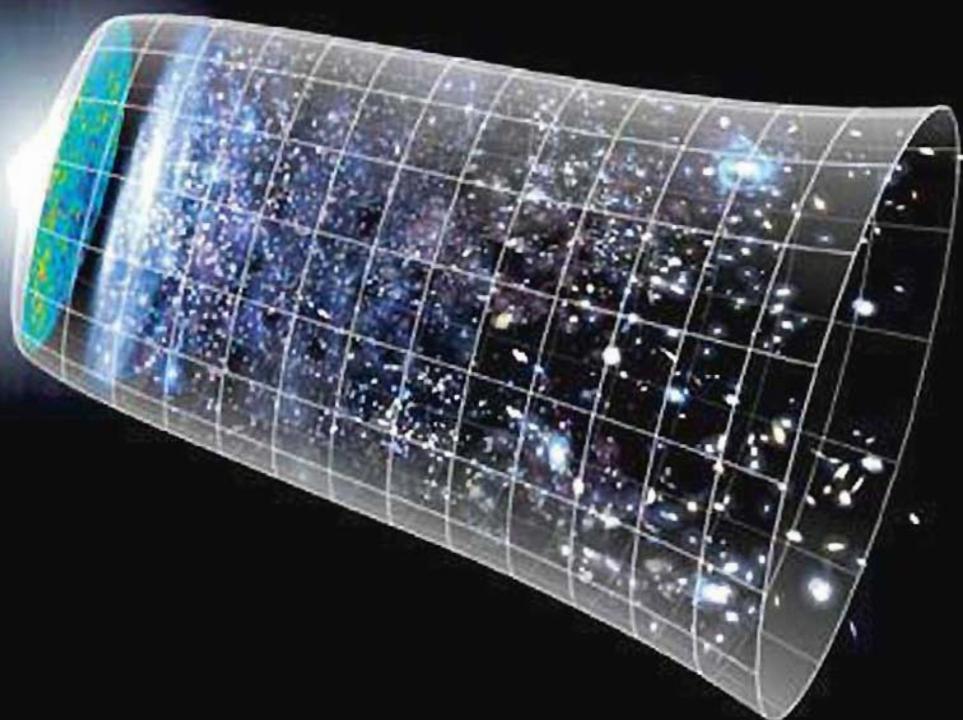
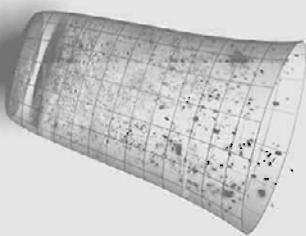


DMITRY S GORBUNOV • VALERY A RUBAKOV



INTRODUCTION TO
THE THEORY OF THE
EARLY UNIVERSE

Cosmological Perturbations
and Inflationary Theory



INTRODUCTION TO
THE THEORY OF THE
EARLY UNIVERSE

*Cosmological Perturbations
and Inflationary Theory*

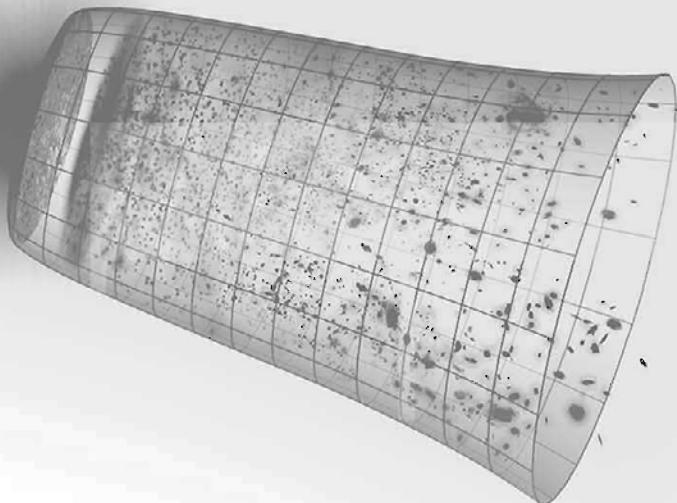
This page is intentionally left blank

DMITRY S GORBUNOV

Institute for Nuclear Research of the Russian Academy of Sciences

VALERY A RUBAKOV

Institute for Nuclear Research of the Russian Academy of Sciences
& Moscow State University



INTRODUCTION TO
THE THEORY OF THE
EARLY UNIVERSE

Cosmological Perturbations
and Inflationary Theory

 World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in Publication Data

A catalogue record for this book is available from the British Library.

INTRODUCTION TO THE THEORY OF THE EARLY UNIVERSE

Cosmological Perturbations and Inflationary Theory

Copyright © 2011 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN-13 978-981-4322-22-5

ISBN-10 981-4322-22-9

ISBN-13 978-981-4343-78-7 (pbk)

ISBN-10 981-4343-78-1 (pbk)

Typeset by Stallion Press
Email: enquiries@stallionpress.com

Printed in Singapore.

To Olesya and Elvira

This page is intentionally left blank

Preface

Modern cosmology is deeply connected to microphysics which studies elementary particles and their interactions at the most fundamental level. This book is written precisely from this perspective. It accompanies the book “Introduction to the Theory of the Early Universe: Hot Big Bang Theory” containing the material on the homogeneous and isotropic Universe at the hot stage of its evolution and at subsequent stages. This book presents the theory of cosmological perturbations (inhomogeneities in the Universe), inflationary theory and theory of post-inflationary reheating.

This book grew from the lecture course which is being taught for a number of years at the Department of Quantum Statistics and Field Theory, and now at the Department of Particle Physics and Cosmology of the Physics Faculty of the Lomonosov Moscow State University. This course is aimed at undergraduate students specializing in theoretical physics. We decided, however, to add a number of more advanced Chapters and Sections which we mark by asterisks.

For reading the main Chapters of this book, knowledge of material taught in general physics courses is in principle sufficient. So, the main Chapters must be understandable by undergraduate students. The necessary material on General Relativity and theory of random fields is collected in Appendices which, of course, do not pretend to give comprehensive presentation of these areas of physics. However, some parts labeled by asterisks make use of nonequilibrium statistical mechanics and specific methods of quantum field theory, so basic knowledge of these areas is required for reading these parts. Some useful concepts and results are given in Chapter 8 and Appendices D and E.

We tried to make Chapters 11, 12, 13, 14 and 15, where we discuss inflation and post-inflationary reheating, as self-contained as possible. We hope that the reader interested in the inflationary theory will be able to read these Chapters without constantly consulting the rest of the book.

In appropriate places of this book we use the notations and results presented in the accompanying book “Introduction to the Theory of the Early Universe: Hot Big Bang Theory”. References to that book begin with I. As an example, I.3.1 means Section 3.1 of that book, and (I.3.13) means equation (3.13) there.

Literature on cosmology is huge, and presenting systematic and comprehensive bibliography would be way out of the scope of this book. To orient the reader, in the end of this book we give a list of monographs and reviews where the issues we

touch upon are considered in detail. Certainly, this list is by no means complete. We refer to original literature in those places where we present concrete results without detailed derivation.

Both observational cosmology and experimental particle physics develop very fast. Observational and experimental data we quote, the results of their compilations and fits (values of the cosmological parameters, limits on masses and couplings of hypothetical particles, etc.) will most probably get more precise even before this book is published. This drawback can be corrected, e.g., by using the regularly updated material of Particle Data Group at <http://pdg.lbl.gov/>

We would like to thank our colleagues from the Institute for Nuclear Research of the Russian Academy of Sciences F. L. Bezrukov, S. V. Demidov, S. L. Dubovsky, A. A. Khmelnitsky, V. A. Kuzmin, D. G. Levkov, M. V. Libanov, E. Y. Nugaev, G. I. Rubtsov, D. V. Semikoz, I. I. Tkachev and S. V. Troitsky for participation in the preparation of the lecture course and numerous helpful discussions and comments. We are deeply indebted to G. S. Bisnovaty-Kogan, V. N. Lukash, V. F. Mukhanov, I. D. Novikov, K. A. Postnov, M. V. Sazhin, I. L. Shapiro, M. E. Shaposhnikov, Yu. V. Shtanov, R. A. Sunyaev, A. V. Tikhonov and O. V. Verkhodanov for useful comments and criticism on the preliminary version of this book.

Contents

<i>Preface</i>	vii
1. Jeans Instability in Newtonian Gravity	1
1.1 Jeans Instability in Static Background	2
1.2 Development of Instability in Expanding Universe	6
1.3 Linear Sizes of Perturbations and Masses of Objects	11
2. Cosmological Perturbations in General Relativity. Equations of Linearized Theory	13
2.1 Background Metric	14
2.1.1 Metric in conformal time	14
2.1.2 Cosmological parameters and conformal times of various epochs	15
2.2 Generalities	20
2.2.1 Gauge $h_{0i} = 0$	20
2.2.2 Linearized energy-momentum tensor of ideal fluid	22
2.2.3 Helicity decomposition	24
2.3 Equations in Helicity Sectors	27
2.3.1 Tensor perturbations	27
2.3.2 Vector perturbations	29
2.3.3 Scalar perturbations: conformal Newtonian gauge	30
2.4 Regimes of Evolution	33
2.5 *Scalar Field Condensate as Dark Matter	37
3. Evolution of Vector and Tensor Perturbations	47
3.1 Vector Modes	47
3.2 Tensor Modes: Relic Gravity Waves	48
3.2.1 Superhorizon perturbations: constant and decaying modes	48
3.2.2 Subhorizon perturbations. Matching to constant mode	50

4.	Scalar Perturbations: Single-Component Fluids	53
4.1	Master Equation	53
4.2	Relativistic Matter	55
4.3	Non-relativistic Matter	57
4.4	Matter Perturbations at Λ Domination	58
5.	Primordial Perturbations in Real Universe	61
5.1	Adiabatic and Isocurvature Modes	62
5.2	Adiabatic Mode in Superhorizon Regime	65
5.3	Initial Data for Isocurvature Modes	71
5.4	Primordial Spectra: Results from Observations	72
5.5	Evolution of Adiabatic Perturbations: A Preview	77
6.	Scalar Perturbations before Recombination	81
6.1	Adiabatic Modes of Large Wavelengths	83
6.2	Adiabatic Modes Entering the Sound Horizon at Radiation Domination	83
6.2.1	Dark matter perturbations	85
6.2.2	Perturbations in baryon-photon component	91
6.3	Adiabatic Perturbations of Intermediate Momenta	95
6.4	CDM Isocurvature Perturbations	97
6.5	Baryon Isocurvature Perturbations	101
7.	Structure Formation	103
7.1	Matter Perturbations after Recombination: Linear Regime	104
7.1.1	Power spectrum $P(k, z)$	104
7.1.2	Baryon acoustic oscillations	114
7.2	Beginning of Non-linear Regime	121
7.2.1	Preliminaries	121
7.2.2	Mass distribution of structures	122
8.	*Beyond Ideal Fluid Approximation	133
8.1	Distribution Functions and Boltzmann Equation in Curved Space-Time	133
8.2	General Equations for Scalar Perturbations	145
8.3	Warm Dark Matter	149
8.3.1	Suppressed growth of perturbations	149
8.3.2	Bound on WDM particle mass from phase space density	162
8.4	Neutrino Free Streaming	165
8.4.1	Early evolution: relativistic neutrinos	167
8.4.2	Non-relativistic neutrinos	174

8.4.3	Neutrino effect on matter perturbations. Cosmological bound on neutrino masses	177
8.5	Photons and Baryons at Recombination Epoch	185
8.5.1	Thickness of last scattering sphere	186
8.5.2	Silk damping	189
9.	Temperature of Cosmic Microwave Background	197
9.1	CMB Temperature Anisotropy	197
9.2	Temperature Anisotropy in Instant Photon Decoupling Approximation	201
9.2.1	General formalism	201
9.2.2	Large angular scales	208
9.2.3	Intermediate angular scales	218
9.3	Small Angular Scales	228
9.4	Anisotropy Spectrum and Cosmological Parameters	231
9.5	Temperature Anisotropy Generated by Isocurvature Modes	235
10.	*CMB Polarization	239
10.1	Sources of CMB Polarization	239
10.2	Polarization Tensor. E - and B -modes	242
10.3	Generation of CMB Polarization	250
10.3.1	Generalities	251
10.3.2	Scalar perturbations	256
10.3.3	Tensor perturbations	266
10.4	Discussion	276
11.	Drawbacks of the Hot Big Bang Theory. Inflation as Possible Way Out	279
11.1	Drawbacks of the Hot Big Bang Theory	279
11.1.1	Horizon problem	281
11.1.2	Flatness problem	283
11.1.3	Entropy problem	284
11.1.4	Primordial perturbation problem	285
11.2	Inflation: The Basic Idea	285
12.	Inflation in Slow Roll Regime	291
12.1	Slow Roll Conditions	291
12.2	Inflationary Models	296
12.2.1	Large field inflation (“chaotic”)	297
12.2.2	Small field inflation (“new”)	302
12.2.3	Hybrid inflation	305
12.2.4	Concluding remarks	309

13.	Generation of Cosmological Perturbations at Inflation	313
13.1	Simplified Analysis: Inflaton Fluctuations	314
13.1.1	Vacuum fluctuations in flat space-time	314
13.1.2	Generation of inflaton perturbations	316
13.1.3	Primordial scalar perturbations	323
13.2	Scalar Perturbations in Full Linear Theory	326
13.3	Tensor Perturbations	331
13.4	Amplitudes and Tilts of Power Spectra	332
13.4.1	Scalar amplitude: flatness of inflaton potential	332
13.4.2	Tensor amplitude	334
13.4.3	Spectral tilts	339
13.5	Discussion	343
14.	*Further Aspects of Inflationary Theory	347
14.1	Eternal Inflation	347
14.2	Generation of Scalar Perturbations by Curvaton Mechanism	349
14.2.1	Non-Gaussian adiabatic perturbations	349
14.2.2	Isocurvature modes	354
14.3	Light Scalar Field in Inflating Universe	356
14.3.1	Light field without self-interaction	356
14.3.2	Model with quartic potential	362
14.4	Axion as Dark Matter Candidate: CDM Isocurvature Mode	364
15.	*Preheating after Inflation	371
15.1	Inflaton Decay in Weakly Coupled Models	371
15.2	Inflaton Decay in a Model with Quadratic Potential	375
15.2.1	Decay of large amplitude oscillations	377
15.2.2	Intermediate coupling	382
15.3	Peculiarities of ϕ^4	385
15.4	Creation of Heavy Fermions	390
15.5	Physics Applications	393
15.5.1	Generation of baryon asymmetry at reheating	393
15.5.2	Non-thermal phase transitions	396
15.5.3	Small wavelength gravity waves	397
16.	*Bouncing Universe	401
17.	Color Pages	409
Appendix A	Exact Solutions for Gravitating Matter	415
A.1	Gravitational Collapse in Newtonian Theory	415
A.2	Spherical Collapse of Pressureless Fluid in General Relativity	418

