18
Thermal equilibrium, 197, 200, 206, 208
Tilt, 254, 263
Top-down structure formation, 280
Topological defects, 278
Tully–Fisher relation, 54, 56, 74

V

- Vacuum epoch, 30
Virial theorem, 82
Virialization, 242, 246, 264

W

- w , 17
Weak interactions, 305
White dwarfs, 50, 68, 168
WIMPs, 12, 225
detection, 64
WMAP, 2, 174, 191, 266, 273, 276

X

- X-rays, 61, 62, 77, 82, 87, 88
XMM, 88