Appendix B Derivation of Linearized Einstein Equations	425
Appendix C Gaussian Random Variables and Gaussian Random Fields	433
C.1 Properties of Gaussian Random Variables	433
C.2 Gaussian Random Fields	438
Appendix D Fermions in Gravitational Fields	443
D.1 Lorentz Group as Gauge Group	443
D.2 Fermion Action and Dirac Equation	447
Appendix E Particle Creation in Background Fields. Method of Bogoliubov Transformations	451
E.1 Bosons	451
E.2 Fermions	457
Appendix F Some Special Functions and Their Properties	463
F.1 Spherical Bessel Functions $j_l(x)$ of Integer Order	463
F.2 Legendre Polynomials $P_n(x)$ and Spherical Harmonics $Y_{lm}(\mathbf{n})$	469
<i>Books and Reviews</i>	475
<i>Bibliography</i>	479
<i>Index</i>	487

Chapter 1

Jeans Instability in Newtonian Gravity

In this and following Chapters we study the evolution of matter density inhomogeneities and metric perturbations in the expanding Universe. Several inter-related aspects are of interest here. Let us mention two of them. The first is the growth of density perturbations which in the end leads to formation of structures (galaxies, clusters of galaxies, etc.). The second is the temperature anisotropy and polarization of the Cosmic Microwave Background (CMB). By comparing theory with observations, one investigates the key cosmological issues: (i) what were the properties of the primordial cosmological perturbations, which most probably existed before the Hot Big Bang epoch; (ii) what are the properties of the medium in the Universe (usual matter, dark matter, dark energy). Primordial perturbations serve as initial data for the further evolution; if not for primordial perturbations, structures in the Universe would never appear. In its turn, the evolution of perturbations depends on the properties of the cosmic medium and on the cosmological expansion rate at various epochs, so the picture consistent with observations is obtained only for a certain composition of this medium. We will see, in particular, that this picture requires cold (or possibly warm) dark matter whose particles were non-relativistic already at fairly high temperatures.

In most of this book we consider the *linearized* theory of density and metric perturbations. This theory works when the perturbations $\delta\rho$ and $\delta g_{\mu\nu}$ are small compared to average density and metric,

$$\delta\rho \ll \rho, \quad \delta g_{\mu\nu} \ll g_{\mu\nu}.$$

Linearized theory is adequate for fairly early cosmological epoch. It is sufficient for the analysis of the issues related to CMB: at the time of photon last scattering (recombination), density and metric perturbations were, roughly speaking, of the order of the temperature fluctuation,

$$\frac{\delta\rho}{\rho} \sim \delta g_{\mu\nu} \sim \frac{\delta T}{T},$$

while CMB anisotropy data give $\delta T/T \sim 10^{-4} - 10^{-5}$.

On the other hand, perturbations of relatively small spatial sizes are strong in the *present* Universe. As an example, mass density in galaxies exceeds the average mass density in the Universe by a factor of $10^5 - 10^6$. Obviously, linearized theory is not applicable for complete analysis of these gravitationally bound systems. Still, the linearized theory not only enables one to understand how perturbations grew from tiny initial values to $\delta\rho/\rho \sim 1$, but also provides the tool for semi-quantitative calculations of some properties like the mass spectrum of structures at various redshifts. We will see, in particular, that perturbations at relatively small spatial scales indeed reach $\delta\rho/\rho \sim 1$, go non-linear and form structures, whereas perturbations at larger scales remain small (linear) up to the present epoch. This is in accordance with observations, at least at the qualitative level. The full quantitative theory of structure formation involves methods suitable for the analysis of non-linear dynamics, which are beyond the scope of this book. Most popular of these methods are large scale numerical simulations.

The Hot Big Bang theory does not explain the origin of the primordial density and metric perturbations themselves; this is one of the problems of that theory. An elegant solution to this problem is given by the inflationary theory which we present in Chapters 11–14 of this book.

1.1 Jeans Instability in Static Background

In general terms, the mechanism of the growth of density perturbations is as follows. An overdense region generates gravitational field that attracts matter from surrounding regions. The matter falls into the overdense region, and the density contrast in the latter becomes even higher. The process continues until the density contrast $\delta\rho/\rho$ becomes of order 1. Then the dynamics enters the non-linear regime, the overdense region collapses and forms a compact object like galaxy.

Let us consider this process by making use of the Newtonian theory of gravity and classical hydrodynamics. In this Section we present, for the purpose of illustration only, theoretically inconsistent case of static medium in flat space-time. The phenomenon we are going to observe is called Jeans instability; our naive treatment will be helpful for understanding the spatial and temporal scales of this instability. We study more realistic case of the expanding Universe in the next Section, and present the complete analysis within General Relativity in subsequent Chapters.

So, let us consider space filled with non-relativistic ideal fluid. This fluid induces gravitational potential according to Newton's law,

$$\Delta\phi = 4\pi G\rho, \quad (1.1)$$

where $\rho(\mathbf{x}, t)$ is the local mass density and Δ is the flat space Laplacian, $\Delta \equiv \nabla^2$. Fluid dynamics is described by its equation of state $p = p(\rho)$ and

hydrodynamic equations,

$$\frac{\partial \rho}{\partial t} + \nabla(\rho \mathbf{v}) = 0 \quad (1.2)$$

$$\rho \frac{\partial \mathbf{v}}{\partial t} + (\rho \mathbf{v} \nabla) \cdot \mathbf{v} = -\nabla p - \rho \nabla \phi, \quad (1.3)$$

where $p(\mathbf{x}, t) = p[\rho(\mathbf{x}, t)]$ is the local pressure and \mathbf{v} is the velocity of fluid. The first equation here (continuity equation) is obtained by considering mass balance in a given volume, while the second equation (generalized Euler equation) follows from the second Newton's law of classical mechanics applied to this volume.¹

Problem 1.1. Derive the hydrodynamic equations (1.2) and (1.3).

Let us now assume that in the absence of perturbations, the fluid is homogeneous and static,

$$\rho(\mathbf{x}) = \text{const}, \quad p(\mathbf{x}) = \text{const}, \quad \mathbf{v}(\mathbf{x}) = 0, \quad (1.4)$$

and also that the Newtonian potential vanishes. This assumption is *not*, in fact, justified, as it contradicts Eq. (1.1). In other words, the notion of homogeneous and static gravitating fluid contradicts the Newtonian mechanics. We will repair this in Section 1.2, and here we proceed under the assumption made.

The evolution of small perturbations $(\delta\rho, \delta p, \delta\mathbf{v} = \mathbf{v}, \delta\phi = \phi)$ is described by the system of equations obtained by linearizing Eqs. (1.1), (1.2) and (1.3):

$$\Delta\phi = 4\pi G \delta\rho, \quad (1.5)$$

$$\frac{\partial \delta\rho}{\partial t} + \rho \nabla \mathbf{v} = 0, \quad (1.6)$$

$$\frac{\partial \mathbf{v}}{\partial t} = -\frac{\nabla \delta p}{\rho} - \nabla \phi. \quad (1.7)$$

Hereafter ρ and p denote homogeneous background quantities. We use the equation of state $p = p(\rho)$ to eliminate the pressure perturbation by writing

$$\delta p = \frac{\partial p}{\partial \rho} \delta\rho \equiv u_s^2 \delta\rho, \quad (1.8)$$

where $u_s = \text{const}$ is the sound speed in the fluid. Upon differentiating Eqs. (1.6) and (1.7) with respect to time and spatial coordinates, respectively, we obtain, after obvious manipulations, the equation for the density perturbation $\delta\rho(\mathbf{x}, t)$:

$$\left(\frac{\partial^2}{\partial t^2} - u_s^2 \Delta - 4\pi G \rho \right) \delta\rho = 0. \quad (1.9)$$

¹As we mention in the accompanying book, Appendix I.A, in the framework of General Relativity, the continuity and Euler equations follow in the appropriate limit from the covariant conservation of energy-momentum tensor.

In the absence of gravity ($G = 0$) this equation describes the sound wave propagation in the fluid. Gravity, however, makes the dynamics quite different at large spatial scales.

We now perform the Fourier transformation,²

$$\delta\rho(\mathbf{x}, t) = \int d^3 q e^{i\mathbf{q}\mathbf{x}} \delta\rho(\mathbf{q}, t),$$

and write the solution as follows,

$$\delta\rho(\mathbf{q}, t) = e^{-i\omega t} \delta\rho(\mathbf{q}).$$

In this way we obtain the dispersion equation

$$-\omega^2 + u_s^2 \mathbf{q}^2 - 4\pi G \rho = 0. \quad (1.10)$$

It follows from (1.10) that Eq. (1.9) has solutions of two types, depending on the value of the momentum $q \equiv |\mathbf{q}|$. Perturbations of high momenta have positive frequency squared, and these perturbations are indeed sound waves. This occurs for $q > q_J$, where

$$q_J \equiv \sqrt{\frac{4\pi G \rho}{u_s^2}}. \quad (1.11)$$

For lower momenta, $q < q_J$, frequency ω is pure imaginary, and Eq. (1.9) does not admit propagating wave solutions. One of the modes grows exponentially,

$$\delta\rho(\mathbf{q}, t) \propto \exp(\Omega_q t),$$

where $\Omega_q = \sqrt{\Omega_J^2 - u_s^2 q^2}$ and $\Omega_J = \sqrt{4\pi G \rho}$. Another mode exponentially decays. Note that perturbations in pressureless fluid (dust) grow exponentially at arbitrarily small wavelengths.

The exponential growth is precisely the Jeans instability of perturbations of the gravitating fluid at $\lambda > \lambda_J$, where

$$\lambda_J \equiv \frac{2\pi}{q_J} = \sqrt{\frac{u_s^2 \pi}{G \rho}}, \quad (1.12)$$

is the Jeans length. Perturbations of initially small amplitudes get amplified, the density contrast becomes high, $\delta\rho \sim \rho$, and the system enters non-linear regime. The characteristic time of this process is

$$t_J = \Omega_J^{-1} = (4\pi G \rho)^{-1/2}.$$

²In this book, the argument of the spatial Fourier transform is called momentum (physical or conformal, depending on the context). We denote the physical momentum by q in what follows, and reserve p for pressure.

Once the non-linear regime sets in, the overdense region undergoes fast collapse resulting in the formation of a compact object. For perturbation of wavelength λ the mass of this object is estimated by

$$M \sim \rho \cdot \lambda^3.$$

Note that the Jeans time is of the order of the free fall time for an object of mass density ρ . Indeed, the free fall acceleration in a ball of this density at distance r from the center is

$$\mathbf{w}(r) = -G \frac{M(r)}{r^2} \frac{\mathbf{r}}{r} = -\frac{4\pi}{3} G \rho \mathbf{r}.$$

If a massive particle falls with the vanishing initial velocity into the ball from the distance R , then the free fall time t_{ff} is estimated from the relation $|\mathbf{w}(R)| \cdot t_{ff}^2 \sim R$, which gives precisely $t_{ff} \sim t_J$.

Problem 1.2. Find the exact expression for the free fall time.

It is instructive to compare the Jeans time with the Hubble time (see Chapter I.3). Recalling the Friedmann equation $H^2 = (8\pi/3)G\rho$, we see that $t_J \sim H^{-1}$ if the fluid considered dominates the cosmological expansion, and $t_J \gg H^{-1}$ in the opposite situation. In either case, long Jeans time implies that the static background approximation used in this Section is not valid. We will see in the next Section that the expansion of the Universe changes the way the Jeans instability develops. Nevertheless, Eq. (1.12) gives the right estimate for the Jeans length of non-relativistic matter perturbations in the expanding matter-dominated Universe.

The mass of a ball whose diameter equals the Jeans length is called the Jeans mass. It estimates the minimum mass of an object formed due to the Jeans instability. Let us give an example by estimating the Jeans mass right after the recombination epoch in the hypothetical Universe filled with hydrogen. We obtain

$$M_J = \frac{\pi}{6} \rho_H \lambda_J^3 = \frac{\pi^{5/2}}{6G^{3/2}} \frac{u_s^3}{\sqrt{\rho_H}},$$

where ρ_H is the mass density of hydrogen. Right after recombination,

$$\rho_H \approx m_p \eta_B n_\gamma(T_r) = \frac{2\zeta(3)}{\pi^2} m_p \eta_B T_r^3,$$

(we neglect helium for the estimate), where the temperature at recombination T_r is approximately 0.26 eV (see Section I.6.2), $\eta_B \simeq 6 \cdot 10^{-10}$ is the baryon-to-photon ratio (see Section I.5.2), and $u_s^2 = 5T_r/3m_p$ for the gas of non-relativistic particles. Thus, the numerical value of the Jeans mass is

$$M_J \simeq 10^6 M_\odot,$$

where the Solar mass in natural units is

$$M_\odot = 1.2 \cdot 10^{57} \text{ GeV.} \quad (1.13)$$

M_J would be the mass of the lightest objects that would start forming right after recombination in the hypothetical Universe. Curiously, this mass is of the order of the mass of the lightest dwarf galaxies, as well as heaviest globular clusters.

Let us emphasize that this calculation has nothing to do with the real Universe filled with dark matter. If dark matter is cold (rather than warm), it is the collisionless gas of particles with very low effective temperature. Hence, the effective pressure of dark matter is negligibly small, and Jeans instability develops in dark matter well before recombination. Perturbations of dark matter to large extent dominate the evolution of the density perturbations and gravitational potential in the real Universe after radiation-matter equality, while baryons fall after recombination into gravitational wells prepared by dark matter. These issues are discussed in detail in Chapters 6 and 7.

Consider now the gas of relativistic particles, whose equation of state is $p = \rho/3$ and the sound speed equals $u_s = 1/\sqrt{3}$. In this case, the Jeans length λ_J is comparable to the horizon size at radiation domination. The Newtonian approach is completely irrelevant in this situation; we will see in Section 4.2 that perturbations in relativistic gas do not grow at radiation domination.

To end this Section, we note that the spherically symmetric problem of gravitational collapse of pressureless gas (dust) admits exact solution both in the Newtonian gravity and in General Relativity; in the latter case it is known as the Tolman solution. We present these solutions in Appendix A. In both cases, a singularity develops in the center of a ball in the proper time of infalling particles of the order of the free fall time. In General Relativity, the solution describes black hole formation. The fact that black hole is formed from dust cloud of arbitrary mass is due to the exact spherical symmetry. In realistic situations, the black hole is not necessarily created. Also, pressure and dissipation are important in the formation of real astrophysical objects.

1.2 Development of Instability in Expanding Universe

The adequate theory for the full analysis of cosmological perturbations in the expanding Universe is General Relativity; this analysis is presented in subsequent Chapters. However, small perturbations of non-relativistic matter at length scales well below the Hubble size must be correctly described by the Newtonian theory. Indeed, space-time is effectively almost flat in that case, the fluid motion is slow, and the gravitational interaction is mostly due to particle masses. Hence, it must be possible to study both the expansion of the homogeneous and isotropic Universe and perturbations of matter of low pressure on the basis of Eqs. (1.1)–(1.3). This is indeed the case; here we give the details of the analysis.

The homogeneous and isotropic Universe is described in the Newtonian approach as follows,

$$\rho = \rho(t), \quad p = p(t), \quad \bar{\mathbf{V}} = \mathbf{x} \cdot \mathbf{H}(t), \quad \bar{\phi} = \frac{2\pi G}{3} \rho \mathbf{x}^2. \quad (1.14)$$

Hereafter bar denotes background values; the total velocity is denoted by \mathbf{V} in this Section. We do not use bar for background energy density and pressure; ρ and p always denote background quantities.

The potential $\bar{\phi}$ is indeed a solution to Eq. (1.1), while the functions $\rho(t)$ and $H(t)$ obey the system of equations

$$\frac{d\rho}{dt} + 3H\rho = 0, \quad (1.15a)$$

$$\frac{dH}{dt} + H^2 = -\frac{4\pi}{3}G\rho. \quad (1.15b)$$

This system is the consequence of Eqs. (1.2) and (1.3) in the Newtonian approach. In General Relativity, Eq. (1.15a) is the equation of the covariant energy-momentum conservation with pressure neglected, while Eq. (1.15b) is a consequence of the covariant conservation and Friedmann equations for the spatially flat Universe. It is thus not surprising that one of the solutions to (1.15) is the well-known cosmological solution for the matter dominated Universe,

$$H = \frac{2}{3t}, \quad \rho = \frac{1}{6\pi G t^2}. \quad (1.16)$$

We will not use other solutions in this Section.

Problem 1.3. *Find the general solution to the system (1.15). Compare to the result of General Relativity and explain the difference.*

Let us now consider the evolution of small perturbations

$$\delta\rho, \quad \delta p = u_s^2 \delta\rho, \quad \delta\phi = \Phi, \quad \delta\mathbf{V} = \mathbf{v}.$$

The notations for velocity perturbation and gravitational potential here are similar to the ones we use in the subsequent Chapters. We emphasize that the Newtonian approximation is not valid for distances of the order of the Hubble size, so we consider perturbations of wavelengths $\lambda \ll H^{-1}$. We obtain the relevant system of equations by expanding Eqs. (1.1), (1.2), (1.3) to the linear order in perturbations about the background (1.14), (1.15). This is similar to what we have done in Section 1.1. We note that even if pressure is small, it cannot be neglected when studying perturbations; as an example, the Jeans length (1.12) is determined by the sound speed $u_s = (\delta p / \delta\rho)^{1/2}$. Hence, we are not going to neglect δp . At the same time, pressure *can* be neglected in Eq. (1.15) for homogeneous background as long as $p \ll \rho$.

Instead of the absolute energy density perturbation $\delta\rho$ one conveniently introduces the relative perturbation

$$\delta(\mathbf{x}, t) \equiv \frac{\delta\rho(\mathbf{x}, t)}{\rho(t)}.$$

Then the linearized equations become

$$\Delta\Phi = 4\pi G\rho \delta, \quad (1.17a)$$

$$\frac{\partial\delta}{\partial t} + H\mathbf{x}\nabla\delta + \nabla\mathbf{v} = 0, \quad (1.17b)$$

$$\frac{\partial\mathbf{v}}{\partial t} + H\mathbf{v} + H \cdot (\mathbf{x}\nabla) \mathbf{v} = -u_s^2 \nabla\delta - \nabla\Phi. \quad (1.17c)$$

To solve this system, we introduce the following Fourier decomposition,

$$\delta(\mathbf{x}, t) = \int d^3k \exp \left[i\mathbf{k} \frac{\mathbf{x}}{a(t)} \right] \cdot \delta(\mathbf{k}, t), \quad (1.18)$$

and similarly for other quantities. Here $a(t)$ is the scale factor, which is defined in the Newtonian theory in analogy to General Relativity,

$$H(t) = \frac{1}{a} \frac{da}{dt}.$$

In General Relativity, the quantities \mathbf{k} and \mathbf{x}/a are conformal momentum and comoving coordinate; in the Newtonian theory, the Fourier transformation in the form (1.18) is merely a convenient trick that enables one to get rid of the explicit dependence of Eq. (1.17) on spatial coordinates. The time derivatives are then

$$\frac{\partial\delta}{\partial t} = \int d^3k \exp \left[i\mathbf{k} \frac{\mathbf{x}}{a(t)} \right] \cdot \left[-i\frac{\mathbf{kx}}{a} H\delta(\mathbf{k}, t) + \frac{\partial\delta(\mathbf{k}, t)}{\partial t} \right],$$

and similarly for $\partial\mathbf{v}/\partial t$. Therefore, the system of equations for the Fourier transforms is

$$k^2\Phi = -4\pi G\rho a^2\delta, \quad (1.19a)$$

$$\frac{\partial\delta}{\partial t} + i\frac{\mathbf{k}}{a}\mathbf{v} = 0, \quad (1.19b)$$

$$\frac{\partial\mathbf{v}}{\partial t} + H\mathbf{v} + i\frac{\mathbf{k}}{a}(u_s^2\delta + \Phi) = 0. \quad (1.19c)$$

Hereafter we use the same notations for functions of spatial coordinates and their Fourier transforms; we indicate their dependence on \mathbf{x} or \mathbf{k} when necessary, cf. (1.18).

Let us consider the evolution of the velocity perturbation. To this end, we use Eq. (1.19a) and write Eq. (1.19c) as follows,

$$\frac{\partial}{\partial t} (a\mathbf{v}) + i\mathbf{k} \left(u_s^2 - 4\pi G\rho \frac{a^2}{k^2} \right) \delta = 0. \quad (1.20)$$

For given \mathbf{k} , the velocity perturbation vector is conveniently decomposed into transverse and longitudinal components,

$$\mathbf{v} = \mathbf{V}^T + i\mathbf{k}v,$$

where the transverse component obeys $\mathbf{k} \cdot \mathbf{V}^T = 0$. Note that the longitudinal component in the coordinate representation is $\mathbf{v}^L = \nabla v$, so $v(\mathbf{x}, t)$ is the *velocity potential*.

Transverse and longitudinal components evolve independently. This is a consequence of the symmetry of the background under spatial rotations, see details in Section 2.2.3. It is clear from (1.20) that transverse velocity perturbations decouple from other perturbations. To study them, one consistently sets $\delta\rho = 0$. This type of perturbations is called *vector modes*. They are characterized by vanishing perturbations in energy density, pressure and gravitational field. The fluid velocity is orthogonal to the wave vector \mathbf{k} . Hence, we are dealing with purely rotational perturbations carrying angular momentum.

The transverse velocity perturbations obey

$$\frac{\partial}{\partial t} (a\mathbf{V}^T) = 0,$$

so they decay in time,

$$V^T = \frac{\text{const}}{a(t)}. \quad (1.21)$$

In matter dominated Universe we consider, the scale factor grows as $a(t) \propto t^{2/3}$, hence $V^T \propto t^{-2/3}$. The solution (1.21) has simple interpretation in terms of the angular momentum conservation in the expanding Universe. The angular momentum of fluid in a region of physical size l with rotational velocity V^T is estimated as the product of mass, velocity and size, $(\rho l^3) \cdot V^T \cdot l$. As the Universe expands, the physical size l grows proportionally to the scale factor a . Hence, the angular momentum conservation implies

$$\rho V^T a^4 = \text{const}. \quad (1.22)$$

Energy density ρ of non-relativistic matter decays as a^{-3} , so the latter relation reduces to (1.21).

Vector perturbations most probably do not play any role in the real Universe, since they rapidly decay at matter domination, see also Section 3.1.

Now, let us turn to the longitudinal component of velocity. According to (1.19b), its potential v is related to the density perturbation,

$$v = \frac{a}{k^2} \cdot \frac{\partial \delta}{\partial t}.$$

Let us make use of this relation to obtain a closed equation for the density perturbation. We express Φ by making use of (1.19a) and obtain from (1.19c) the following equation,

$$\frac{\partial^2 \delta}{\partial t^2} + 2H \frac{\partial \delta}{\partial t} + \left(\frac{k^2}{a^2} u_s^2 - 4\pi G \rho \right) \delta = 0. \quad (1.23)$$

This equation generalizes Eq. (1.9) to the expanding Universe. In the static case, $H = 0$, we come back to the result we already know: modes of physical momenta $q = k/a$ smaller than the Jeans momentum are exponentially unstable.

The Jeans instability occurs in the expanding Universe too, but now it has power-law, rather than exponential character. Let us show this by considering short wavelength modes, $q \ll q_J$. The Hubble parameter and energy density evolve at the matter dominated epoch according to (1.16), so the equation for short wavelength modes is

$$\frac{\partial^2 \delta}{\partial t^2} + \frac{4}{3t} \frac{\partial \delta}{\partial t} - \frac{2}{3t^2} \delta = 0. \quad (1.24)$$

The growing solution to Eq. (1.24) is

$$\delta \propto t^{2/3}, \quad (1.25)$$

so the density contrast grows linearly with the scale factor, $\delta \propto a$. The growth of perturbations is suppressed as compared to the static background; the physical reason is the expansion of the Universe.

Linear growth of perturbations with the scale factor is the property of matter domination only. The perturbations grow even slower at radiation domination, see below in this Section and Chapters 4, 6. Nevertheless, the overall relatively mild growth is sufficient for short wavelength modes to develop from the initial values $\delta \sim 10^{-4}$ to perturbations of order 1, enter the non-linear regime and form structures. We discuss this issue in some detail in Chapter 7.

Equation (1.23) can be generalized to the case of multi-component fluid. Namely, consider the medium consisting of several non-interacting components labeled by subscript λ . Each of these components has its own energy density, pressure and velocity ρ_λ , p_λ and \mathbf{v}_λ . Then Eq. (1.23) for the component λ is

$$\frac{\partial^2 \delta_\lambda}{\partial t^2} + 2H \frac{\partial \delta_\lambda}{\partial t} + \frac{k^2}{a^2} u_{s,\lambda}^2 \delta_\lambda - 4\pi G \sum_{\lambda'} \rho_{\lambda'} \delta_{\lambda'} = 0, \quad (1.26)$$

where the sum runs over all components. Let us use this equation for non-relativistic dark matter at radiation domination, assuming that the perturbations in radiation are negligible. The latter assumption is justified by the fact that perturbations of relativistic matter do not grow for $\lambda \ll H^{-1}$ (the Jeans length for relativistic matter is formally of order H^{-1} , see the end of Section 1.1). Hereafter we label dark matter by CDM . We have at radiation domination,

$$4\pi G \rho_{CDM} \ll H^2 = \frac{8\pi}{3} G \rho_{rad},$$

where rad denotes the relativistic component. Because of this inequality, the dark matter contribution to the sum in (1.26) is negligible, so the entire last term is

absent. Also, $u_s^2 = 0$ for cold dark matter. The Hubble parameter at radiation domination is $H(t) = (2t)^{-1}$, hence the dark matter perturbation obeys

$$\frac{\partial^2 \delta_{CDM}}{\partial t^2} + \frac{1}{t} \frac{\partial \delta_{CDM}}{\partial t} = 0,$$

The general solution to this equation is

$$\delta_{CDM}(t) = C_1 + C_2 \log \frac{t}{t_1}. \quad (1.27)$$

Thus, perturbations grow in time logarithmically. This growth is even slower than at matter domination; the reason is that the Universe expands rapidly at radiation domination. The general tendency is that the faster the expansion, the slower the growth of perturbations of non-relativistic matter.

Problem 1.4. *Find the evolution of perturbations of non-relativistic matter in the hypothetical Universe whose expansion is dominated by spatial curvature. Find the same for the epoch of the cosmological constant domination.*

To end this Section, we emphasize that our analysis here is of preliminary character. The main drawback of the Newtonian theory is that it cannot be used for perturbations whose wavelengths are of the order of the Hubble size and longer. This makes it impossible to relate the amplitudes of perturbations at late stages to the primordial perturbations. The second drawback is that this theory is not applicable to matter consisting of relativistic particles, e.g., baryon-photon plasma before recombination. Finally, the Newtonian theory lacks tensor perturbations. The full analysis of cosmological perturbations is possible only within General Relativity.

1.3 Linear Sizes of Perturbations and Masses of Objects

Once density perturbation of the present linear size R goes non-linear and collapses, a compact object is formed. The total mass of this object is

$$M(R) \simeq \frac{4\pi}{3} R^3 \rho_{M,0}, \quad (1.28)$$

where $\rho_{M,0} = \Omega_M \rho_c$ is the present average mass density of dark matter together with baryons. It is useful for what follows to get an idea of spatial sizes corresponding to one or another structure in the Universe. The values of the cosmological parameters are given in Sections I.1.3 and 2.1.2; in particular, $\rho_c = 1.5 \cdot 10^{11} \cdot M_\odot/\text{Mpc}^3$, $\Omega_M = 0.27$. We find from (1.28) that the size of linear perturbation $R = 1 \text{ Mpc}$ corresponds to an object of mass $M \simeq 1.6 \cdot 10^{11} M_\odot$. This mass scale is characteristic of usual (not dwarf) galaxies. Note that we are talking about the total mass that includes the mass of dark galactic halo. Thus, the length scales relevant for galaxies are

$$\text{galaxies: } R \sim (1 - 3) \text{ Mpc} \iff M \sim (10^{11} - 4 \cdot 10^{12}) M_\odot.$$

Our Galaxy is fairly large, its mass is about $10^{12} M_{\odot}$. We emphasize again that the size R is the present size of a region from which matter gathered in a galaxy; the sizes of galaxies themselves are much smaller, 10 – 100 kpc, and the mass densities in their central parts exceed the average density by a factor of 10^5 to 10^6 .

Clusters of galaxies correspond to much larger values of R and M :

$$\text{clusters of galaxies: } R \sim (10 - 30) \text{ Mpc} \iff M \sim (10^{14} - 4 \cdot 10^{15}) M_{\odot}.$$

These are the largest gravitationally bound systems in the Universe; their actual sizes are in the range of 1 – 3 Mpc, and the mass densities in the central regions are 100 – 1000 times higher than the average density.

The smallest objects containing stars embedded into dark matter halos are, presumably, dwarf galaxies. The corresponding estimates for them are

$$\text{dwarf galaxies: } R \sim (40 - 400) \text{ kpc} \iff M \sim (10^7 - 10^{10}) M_{\odot}.$$

It is likely that smaller halos cannot keep baryons inside, as baryons get blown out by the radiation from the first stars and by flows of matter generated in the bursts of these stars. The mass scale $M \sim 10^5 M_{\odot}$ is nevertheless of interest: the first stars were formed in halos of precisely these masses (see also the beginning of Section 7.2). The parameters of these halos are

$$\text{halos of protostars: } R \sim 10 \text{ kpc} \iff M \sim 10^5 M_{\odot}. \quad (1.29)$$

Clearly, the above classification is rather vague: there are objects of all masses in the Universe, from dwarf galaxies to rich clusters. We discuss the mass distribution of structures in Section 7.2.2.

Chapter 2

Cosmological Perturbations in General Relativity. Equations of Linearized Theory

We begin the study of the linear theory of cosmological perturbations within General Relativity. In the first place, we have to find out its equations in explicit form. These are obtained by linearizing the Einstein equations and covariant conservation equations for the energy-momentum tensor in the background of the time dependent homogeneous and isotropic metric and homogeneous energy density. We consider in the main text the major points of the derivation which are important for understanding of what is going on; some details are left for Appendix B.

In the subsequent Chapters we use the equations of the linear theory to study the evolution of the cosmological perturbations in the framework of the Hot Big Bang theory. The initial data for the evolution do not follow from this theory, so they have to be specified *ad hoc* in this approach. We postpone the discussion of primordial perturbations to Section 5.4, and before that we consider fairly general initial data (the only condition we impose is the absence of decaying mode, see Section 3.2.1). This way of presentation reflects the fact that the precise formulation of the initial value problem needs the knowledge of some properties of solutions to the evolution equations.

Literature on cosmology often invokes the notion of transfer functions: if $h(t_i)$ is the set of initial data, then the perturbation at time t can be written as

$$h(t) = T(t, t_i)h(t_i). \quad (2.1)$$

Here $T(t, t_i)$ is the transfer function. In the linear regime, this mapping is linear. In this language, the subsequent Chapters deal precisely with the transfer functions. We will rarely use this language, however.

2.1 Background Metric

2.1.1 Metric in conformal time

It is convenient to use conformal coordinates and write the metric of spatially flat¹ homogeneous and isotropic Universe as follows,

$$ds^2 = a^2(\eta)[d\eta^2 - dx^i dx^i] = a^2(\eta) \cdot \eta_{\mu\nu} dx^\mu dx^\nu, \quad (2.2)$$

where $\eta_{\mu\nu} = \text{diag}(+1, -1, -1, -1)$ is the Minkowski metric. We recall the relation between the conformal time η and cosmic time t ,

$$a(\eta)d\eta = dt.$$

In what follows, the derivative with respect to η is denoted by prime, while dot denotes the derivative with respect to t . As an example, the Hubble parameter is

$$H \equiv \frac{\dot{a}}{a} = \frac{a'}{a^2}. \quad (2.3)$$

The Friedmann equation is written as (see Section I.3.1)

$$\frac{a'^2}{a^4} = \frac{8\pi}{3} G\rho, \quad (2.4)$$

while (ij) -components of the Einstein equations reduce to

$$2\frac{a''}{a^3} - \frac{a'^2}{a^4} = -8\pi Gp, \quad (2.5)$$

where ρ and p are homogeneous energy density and pressure (we do not consider perturbations yet). We also recall the equation of the covariant energy conservation for the homogeneous background, see Section I.3.1,

$$\rho' = -3\frac{a'}{a}(\rho + p). \quad (2.6)$$

In terms of the conformal time, the solutions studied in Section I.3.2 are

— radiation dominated (RD) epoch:

$$a(\eta) = \text{const} \cdot \eta, \quad \eta = \text{const} \cdot t^{1/2}; \quad (2.7)$$

— epoch of the domination of non-relativistic matter (matter domination, MD):

$$a(\eta) = \text{const} \cdot \eta^2, \quad \eta = \text{const} \cdot t^{1/3}; \quad (2.8)$$

— epoch of dark energy domination (Λ -domination, Λ D):

$$a(\eta) = -\frac{1}{H_{dS}\eta}, \quad \eta = -\text{const} \cdot e^{-H_{dS}t}, \quad \eta < 0. \quad (2.9)$$

¹We recall that spatial curvature is negligible at all known cosmological epochs. We consider possible effects due to spatial curvature in appropriate places.

Here and in most other places in this book we assume that dark energy density is time-independent; in other words, dark energy is assumed to be cosmological constant = vacuum energy density. Hence, the Hubble parameter H_{dS} is independent of time, $H_{dS}^2 = (8\pi/3)G\rho_\Lambda$.

2.1.2 Cosmological parameters and conformal times of various epochs

Let us make a digression to summarize the properties of our Universe relevant throughout this book.

The energy content in the Universe is characterized by parameters Ω_λ which are the ratios of the present energy density of component λ to the critical density. We emphasize that we define Ω_λ for the present epoch only; hence, these parameters are independent of time by definition. Unless the opposite is explicitly stated, we use the following values in numerical estimates [2],

$$\begin{aligned}\Omega_{CDM} &= 0.223, & \Omega_B &= 0.046, \\ \Omega_M &= \Omega_B + \Omega_{CDM} = 0.27, & \Omega_\Lambda &= 0.73\end{aligned}\tag{2.10a}$$

$$H_0 = h \cdot 100 \frac{\text{km}}{\text{s} \cdot \text{Mpc}}, \quad h = 0.705\tag{2.10b}$$

$$\begin{aligned}\rho_c &= \frac{3}{8\pi} M_{Pl}^2 H_0^2 \\ &= 1.054 \cdot 10^{-5} h^2 \text{ GeV cm}^{-3} = 2.775 \cdot 10^{11} h^2 \cdot M_\odot \text{ Mpc}^{-3}\end{aligned}\tag{2.10c}$$

$$\eta_B \equiv \frac{n_B}{n_\gamma} = 6.2 \cdot 10^{-10}, \quad n_{\gamma,0} = 411 \text{ cm}^{-3}.\tag{2.10d}$$

We denote the present values by the subscript 0; in particular, H_0 is the present value of the Hubble parameter, $n_{\gamma,0}$ is the present number density of CMB photons. Notations CDM , B , M and Λ refer to dark matter, baryons, all non-relativistic matter and dark energy, respectively. We recall that dark matter may be cold or warm. The distinction between cold and warm dark matter is irrelevant for most of this book, and we use the notation CDM for definiteness. We will explicitly distinguish cold and warm dark matter in places where this distinction is important. We have not included neutrinos into the list in (2.10a); we will briefly say more about neutrinos in this Section and study specific effects due to neutrinos in Section 8.4. Also, we assume throughout most of this book that the Universe is spatially flat, $\Omega_{curv} = 0$; this is perfectly consistent with the observational bound [2]

$$-0.0175 < \Omega_{curv} < 0.0085.$$

Practically zero value of $\Omega_{curv} = 0$ is also preferred theoretically, see Section 11.2.

The precision at which these parameters are determined is continuously being improved, the current values can be found in Ref. [3]. We recall that [2]

$$\begin{aligned} H_0^{-1} &= 1.0 h^{-1} \cdot 10^{10} \text{ yrs} = 3.0 h^{-1} \cdot 10^3 \text{ Mpc} \\ &= 1.4 \cdot 10^{10} \text{ yrs} = 4.3 \cdot 10^3 \text{ Mpc}, \quad h = 0.705, \end{aligned}$$

or

$$H_0^{-1} = 3.1 h^{-1} \cdot 10^{17} \text{ s} = 0.93 h^{-1} \cdot 10^{28} \text{ cm} \quad (2.11a)$$

$$= 4.4 \cdot 10^{17} \text{ s} = 1.3 \cdot 10^{28} \text{ cm}, \quad h = 0.705. \quad (2.11b)$$

Let us make a comment on the parameter Ω_{rad} . We use the convention that Ω_{rad} is the relative energy density that photons and neutrino would have *in the case of zero masses of all neutrino species*. Since neutrino masses are not negligible at the present epoch, Ω_{rad} is *not* a true characteristic of the present Universe. Still, our convention is very useful. The point is that we are often interested in the energy density of the relativistic component at radiation domination or recombination. At that time, neutrino masses can be neglected. Indeed, the fairly conservative cosmological bound on neutrino masses is [2] $\sum_{e,\mu,\tau} m_{\nu_i} < 0.6 \text{ eV}$ (some analyses give even stronger bounds), so the neutrino oscillation data imply that the mass of every neutrino species is smaller than 0.2 eV. This is smaller than the temperature at photon last scattering $T_r = 0.26 \text{ eV}$ (see Section I.6.2), let alone the temperature at radiation-matter equality. With our definition, the energy density of the relativistic component at redshift z is obtained by simple rescaling, $\rho_{rad}(z) \propto (1+z)^4 \Omega_{rad} h^2$. The value of $\Omega_{rad} h^2$ defined according to our convention is well-known (assuming that there are no hypothetical new relativistic particles in our Universe), see Sections I.4.1 and I.4.4,

$$\Omega_{rad} h^2 = 1.68 \cdot \Omega_\gamma h^2 = 4.16 \cdot 10^{-5}, \quad (2.12)$$

so with the quoted value of h we have

$$\Omega_{rad} = 8.4 \cdot 10^{-5}, \quad h = 0.705.$$

With our values of parameters, the radiation-matter equality, that occurs when $\rho_M = \rho_{rad}$, is at redshift

$$z_{eq} = 3.2 \cdot 10^3.$$

It is useful for what follows to find the proportionality coefficients in (2.7) and (2.8). Deep at radiation domination, one has

$$H^2 = \frac{8\pi}{3} G \frac{\pi^2}{30} g_* T^4 = \left(\frac{g_{*,0}}{g_*} \right)^{1/3} \Omega_{rad} H_0^2 \left(\frac{a_0}{a} \right)^4,$$

where g_* is the effective number of degrees of freedom at temperature T and $g_{*,0} = 43/11$ is its present value, see Eqs. (I.5.13) and (I.5.33). Here we use the entropy conservation, $g_* a^3 T^3 = \text{const}$, and recall that the present entropy (again with massless neutrinos) is $s_0 = (2\pi^2/45) g_{*,0} T_0^3$. We find at radiation domination

$$\text{RD: } \eta = \int_0^t \frac{d\tilde{t}}{a(\tilde{t})} = \int_0^a \frac{d\tilde{a}}{\tilde{a}^2 H(\tilde{a})} \quad (2.13a)$$

$$= \left(\frac{g_*}{g_{*,0}} \right)^{1/6} \frac{1}{a_0 H_0 h^{-1} \sqrt{\Omega_{rad} h^2}} \frac{a}{a_0} \quad (2.13b)$$

This expression is basically free of uncertainties, if one considers temperatures well below 100 GeV, since the number of degrees of freedom is known at these temperatures, while $\Omega_{rad} h^2$ and $H_0 h^{-1}$ are given by (2.12) and (2.11a). Deep at matter domination, we have $H^2 = H_0^2 \Omega_M (a_0/a)^3$ and

$$\eta = \frac{2}{a_0 H_0 \sqrt{\Omega_M}} \left(\frac{a}{a_0} \right)^{1/2}. \quad (2.14)$$

This expression does have some uncertainty coming from $\Omega_M h^2$.

We need in a number of places fairly accurate estimates of conformal times of some cosmological epochs. Let us make these estimates for time-independent dark energy. In that case, the Friedmann equation for the spatially flat Universe can be cast into the relation between the redshift and Hubble parameter,

$$H = H_0 \sqrt{\Omega_\Lambda + \Omega_M (1+z)^3 + \Omega_{rad} (1+z)^4}.$$

The conformal time at redshift z is then

$$\eta = \int_z^\infty \frac{d\tilde{z}}{a_0 H_0 \sqrt{\Omega_\Lambda + (1+\tilde{z})^3 \Omega_M + (1+\tilde{z})^4 \Omega_{rad}}} \quad (2.15)$$

In particular, the conformal time of the present epoch is

$$\eta_0 = \frac{2}{a_0 H_0 \sqrt{\Omega_M}} I(\Omega_M), \quad (2.16)$$

where the integral $I(\Omega_M)$ is equal to (we encountered this integral in Section I.6.4, where we neglected Ω_{rad})

$$I(\Omega_M) = \frac{1}{2} \int_0^\infty \frac{dz}{\sqrt{(1+z)^3 + \frac{\Omega_\Lambda}{\Omega_M} + \frac{\Omega_{rad}}{\Omega_M} (1+z)^4}}.$$

Its numerical value at $\Omega_M = 0.27$, $\Omega_\Lambda = 0.73$ is

$$I(0.27) = 0.88. \quad (2.17)$$

To calculate the conformal time at recombination η_r , we note that ρ_Λ is negligibly small at $\eta \lesssim \eta_r$ and write

$$\begin{aligned}\eta_r &= \int_{z_r}^{\infty} \frac{dz}{a_0 H_0 \sqrt{(1+z)^3 \Omega_M + (1+z)^4 \Omega_{rad}}} \\ &= \frac{2}{a_0 H_0 \sqrt{\Omega_M}} \cdot \mathcal{F} \left(\frac{\Omega_{rad}}{\Omega_M} \right),\end{aligned}\quad (2.18)$$

where

$$\begin{aligned}\mathcal{F} \left(\frac{\Omega_{rad}}{\Omega_M} \right) &= \sqrt{\frac{1}{1+z_r} + \frac{\Omega_{rad}}{\Omega_M}} - \sqrt{\frac{\Omega_{rad}}{\Omega_M}} \\ &= \sqrt{\frac{1}{1+z_r} + \frac{1}{1+z_{eq}}} - \frac{1}{\sqrt{1+z_{eq}}}.\end{aligned}\quad (2.19)$$

Here

$$1+z_r = \frac{T_r}{T_0} = 1100$$

is the redshift at recombination. The second equality in (2.19) is obtained by making use of the relation $\Omega_M/\Omega_{rad} = 1+z_{eq}$. We note that the ratio Ω_{rad}/Ω_M can be viewed as a function of the only somewhat uncertain parameter $\Omega_M h^2$, since $\Omega_{rad} h^2$ is well-known, see (2.12). For our values of the parameters we have numerically

$$\mathcal{F} = 0.017.$$

This gives an important relation

$$\frac{\eta_0}{\eta_r} = \frac{I(\Omega_M)}{\mathcal{F}}, \quad (2.20)$$

so that for our values of the parameters

$$\frac{\eta_0}{\eta_r} = 51. \quad (2.21)$$

Note that the inverse of this quantity is the angle at which we see today the horizon of the recombination epoch, $\Delta\theta_r = \eta_r/\eta_0 \simeq 0.02$. Thus, we have refined the estimate² given in Section I.6.4.

Finally, the conformal time of matter-radiation equality is

$$\eta_{eq} = \frac{2}{a_0 H_0 \sqrt{\Omega_M}} \frac{1}{\sqrt{1+z_{eq}}} (\sqrt{2} - 1). \quad (2.22)$$

²In Section I.6.4 we used, instead of (2.19), the expression $(1+z_r)^{-1/2}$, i.e., we employed the approximation $z_r \ll z_{eq}$. This was not important for the discussion in Section I.6.4 of the dependence of the angle $\Delta\theta_r$ on cosmological parameters.

Numerically, we have

$$\frac{\eta_r}{\eta_{eq}} = 2.4, \quad \frac{\eta_0}{\eta_{eq}} = 1.2 \cdot 10^2. \quad (2.23)$$

We see that recombination and equality are not very distant in conformal time. We emphasize that the notion of equality is actually rather vague: the ratio ρ_M/ρ_{rad} depends on the scale factor rather weakly, $\rho_M/\rho_{rad} \propto a(\eta)$, so non-relativistic and relativistic matter have comparable energy densities during fairly long time interval around η_{eq} . The formal definition of equality as the time when $\rho_M = \rho_{rad}$ is precise, nevertheless.

Traditionally, the present value of the scale factor is set equal to 1 (unless possible spatial curvature is considered). We will *not* use this convention. Nevertheless, we note that with this convention, η_0 , η_r and η_{eq} have the meaning of the present size of the horizon of pertinent epochs. Numerically,

$$a_0\eta_0 = 1.4 \cdot 10^4 \text{ Mpc}, \quad a_0\eta_r = 2.8 \cdot 10^2 \text{ Mpc}, \quad a_0\eta_{eq} = 1.2 \cdot 10^2 \text{ Mpc}. \quad (2.24)$$

Problem 2.1. Repeat the above analysis of conformal times for dark energy equation of state $p_{DE} = w_{DE}\rho_{DE}$, where w_{DE} is time-independent. Make numerical estimates for $w_{DE} = -0.9$ and $w_{DE} = -1.1$. Take the same present values of parameters as in (2.10); in particular $\Omega_{DE} = \Omega_\Lambda = 0.73$.

To end this Section we note that expressions (2.18) and (2.22) contain the quantities referring to the present Universe (e.g., the present value of the Hubble parameter). Nevertheless, the conformal times of epochs in question are independent of the time we live, as should be the case. To see this explicitly, we again use (2.13a) and change the integration variable to temperature T . We make use of the relations $\rho_M(T) = \rho_{rad}(T)T_{eq}/T$, $a(T) = a_{eq}T_{eq}/T$ and write the Friedmann equation at the epoch of interest as follows,

$$H^2(T) = \frac{8\pi}{3}G\rho_{rad}(T) \cdot \left(1 + \frac{T_{eq}}{T}\right),$$

where $\rho_{rad}(T) = (\pi^2/30) \cdot \hat{g}_{*,0}T^4$, $\hat{g}_{*,0} = 3.36$ (see Section I.4.4). Upon integration, we obtain

$$\eta(T) = \frac{2 \left(\sqrt{1 + \frac{T_{eq}}{T}} - 1 \right)}{a_{eq} \sqrt{\frac{8\pi}{3}G \cdot \rho_{rad}(T_{eq})}} = \frac{2 \left(\sqrt{1 + \frac{T_{eq}}{T}} - 1 \right)}{a_{eq} T_{eq}^2 \sqrt{\frac{8\pi}{3}G \cdot \frac{\pi^2}{30} \hat{g}_{*,0}}} \quad (2.25)$$

This expression involves only quantities referring to the early Universe. It shows, in particular, that barring exotic relativistic component, the only relevant parameter is $\Omega_M/\Omega_{rad} \equiv T_{eq}/T_0$. Of course, the formula (2.25) is equivalent to the formulas presented above.

Problem 2.2. Show that (2.25) is equivalent to (2.18), (2.19) at $T = T_r$.

2.2 Generalities

2.2.1 Gauge $h_{0i} = 0$

Let us now consider metric with small perturbations,

$$ds^2 = a^2(\eta)\gamma_{\mu\nu}dx^\mu dx^\nu, \quad (2.26)$$

where

$$\gamma_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}(x).$$

We are always interested in metric perturbations $h_{\mu\nu}$ which depend on spatial coordinates in a non-trivial way³. The energy-momentum tensor is also a sum of the homogeneous background part and small perturbation (we denote the background energy-momentum tensor and Einstein tensor by overbar for the time being),

$$T_\nu^\mu = \bar{T}_\nu^\mu(\eta) + \delta T_\nu^\mu(x).$$

We always assume that the background metric $a^2\eta_{\mu\nu}$ and background energy-momentum tensor \bar{T}_ν^μ obey the Einstein equations and covariant conservation equations $\bar{\nabla}_\mu \bar{T}_\nu^\mu = 0$. The general structure of the linearized Einstein equations is

$$\delta G_\nu^\mu = 8\pi G \delta T_\nu^\mu, \quad (2.27)$$

where δG_ν^μ is the linear in $h_{\mu\nu}$ part of the Einstein tensor

$$G_\nu^\mu = R_\nu^\mu - \frac{1}{2}\delta_\nu^\mu R.$$

Our purpose in this Chapter is to find the explicit form of these equations, as well as the linearized equations of covariant energy-momentum conservation,

$$\delta(\nabla_\mu T_\nu^\mu) = 0. \quad (2.28)$$

A useful convention is that indices of objects describing small perturbations are raised and lowered by Minkowski metric. In particular

$$h_\nu^\mu = \eta^{\mu\lambda} h_{\lambda\nu}, \quad h^{\mu\nu} = \eta^{\mu\lambda} \eta^{\nu\rho} h_{\lambda\rho}.$$

To the linear order in $h_{\mu\nu}$, the tensor inverse to $\gamma_{\mu\nu}$, i.e., obeying $\gamma^{\mu\lambda} \gamma_{\lambda\nu} = \delta_\nu^\mu$, is

$$\gamma^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu}.$$

³Spatially homogeneous $h_{\mu\nu}$ would describe homogeneous but, generally speaking, anisotropic cosmological model.

We have to employ in what follows the $(3 + 1)$ -decomposition of four-dimensional tensors and write spatial and temporal indices separately. The reason is that the background is not invariant under Lorentz transformations but is invariant under spatial translations and rotations. We use Latin letters i, j, k, \dots for spatial indices. Yet another convention is that when performing $(3 + 1)$ -decomposition, we lower the spatial indices, so the $(3 + 1)$ -forms involve quantities with lower spatial indices only. Summation over repeated spatial indices is performed with Euclidean metric. As an example, the $(3 + 1)$ -form of the trace of metric perturbation is

$$h_\mu^\mu = h_{00} - h_{ii}.$$

The general expressions for the components of δG_ν^μ are given in Appendix B. We do not reproduce them here, and turn to the gauge choice instead.

General Relativity is invariant under infinitesimal gauge transformations, see Appendix I.A (we denote gauge transformed quantities by tilde since prime is reserved for the derivative with respect to η),

$$g^{\mu\nu} \rightarrow \tilde{g}^{\mu\nu} = g^{\mu\nu} + \nabla^\mu \xi^\nu + \nabla^\nu \xi^\mu \equiv g^{\mu\nu} + g^{\mu\lambda} \nabla_\lambda \xi^\nu + g^{\nu\lambda} \nabla_\lambda \xi^\mu, \quad (2.29)$$

where $\xi^\mu(x)$ are arbitrary functions. We make use of

$$g^{\mu\nu} = \frac{1}{a^2} (\eta^{\mu\nu} - h^{\mu\nu}) \quad (2.30)$$

and write

$$\tilde{g}^{\mu\nu} = \frac{1}{a^2} (\eta^{\mu\nu} - \tilde{h}^{\mu\nu}).$$

Hence, Eq. (2.29) gives the following linearized form of gauge transformations,

$$\tilde{h}^{\mu\nu} = h^{\mu\nu} - \partial^\mu \xi^\nu - \partial^\nu \xi^\mu - 2\eta^{\mu\nu} \xi^\lambda \frac{\partial_\lambda a}{a} \quad (2.31)$$

Both $h^{\mu\nu}$ and ξ^μ here should be treated as small quantities of the same order.

One can use the gauge freedom (2.31) to impose 4 gauge conditions on $h_{\mu\nu}$. In physical terms, this means the choice of the reference frame. Let us use part of the gauge freedom to set

$$h_{0i} = 0. \quad (2.32)$$

In a reference frame with this property, the proper time of an observer at rest is proportional to the coordinate time, $d\tau = a(\eta)[1 + (1/2)h_{00}]d\eta$. The three conditions (2.32) do not fix the gauge freedom completely: there remains invariance under gauge transformations parameterized by one function of spatial coordinates and time. The residual gauge transformations obey

$$\partial_0 \xi_i + \partial_i \xi_0 = 0. \quad (2.33)$$

Furthermore, there remains invariance under time-independent gauge transformations⁴ with

$$\xi_i = \xi_i(\mathbf{x}), \quad \xi_0 = 0. \quad (2.34)$$

The residual gauge freedom will be employed later on.

2.2.2 Linearized energy-momentum tensor of ideal fluid

Let us now consider the right hand side of the linearized Einstein equations. In most cases, the ideal fluid approximation is sufficient for our purposes; we introduce in Chapter 8 (see also Appendix B) the expressions valid beyond the ideal fluid approximation. The energy-momentum tensor of ideal fluid is, see Section I.A.10,

$$T_\nu^\mu = (\hat{\rho} + \hat{p})u^\mu u_\nu - \delta_\nu^\mu \hat{p}, \quad (2.35)$$

where

$$\hat{p} = p + \delta p, \quad \hat{\rho} = \rho + \delta \rho$$

are pressure and energy density with small perturbations; to simplify formulas below we denote the background pressure and energy by p and ρ . The 4-velocity u^μ also contains both background part and small perturbation. The background part is non-zero for the temporal component of the 4-velocity only. The 4-velocity always obeys

$$g_{\mu\nu} u^\mu u^\nu = 1, \quad (2.36)$$

so the non-vanishing components \bar{u}^μ of the background 4-velocity are $\bar{u}^0 = a^{-1}$, $\bar{u}_0 = a$. Hence, we write

$$u^0 = \frac{1}{a}(1 + \delta u^0) \quad (2.37)$$

$$u^i = \frac{1}{a}v^i, \quad (2.38)$$

where δu^0 and v^i are linear perturbations. Equation (2.36) with perturbations included reads

$$(1 + h_{00})(1 + \delta u^0)^2 = 1 + O(v^i v^i),$$

so to the linear order we have

$$\delta u^0 = -\frac{1}{2}h_{00}.$$

It follows from the definition (2.37) that

$$u_0 = a^2(1 + h_{00})u^0 = a \left(1 + \frac{1}{2}h_{00}\right).$$

⁴Gauge transformations depending on time only are of no interest to us, since we consider metric perturbations that non-trivially depend on spatial coordinates.

The spatial components v^i remain undetermined. We note that the spatial velocity v^i defined according to (2.38) is the physical velocity of fluid: according to the definition of 4-velocity one has

$$u^i = \frac{dx^i}{ds} \approx \frac{dx^i}{dt}$$

(the latter equality is exact to the linear order), while the physical velocity is

$$v^i = \frac{\delta l^i}{\delta t} = \frac{adx^i}{dt},$$

where $\delta l^i = adx^i$ is the physical line element. Making use of the above formulas, we find for the linearized energy-momentum tensor

$$\delta T_0^0 = \delta \rho \quad (2.39a)$$

$$\delta T_i^0 = -(\rho + p)v_i \quad (2.39b)$$

$$\delta T_j^i = -\delta_j^i \delta p, \quad (2.39c)$$

where we use the convention $v_i = \delta_{ij} v^j$.

Let us make a comment here. If the medium is *not* ideal fluid, one can *define* the quantities $\delta \rho$ and v_i by the relations (2.39a) and (2.39b), and *define* $\delta p = -(1/3)\delta T_i^i$. Then the perturbation δT_j^i will contain an extra traceless part Π_j^i called anisotropic stress. In the cosmological context, these properties are inherent, e.g., in the neutrino component after neutrino decoupling. We postpone to Chapter 8 the analysis of perturbations in situations where ideal fluid approximation is not adequate, and before that Chapter we stick to ideal fluid.

The linearized covariant conservation equations (2.28) have the following form in the gauge $h_{0i} = 0$,

$$\delta \rho' + 3 \frac{a'}{a} (\delta \rho + \delta p) + (\rho + p) \left(\partial_i v_i - \frac{1}{2} h' \right) = 0 \quad (2.40)$$

$$\partial_i \delta p + (\rho + p) \left(4 \frac{a'}{a} v_i + \frac{1}{2} \partial_i h_{00} \right) + [v_i (\rho + p)]' = 0, \quad (2.41)$$

where $h = h_{ii}$. Their derivation is given in Appendix B.

We make an important comment here. If the medium consists of several non-interacting components, the covariant conservation laws (2.40) and (2.41) are valid *for each component separately*. On the other hand, the right hand side of the Einstein equations contains the *total* energy-momentum tensor, i.e., the sum of energy-momentum tensors of all components. Hence, even though covariant conservation of *total* energy-momentum tensor is guaranteed by the Einstein equations, part of the covariant conservation equations (2.40), (2.41) is independent of the Einstein equations. This is relevant for various cosmological epochs: as an example, the main independent components before recombination are dark matter and baryon-electron-photon plasma (the latter is a single component before

recombination⁵ due to photon-electron scattering and Coulomb interaction of electrons and baryons). Neutrino gas is yet another independent component.

2.2.3 Helicity decomposition

The background metric (2.2) is invariant under spatial translations. So, it is convenient to perform Fourier transformation

$$h_{\mu\nu}(\eta, \mathbf{x}) = \int d^3k e^{i\mathbf{k}\mathbf{x}} h_{\mu\nu}(\mathbf{k}),$$

and similarly for other quantities like $\delta\rho$, δp , v_i . We use in what follows the same notations for quantities in coordinate and momentum representations, and write the argument explicitly when needed. The corresponding replacement in equations is

$$\partial_i \longleftrightarrow ik_i.$$

Hence, we often do not even specify in which representation (coordinate or momentum) we work. Since we are going to study linear equations, we consider modes of different spatial momenta independently. Recall that \mathbf{k} is conformal momentum, while the physical momentum \mathbf{q} depends on time (gets redshifted), $\mathbf{q}(\eta) = \mathbf{k}/a(\eta)$.

The background is invariant also under spatial rotations. Picking up a mode of a definite conformal momentum \mathbf{k} explicitly breaks this symmetry. There remains invariance under rotations around \mathbf{k} -axis, the symmetry group being $SO(2)$. Every 3-tensor can be decomposed into irreducible representations of this $SO(2)$. These are tensors of definite helicities (eigenvalues of the operator $\hat{L}_\alpha = -i\partial/\partial\alpha$, where α is the angle in the plane normal to \mathbf{k}). An object which does not transform under rotations around \mathbf{k} -axis has zero helicity. Examples are: 3-scalar; 3-vector directed along \mathbf{k} (i.e., $v_i \propto k_i$); 3-tensors $h_{ij} \propto k_i k_j$ and $h_{ij} \propto \delta_{ij}$. Vector, orthogonal to \mathbf{k} has unit helicity (generally speaking, a linear combination of helicities +1 and -1). Symmetric transverse traceless tensor h_{ij}^{TT} has helicity 2. Transversality and tracelessness of the latter mean that

$$h_{ii}^{TT} = 0, \quad k_i h_{ij}^{TT} = k_j h_{ij}^{TT} = 0. \quad (2.42)$$

Let us discuss these properties in more detail. Let the unit vectors $\mathbf{e}^{(1)}$ and $\mathbf{e}^{(2)}$ be orthogonal to each other and to vector \mathbf{k} and form, together with \mathbf{k} , right-handed triple (analogous to vectors $\mathbf{i_x}, \mathbf{i_y}, \mathbf{i_z}$ of the standard orthogonal coordinate frame in 3-dimensional Euclidean space). Then the rotation by angle α around the vector \mathbf{k} gives

$$\begin{aligned} \mathbf{e}^{(1)\prime} &= \mathbf{e}^{(1)} \cos \alpha - \mathbf{e}^{(2)} \sin \alpha \\ \mathbf{e}^{(2)\prime} &= \mathbf{e}^{(2)} \cos \alpha + \mathbf{e}^{(1)} \sin \alpha. \end{aligned} \quad (2.43)$$

⁵This statement is not precise; we consider this issue in Section 8.5.

Hence, the vectors

$$\mathbf{e}^{(\pm)} = \mathbf{e}^{(1)} \pm i\mathbf{e}^{(2)}$$

transform as follows

$$\mathbf{e}^{(\pm)\prime} = e^{\pm i\alpha} \mathbf{e}^{(\pm)}.$$

This precisely means that the vector $\mathbf{e}^{(+)}$ belongs to the irreducible representation of $SO(2)$ of helicity +1, while helicity of $\mathbf{e}^{(-)}$ is -1. An arbitrary transverse vector is a linear combination of $\mathbf{e}^{(+)}$ and $\mathbf{e}^{(-)}$, so it is a mixture of helicities +1 and -1. Of course, these properties are well-known from electrodynamics, where the vectors $\mathbf{e}^{(+)}$ and $\mathbf{e}^{(-)}$ are circular polarization vectors of photon.

Let us use the vectors $\mathbf{e}^{(1)}$ and $\mathbf{e}^{(2)}$ to construct two symmetric transverse traceless basis tensors

$$\begin{aligned} e_{ij}^{(+)} &= \frac{1}{\sqrt{2}} \left(e_i^{(1)} e_j^{(1)} - e_i^{(2)} e_j^{(2)} \right) \\ e_{ij}^{(\times)} &= \frac{1}{\sqrt{2}} \left(e_i^{(1)} e_j^{(2)} + e_i^{(2)} e_j^{(1)} \right). \end{aligned} \quad (2.44)$$

One can then construct two linear combinations of these tensors $e_{ij}^{(\pm 2)}$ which transform as follows

$$e_{ij}^{(\pm 2)\prime} = e^{\pm 2i\alpha} e_{ij}^{(\pm 2)}, \quad (2.45)$$

i.e., they have helicities +2 and -2. An arbitrary symmetric transverse traceless tensor of the second rank is a linear combination of $e_{ij}^{(+)}$ and $e_{ij}^{(\times)}$, i.e., a mixture of helicities +2 and -2.

Problem 2.3. Prove the last statements above.

Problem 2.4. Show that the third combination $(e_i^{(1)} e_j^{(1)} + e_i^{(2)} e_j^{(2)})$ has zero helicity, i.e., that it transforms trivially under rotations around \mathbf{k} -axis. Express this combination in terms of δ_{ij} and $k_i k_j$.

Problem 2.5. Construct the lowest rank tensors of helicities ± 3 and ± 4 .

Multiplication by ik_j (operation ∂_j in coordinate representation) does not change helicity. Since the linearized⁶ Einstein and covariant conservation equations involve this operation only (together with obviously scalar operations like differentiation with respect to time), these equations break up into independent sets of equations for cosmological perturbations of definite helicities. Thus there are tensor (helicity 2), vector (helicity 1) and scalar (helicity 0) perturbations, which evolve independently.

⁶It is important that we consider the linear theory.

In general, the helicity decomposition of metric perturbation, also called Lifshitz decomposition, is

$$h_{00} = 2\Phi \quad (2.46)$$

$$h_{0i} = ik_i Z + Z_i^T \quad (2.47)$$

$$h_{ij} = -2\Psi\delta_{ij} - 2k_ik_j E + i(k_i W_j^T + k_j W_i^T) + h_{ij}^{TT}, \quad (2.48)$$

where Φ, Z, Ψ, E are scalars (helicity 0), Z_i^T, W_i^T are transverse vectors ($k_i Z_i^T = k_i W_i^T = 0$, helicity 1) and h_{ij}^{TT} is transverse traceless tensor (helicity 2). With our gauge choice $h_{0i} = 0$ we have

$$Z_i^T = 0, \quad Z = 0.$$

Let us also write the helicity decomposition of velocity,

$$v_i(\eta, \mathbf{k}) = V_i^T(\eta, \mathbf{k}) + ik_i v(\eta, \mathbf{k}), \quad v_i(\eta, \mathbf{x}) = V_i^T(\eta, \mathbf{x}) + \partial_i v(\eta, \mathbf{x}), \quad (2.49)$$

where V_i^T is transverse and v is a scalar velocity potential. According to the above analysis, there are three independent sets of equations for tensor modes h_{ij}^{TT} , vector perturbations W_i^T, V_i^T and scalar perturbations $\Phi, \Psi, E, v, \delta\rho$ and δp , respectively. We note that transversality condition in the coordinate representation is

$$\partial_i W_i^T = 0, \quad \partial_i V_i^T = 0. \quad (2.50)$$

Let us make a simple comment on the helicity decomposition of the 3-tensor h_{ij} . The symmetric tensor h_{ij} has 6 independent components. Its decomposition (2.48) also has 6 independent components: one from each of Ψ and E and two from each of W_i^T and h_{ij}^{TT} (according to the number of signs of helicity). As should be the case, helicity decomposition does not change the number of independent functions of momentum and time. Similar comment applies to helicity decomposition of 3-vector v_i .

Problem 2.6. Construct projection operators extracting objects of definite helicity from h_{ij} , i.e., the operators $P_{ijkl}^{(\Psi)}$, $P_{ijkl}^{(TT)}$, etc., such that

$$P_{ijkl}^{(\Psi)} h_{kl} = -2\Psi\delta_{ij},$$

$$P_{ijkl}^{(TT)} h_{kl} = h_{ij}^{TT},$$

etc. Show that they obey the defining property of projectors, $P \cdot P = P$. Prove orthogonality and completeness of this system of operators.

To end this Section, we recall that general medium has non-vanishing anisotropic stress Π_j^i . It also has to be decomposed into helicity components. This is done in Appendix B, see (B.13b).

2.3 Equations in Helicity Sectors

This Section introduces the explicit forms of the linearized Einstein and covariant conservation equations in sectors of definite helicity. We still approximate the cosmic medium by ideal fluid(s) and work in the gauge $h_{0i} = 0$. The generalization to the case in which the ideal fluid approximation does not apply is given in Appendix B.

2.3.1 Tensor perturbations

We begin with helicity-2 modes, i.e., tensor perturbations. It is clear from (2.39) that the energy-momentum tensor does not have helicity-2 components. Hence, tensor modes are purely gravitational; these are gravity waves without sources⁷ propagating through the expanding Universe (the latter interpretation is not valid, though, for wavelength exceeding the current cosmological horizon, see Section 3.2).

Tensor perturbations are described by the transverse traceless tensor h_{ij}^{TT} . For $h_{00} = 0$, $h_{ij} = h_{ij}^{TT}$, the linearized Einstein tensor is (see Appendix B)

$$\begin{aligned}\delta G_{00} &= 0, \quad \delta G_i^0 = 0, \\ \delta G_j^i &= \frac{1}{2a^2} \left(\partial_\eta^2 h_{ij}^{TT} + 2\frac{a'}{a} \partial_\eta h_{ij}^{TT} - \Delta h_{ij}^{TT} \right),\end{aligned}\tag{2.51}$$

where

$$\Delta = \partial_i \partial_i.$$

Thus, the equation for tensor perturbations is

$$\partial_\eta^2 h_{ij}^{TT} + 2\frac{a'}{a} \partial_\eta h_{ij}^{TT} - \Delta h_{ij}^{TT} = 0.\tag{2.52}$$

It is useful for what follows to obtain the action for tensor modes. The lowest non-trivial order here is quadratic in h_{ij} . In general, this action can be obtained by expanding the complete action to the second order in perturbations. It is convenient, however, to simplify the calculation by using the following trick. Let us recall that the variation of the complete Einstein–Hilbert action is (see Appendix I.A)

$$\delta S = \frac{1}{16\pi G} \int d^4x \sqrt{-g} \delta g_{\mu\nu} (G^{\mu\nu} - 8\pi G T^{\mu\nu})\tag{2.53}$$

⁷This is a property of medium in the ideal fluid approximation. In general, gravity waves are sourced by energy-momentum tensor (anisotropic stress in the cosmological context). We note that gravity wave emission has been observed, albeit indirectly, by measuring the secular decrease of periods of binary pulsars PSR B1913+16, B1534+12, J0737–3039, etc.

We consider the variation of metric as the variation of $h_{\mu\nu}$ and write the variation of the action (2.53) up to the second order, treating $h_{\mu\nu}$ and $\delta h_{\mu\nu}$ as the first order quantities,

$$\delta S = \frac{1}{16\pi G} \int d^4x a^6 \delta h_{\mu\nu} [(\bar{G}^{\mu\nu} - 8\pi G \bar{T}^{\mu\nu}) + (\delta G^{\mu\nu} - 8\pi G \delta T^{\mu\nu}) + O(h^2)].$$

We made use here of the fact that $\sqrt{-g} = a^4$ for the tensor modes up to the second order, and expanded $G^{\mu\nu}$ and $T^{\mu\nu}$ to the linear order in metric perturbations. In fact, the expression $\delta h_{\mu\nu}(\bar{G}^{\mu\nu} - 8\pi G \bar{T}^{\mu\nu})$ vanishes, since the background metric obeys the Einstein equations. Now, we have for the tensors with upper indices

$$\begin{aligned} \delta G^{\mu\nu} - 8\pi G \delta T^{\mu\nu} &= a^{-2} \eta^{\mu\lambda} (\delta G_\lambda^\nu - 8\pi G \delta T_\lambda^\nu) - a^{-2} h^{\mu\lambda} (\bar{G}_\lambda^\nu - 8\pi G \bar{T}_\lambda^\nu) \\ &= a^{-2} \eta^{\mu\lambda} \delta G_\lambda^\nu \end{aligned} \quad (2.54)$$

When writing the second equality we again made use of the equations for the background metric and set $\delta T_\nu^\mu = 0$ for tensor perturbations. Thus,

$$\delta S = -\frac{1}{16\pi G} \int d^4x a^4 \delta h_{ij}^{TT} [\delta G_j^i + O(h^2)]. \quad (2.55)$$

We know that the term $\delta h \cdot h$ here is the variation of the quadratic action in question. Hence, the quadratic action is half of the right hand side of (2.55) with h_{ij}^{TT} substituted for δh_{ij}^{TT} ,

$$\begin{aligned} S_{TT} &= -\frac{1}{32\pi G} \int d^4x a^4 h_{ij}^{TT} \delta G_j^i \\ &= -\frac{1}{64\pi G} \int d^4x a^2 h_{ij}^{TT} \left(\partial_\eta^2 h_{ij}^{TT} + 2\frac{a'}{a} \partial_\eta h_{ij}^{TT} - \Delta h_{ij}^{TT} \right). \end{aligned}$$

Notice that the expression in parenthesis is the left hand side of the linearized field equation. Integrating by parts, we obtain finally

$$S_{TT} = \frac{1}{64\pi G} \int d^4x a^2 \left[(\partial_\eta h_{ij}^{TT})^2 - \partial_k h_{ij}^{TT} \partial_k h_{ij}^{TT} \right] \quad (2.56)$$

This is the desired expression for the quadratic action for tensor perturbations.

We note that the above trick is quite general. It can be used in many other cases for obtaining the *quadratic* actions in terms of the *linear* expressions entering equations of motion. The quadratic Lagrangian obtained in this way may differ by a total derivative from the “true” expression (the latter is obtained by expanding the complete action up to the second order); this difference, however, is irrelevant at least in perturbation theory.⁸

⁸At the classical level, adding total derivative to the Lagrangian never alters the dynamics; this is not necessarily the case in quantum theory. The famous example is the θ -term in Quantum Chromodynamics, see Section I.9.7.1.

Continuing with tensor modes, let us decompose (in momentum representation) the tensor h_{ij}^{TT} in the basis of helicity-2 tensors (2.44),

$$h_{ij}^{TT} = \sum_A e_{ij}^{(A)} h^{(A)}. \quad (2.57)$$

Hereafter in this Section

$$A = +, \times.$$

We find from (2.52) that each $h^{(A)}$ obeys

$$\partial_\eta^2 h^{(A)} + 2\frac{a'}{a} \partial_\eta h^{(A)} - \Delta h^{(A)} = 0. \quad (2.58)$$

The action (2.56) written in terms of scalar variables $h^{(A)}$ is

$$S_{TT} = \sum_A \frac{1}{64\pi G} \int d^3x d\eta a^2 [(\partial_\eta h^{(A)})^2 - \partial_k h^{(A)} \partial_k h^{(A)}]. \quad (2.59)$$

It describes perturbations of two polarizations in the expanding Universe.

Note that the limiting case of the Minkowski background is obtained by setting $a = 1$. Then η and \mathbf{x} are the usual temporal and spatial coordinates. Equation (2.58) reduces in this limit to

$$\square^{(4)} h^{(A)} = 0,$$

where $\square^{(4)}$ is the 4-dimensional D'Alembertian in the Minkowski space, $\square^{(4)} = \eta^{\mu\nu} \partial_\mu \partial_\nu$. The latter equation describes the propagation of gravity waves of two polarizations with the speed of light. The fact that Eq. (2.58) does not involve source term is due to the special form of the energy-momentum tensor (2.39).

Remarkably, the action (2.59) coincides, modulo the overall factor and two polarizations, with the action for massless scalar field minimally coupled to gravity and propagating in the expanding Universe. Of course, the field equations also coincide. The analogy between tensor perturbations and massless scalar field is often useful.

2.3.2 Vector perturbations

To obtain the linearized equations for vector modes (helicity-1) we set

$$h_{ij} = \partial_i W_j^T + \partial_j W_i^T \quad (2.60)$$

$$v_i = V_i^T \quad (2.61)$$

$$h_{00} = 0 \quad (2.62)$$

$$\delta\rho = \delta p = 0, \quad (2.63)$$

where W_i^T and V_i^T obey the transversality condition (2.50). It follows from (2.39) that the only non-vanishing component of the energy-momentum tensor is

$$\delta T_i^0 = -(\rho + p)V_i^T.$$

The covariant conservation equation (2.40) is identically satisfied, while Eq. (2.41) gives

$$\partial_\eta[(\rho + p)V_i^T] + 4\frac{a'}{a}(\rho + p)V_i^T = 0. \quad (2.64)$$

The $(0i)$ -component of the Einstein equations reads (see Appendix B)

$$\partial_\eta\Delta W_i^T = 16\pi Ga^2(\rho + p)V_i^T. \quad (2.65)$$

The (00) -component is satisfied identically, and (ij) -components give nothing new as compared to (2.64) and (2.65), see problem B.9.

According to our discussion in the end of Section 2.2.2, in the case of multi-component fluid, covariant conservation equation (2.64) must be satisfied by each component separately, while the right hand side of the Einstein equation (2.65) contains the sum of contributions of all components. This point is actually not very relevant here: since W_i^T does not enter Eq. (2.64), velocity V_i^T can be found for each of the components in terms of the initial velocity, and then W_i^T can be found from Eq. (2.65) as a sum of contributions of all components.

Note that the general solution to Eq. (2.65) contains an arbitrary function of spatial coordinates,

$$W_i^T = W_i^T(\mathbf{x}).$$

This solution, however, is pure gauge, which can be eliminated by the gauge transformation (2.31) with $\xi_i = W_i^T(\mathbf{x})$, $\xi_0 = 0$ (recall that this transformation is consistent with the gauge condition $h_{0i} = 0$, see (2.34)). Thus, no non-vanishing solutions to Eq. (2.65) exist in the absence of matter. This means, in particular, that free gravitational field in the Minkowski space does not have propagating helicity-1 modes.

2.3.3 Scalar perturbations: conformal Newtonian gauge

When studying scalar perturbations, it is convenient to make use of the residual gauge freedom with gauge functions obeying the relation (2.33). According to that relation, ξ_i is a longitudinal vector,

$$\xi_i = -\partial_i\sigma(\eta, \mathbf{x}),$$

while $\xi_0 = \partial_\eta\sigma$. We see from (2.31) that under this gauge transformation the spatial components of metric perturbation transform as follows,

$$\tilde{h}_{ij} = h_{ij} - 2\partial_i\partial_j\sigma - 2\frac{a'}{a}\delta_{ij}\sigma'.$$

Clearly, one can choose σ in such a way that the second term in (2.48) is eliminated. As a result, the non-vanishing perturbations are $\delta\rho$, δp and

$$h_{00} = 2\Phi, \quad (2.66)$$

$$h_{ij} = -2\Psi\delta_{ij}, \quad (2.67)$$

$$v_i = \partial_i v. \quad (2.68)$$

This gauge is called conformal Newtonian. In this gauge, the metric with scalar perturbations is

$$ds^2 = a^2(\eta)[(1 + 2\Phi)d\eta^2 - (1 + 2\Psi)d\mathbf{x}^2]. \quad (2.69)$$

This formula explains the terminology: in the Newtonian limit about the Minkowski background, Φ is the Newtonian potential, see Section I.A.8, while in the expanding Universe the 00-component of metric is conformally related to Φ .

We note that v is not the physical velocity potential, since the left hand side of (2.68) is the physical velocity while the right hand side contains the derivative with respect to the *conformal* coordinate. The physical velocity potential is

$$v_{phys} = av.$$

Indeed, the physical velocity is expressed as

$$v_i = \frac{\delta v_{phys}}{\delta l^i},$$

where $\delta l^i = adx^i$ is the physical length element.

The components of the Einstein tensor in the conformal Newtonian gauge are (see Appendix B)

$$\delta G_0^0 = \frac{2}{a^2} \left(-\Delta\Psi + 3\frac{a'}{a}\Psi' - 3\frac{a'^2}{a^2}\Phi \right) \quad (2.70)$$

$$\delta G_i^0 = \frac{2}{a^2} \left(-\partial_i\Psi' + \frac{a'}{a}\partial_i\Phi \right) \quad (2.71)$$

$$\begin{aligned} \delta G_j^i &= \frac{1}{a^2}\partial_i\partial_j(\Phi + \Psi) \\ &\quad - \frac{2}{a^2}\delta_{ij} \left[-\Psi'' + \frac{1}{2}\Delta(\Phi + \Psi) + \frac{a'}{a}(\Phi' - 2\Psi') + 2\frac{a''}{a}\Phi - \frac{a'^2}{a^2}\Phi \right] \end{aligned} \quad (2.72)$$

The ij -components of the energy-momentum tensor (2.39) have the structure δ_{ij} ; they do not contain longitudinal part proportional to $\partial_i\partial_j$ (in other words, to k_ik_j in the momentum representation). Since the tensors δ_{ij} and k_ik_j are independent, the longitudinal part of the ij -components of the Einstein equations, i.e., the first term in (2.72), should vanish,

$$\Psi = -\Phi. \quad (2.73)$$

Then the rest of the Einstein equations becomes

$$\Delta\Phi - 3\frac{a'}{a}\Phi' - 3\frac{a'^2}{a^2}\Phi = 4\pi Ga^2 \cdot \delta\rho_{tot} \quad (2.74)$$

$$\Phi' + \frac{a'}{a}\Phi = -4\pi Ga^2 \cdot [(\rho + p)v]_{tot} \quad (2.75)$$

$$\Phi'' + 3\frac{a'}{a}\Phi' + \left(2\frac{a''}{a} - \frac{a'^2}{a^2} \right)\Phi = 4\pi Ga^2 \cdot \delta p_{tot} \quad (2.76)$$

The covariant conservation equations are

$$\delta\rho'_\lambda + 3\frac{a'}{a}(\delta\rho_\lambda + \delta p_\lambda) + (\rho_\lambda + p_\lambda)(\Delta v_\lambda - 3\Phi') = 0 \quad (2.77)$$

$$[(\rho_\lambda + p_\lambda)v_\lambda]' + 4\frac{a'}{a}(\rho_\lambda + p_\lambda)v_\lambda + \delta p_\lambda + (\rho_\lambda + p_\lambda)\Phi = 0, \quad (2.78)$$

where λ labels the components of the cosmic medium. The latter equations are obtained from (2.40), (2.41) by inserting (2.66), (2.67), (2.68) there, and using (2.73).

The system of equations (2.74), (2.75), (2.76) and (2.77), (2.78) is the complete system for scalar perturbations in the ideal fluid approximation.

Let us make a few remarks in this regard. First, the simple relation (2.73) between the gravitational potentials Ψ and Φ is valid in the ideal fluid approximation only. The generalization beyond the ideal fluid approximation is given in Appendix B; it is used in Chapter 8. Second, as we already emphasized, in the case of multi-component medium, the right hand side of the Einstein equations (2.74), (2.75), (2.76) contains the sum of contributions of all components; this is reflected in the notation “tot”:

$$[(\rho + p)v]_{tot} = \sum_\lambda [(\rho_\lambda + p_\lambda)v_\lambda],$$

etc. On the other hand, the covariant conservation equations (2.77), (2.78) are valid for each component λ separately. Third, not all equations of the system (2.74)–(2.78) are independent because of the Bianchi identity. Recall that the similar property holds for the system of the Einstein and covariant conservation equations for homogeneous medium.

Problem 2.7. Show that in the case of one-component fluid, Eqs. (2.77), (2.78) are satisfied automatically, provided that perturbations obey Eqs. (2.74)–(2.76). Give the analog of this result for multi-component medium.

On the other hand, the system of equations (2.74)–(2.78) is incomplete: it does not contain information on the cosmic medium itself. In our ideal fluid case this system has to be supplemented with the relation between pressure and energy density for each of the components, i.e., the equation of state. In general, one writes for the background

$$p_\lambda = w_\lambda\rho_\lambda, \quad (2.79)$$

while perturbations obey

$$\delta p_\lambda = u_{s,\lambda}^2\delta\rho_\lambda,$$

where $u_{s,\lambda}$ is the sound speed in the component λ . In general, the parameter w_λ may be different from $u_{s,\lambda}^2$. Also, both w_λ and $u_{s,\lambda}^2$ may depend on the energy density ρ_λ , and hence on time; an important example is baryon-photon plasma

before recombination, whose energy density is due to baryons and photons and pressure due to photons only (see Section 6.2).

Now, in the static Universe limit $a' = a'' = \rho' = 0$, the system of equations we presented is the relativistic generalization of the Newtonian system of equations (1.5)–(1.7), and the two systems are basically the same in the non-relativistic limit $p \ll \rho$. In the Newtonian limit about the Minkowski background, Eq. (2.74) is precisely the Poisson equation for the Newtonian potential Φ . The relation (2.73) then determines the spatial part of the metric (2.69). Furthermore, at $p \ll \rho$ and for modes with $\lambda \ll H^{-1}$, the system (2.74)–(2.78) reduces to the system (1.17) that describes the evolution of perturbations in the Newtonian approximation in the background of the expanding Universe.

Problem 2.8. *Prove the above statements. Hint: When studying unperturbed medium, transform (1.17) to comoving reference frame.*

Problem 2.9. *Consider long distance asymptotics of the Schwarzschild solution. Show that it can be transformed to the form (2.69) with*

$$\Phi = -\Psi = -G \frac{m}{r}.$$

Our final comment is that in the Minkowski background and in the absence of matter, Eqs. (2.74)–(2.76) do not admit non-trivial solutions, so free gravitational field does not have propagating scalar (helicity-0) modes. We have come to the same conclusion for vector modes in Section 2.3.2. Hence, we have reproduced the well-known result that the only propagating modes of free gravitational field are transverse traceless gravity waves of two polarizations $e_{ij}^{(+,\times)}$.

2.4 Regimes of Evolution

Solutions to the evolution equations behave in a very different way depending on the relation between the physical momentum $q = k/a$ and the Hubble parameter $H = a'/a^2$. One can distinguish two regimes:

— *Superhorizon regime.* It occurs when

$$q(\eta) \ll H(\eta). \quad (2.80)$$

This means that the physical wavelength exceeds the current Hubble size H^{-1} . In the Hot Big Bang theory, the Hubble size is of the order of the cosmological horizon size, while in the Universe dominated by vacuum-like energy it is of the order of de Sitter horizon. Hence the term for this regime.

— *Subhorizon regime.* In this regime, the opposite inequality holds,

$$q(\eta) \gg H(\eta). \quad (2.81)$$

The physical wavelength is thus small compared to the current Hubble size.

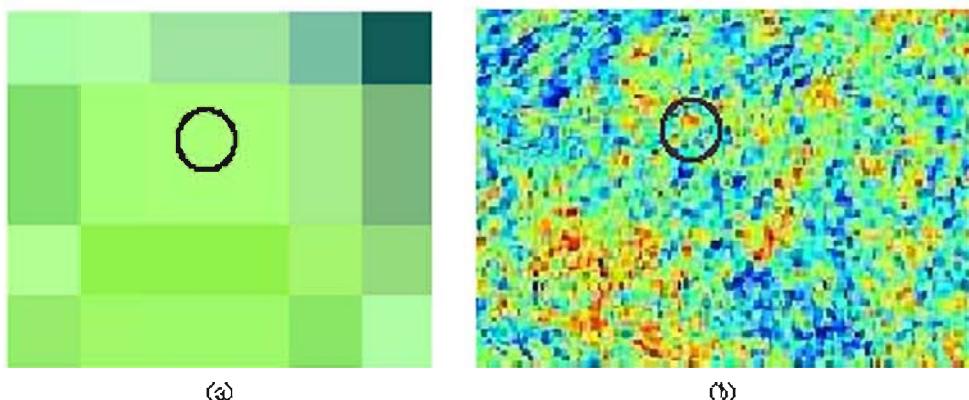


Fig. 2.1 Relationship between typical spatial scales of cosmological perturbations and the horizon size, see Fig. 17.1 for color version. The horizon is shown by circles. a) Superhorizon modes: perturbations are spatially homogeneous at the horizon scale; b) subhorizon modes: spatial inhomogeneities are sizeable inside the horizon.

The difference between the two regimes is as follows. Let us consider a perturbation of a certain wavelength; since the system is linear, we can disregard perturbations of other wavelengths. In the superhorizon regime, variations of metric, energy density and pressure over the horizon are small, see Fig. 2.1a. Recall that in the Hot Big Bang theory, the horizon scale is the maximum size of a causally connected region in the Universe. Hence, inside each of the causally connected regions metric, energy density and pressure are practically constant in space. These regions evolve as independent homogeneous Universes; one cannot talk about propagating sound waves or gravity waves in this case. Phenomena like Jeans instability are also absent.⁹ On the contrary, the expansion rate of the Universe is effectively low for subhorizon modes, and the usual notions of sound waves in the medium, gravity waves, etc., apply, see Fig. 2.1b.

At radiation and matter domination (as well as at any stage of decelerated expansion), the scale factor grows slower than linearly in t , while $H \propto 1/t$. Therefore, the ratio

$$\frac{a}{H} \propto \frac{t}{a(t)}$$

increases in time. Hence, at the Hot Big Bang epoch,¹⁰ a mode is first superhorizon, then it enters the horizon, and the subhorizon regime sets in; the larger the comoving wavelength, the later the perturbation enters the horizon. These properties are illustrated in Fig. 2.2.

⁹Since the evolution equations are local in time, these conclusions do not depend on whether the Hubble distance is the true cosmological horizon, or the actual size of a causally connected region is much greater than the Hubble size, as in the inflationary theory.

¹⁰The picture is entirely different at inflation, see Chapter 10.

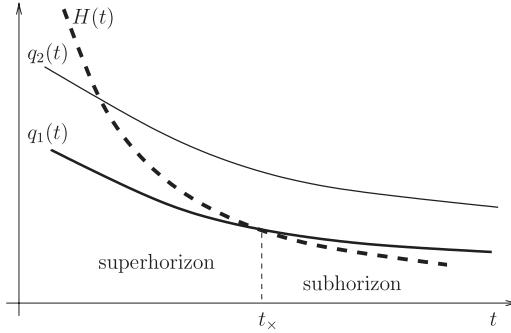


Fig. 2.2 Sketch of the time-dependence of the physical momenta (solid lines) and Hubble parameter (dashed line) at the Hot Big Bang epoch. t_x is the time of the horizon entry. Thick and thin lines show the evolution of the physical momenta for two different conformal momenta k_1 and k_2 , with $k_1 < k_2$.

We note that the ratio $q(t)/H(t)$ *decreases* in time at the present epoch of the accelerated cosmological expansion. Hence, the longest modes have never entered, and will never enter the horizon. The maximum present wavelength of a mode that has ever been subhorizon at the post-inflationary epoch¹¹ equals (for $\Omega_M = 0.27$, $\Omega_\Lambda = 0.73$)

$$\lambda_0^{(max)} = 32 \text{ Gpc.} \quad (2.82)$$

This mode briefly entered the horizon at the epoch of the transition from the decelerated to accelerated expansion, $z = z_{acc} \approx 0.76$, see Section I.4.3. The corresponding momentum is $q_0^{(min)} = 0.2 \text{ Gpc}^{-1}$. We do not consider waves of so large wavelengths in what follows.

Problem 2.10. *Assuming that dark energy density is time-independent, derive (2.82). What is the wavelength of a mode exiting the horizon today?*

Cosmologically interesting perturbations were indeed superhorizon at some early epoch. For a wave entering the horizon at temperature T_x at radiation domination, one has

$$q(T_x) \sim H(T_x) = \frac{T_x^2}{M_{Pl}^*}.$$

An order of magnitude estimate for its present wavelength is

$$\lambda_0 = \frac{a_0}{a(T_x)} \frac{2\pi}{q(T_x)} \sim \frac{2\pi M_{Pl}^*}{T_0 T_x}.$$

¹¹In inflationary theory, even modes of the present wavelengths much exceeding H_0^{-1} were superhorizon early at inflation.

For $T_{\times} \sim 1 \text{ MeV}$ (we know that the Universe had this temperature, since it corresponds to the beginning of the BBN epoch) we find

$$\lambda_0 \sim 1 \cdot 10^{21} \text{ cm} \sim 0.4 \text{ kpc},$$

which is smaller than any of the scales listed in Section 1.3. Thus, perturbations of all interesting wavelengths were indeed subhorizon at the Hot Big Bang epoch. We note that for $T_{\times} \sim 100 \text{ GeV}$ the estimate is

$$\lambda_0 \sim 0.5 \cdot 10^{16} \text{ cm},$$

i.e., somewhat larger than the size of the Solar system.

There is a simple relation between the conformal momentum k of a wave and the conformal time η_{\times} at which this wave enters the horizon. The horizon entry condition is

$$k \sim a(\eta_{\times})H(\eta_{\times}) = \frac{a'(\eta_{\times})}{a(\eta_{\times})}. \quad (2.83)$$

We now recall that $a(\eta)H(\eta) = a'/a \sim \eta^{-1}$, see (2.7), (2.8). Hence, η_{\times} is determined by

$$k\eta_{\times} \simeq 1. \quad (2.84)$$

Of course the relation (2.84) is somewhat vague, since the horizon entry is not an instantaneous event. Still, the simple formula (2.84) is very convenient. It is worth noting that the above discussion shows that the combination $k\eta$ is physical, even though k and η themselves are conformal variables.

Since the expansion rate is different at radiation and matter domination, it is of importance to understand which waves entered the horizon at each of the two epochs. The borderline is determined by

$$k^{(eq)}\eta_{eq} \simeq 1.$$

Waves with $k > k^{(eq)}$ and $k < k^{(eq)}$ enter the horizon at radiation and matter domination, respectively. Making use of (2.24), we find the present physical momentum of a wave with $k = k^{(eq)}$,

$$q_0^{(eq)} = \frac{k^{(eq)}}{a_0} = \frac{1}{a_0\eta_{eq}} \simeq 8.4 \cdot 10^{-3} \text{ Mpc}^{-1}. \quad (2.85)$$

The present wavelength is

$$\lambda_0^{(eq)} = \frac{2\pi a_0}{k^{(eq)}} \simeq 750 \text{ Mpc}. \quad (2.86)$$

We will be able to study analytically modes with either $k \gg k^{(eq)}$ or $k \ll k^{(eq)}$ (with qualifications). We sometimes call them RD-entering and MD-entering, respectively. As the estimate (2.86) shows, the modes relevant for galaxies and clusters of galaxies entered the horizon deep at the radiation domination epoch. On the other hand,

CMB properties are determined by both RD- and MD-entering modes, as well as by modes of intermediate momenta, $k \sim k^{(eq)}$. The latter are often difficult to treat analytically.

2.5 *Scalar Field Condensate as Dark Matter

In the previous Sections, cosmic matter has been described phenomenologically, in terms of the energy density and pressure. There are cases, however, when such a description may be questioned. One of them is the case of effectively massless scalar field, which is accurately studied in Section 13.2. Here we consider another possibility, namely, massive scalar field almost homogeneous in space. The relevant perturbations are scalar (helicity-0), since modes of non-zero helicity are absent for any scalar field. Physical applications include the axion field as dark matter candidate (see Section I.9.7.1), and curvaton field of Section 14.2 (after the onset of its coherent oscillations) as possible source of primordial density perturbations.

As we know from Section I.4.8.1, *homogeneous* scalar field of mass $m \gg H$ rapidly oscillates in time, and behaves as non-relativistic matter: spatial components of the energy-momentum tensor vanish upon averaging over the period of oscillations, while the average energy density decays as a^{-3} . Therefore, one expects that small inhomogeneous perturbations of this field behave in the same way as cold dark matter perturbations. The purpose of this Section is to prove this property by the direct analysis of small perturbations of the scalar field and metric. The only assumption here is that in the absence of the scalar field, metric varies in time slowly as compared to the characteristic time of oscillations m^{-1} . This assumption is valid if there are no other oscillating fields with appreciable energy density.¹² Besides the scalar field perturbations, there may exist perturbations of other matter components (e.g., of relativistic matter); no assumption on their amplitudes has to be made (except, of course, the general assumption that all perturbations are in linear regime).

So, we consider scalar field with the action

$$S_\phi = \int d^4x \sqrt{-g} \left(\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - \frac{m^2}{2} \phi^2 \right).$$

Unlike in Section I.4.8.1, we denote the scalar field by ϕ and reserve the notation φ for the perturbation of this field about the homogeneous background. The energy-momentum tensor of this field is

$$T_\nu^\mu = g^{\mu\lambda} \partial_\nu \phi \partial_\lambda \phi - \delta_\nu^\mu \left(\frac{1}{2} g^{\lambda\rho} \partial_\lambda \phi \partial_\rho \phi - \frac{m^2}{2} \phi^2 \right), \quad (2.87)$$

¹²In fact, a sufficient condition is that other oscillating fields, if any, have period of oscillations considerably different from m^{-1} . Also, one assumes that there are no oscillations of multiple periods.

while the field equation reads

$$\frac{1}{\sqrt{-g}} \partial_\mu (\sqrt{-g} g^{\mu\nu} \partial_\nu \phi) + m^2 \phi = 0. \quad (2.88)$$

With perturbations $\varphi(\mathbf{x}, \eta)$ included, the field has the form

$$\phi(\mathbf{x}, \eta) = \phi_c(\eta) + \varphi(\mathbf{x}, \eta), \quad (2.89)$$

where $\phi_c(\eta)$ is the homogeneous background. Throughout this Section we assume that the mass m is large compared to both physical momenta of perturbations and the Hubble parameter,

$$m \gg q \equiv \frac{k}{a}, \quad m \gg H. \quad (2.90)$$

Furthermore, for the time being we assume that yet another inequality holds,

$$m \gg \frac{q^2}{H}. \quad (2.91)$$

All these inequalities are indeed satisfied in most physically interesting cases. Nevertheless, at the end of this Section we consider the case in which the inequality (2.91) does not hold.

In the first place, let us write the equation for the homogeneous field in the homogeneous Universe,

$$\phi_c'' + 2 \frac{a'}{a} \phi_c' + a^2 m^2 \phi_c = 0. \quad (2.92)$$

As we know from Section I.4.8.1, the solution to this equation in the regime (2.90) is

$$\phi_c = u(\eta) \cdot \cos[mt(\eta)], \quad (2.93)$$

where

$$u(\eta) = \frac{\text{const}}{a^{3/2}} \quad (2.94)$$

is a slowly varying function of time,

$$t(\eta) = \int d\eta a(\eta)$$

is cosmic time, and we have chosen the phase of oscillations arbitrarily.

Let us find the linearized equations for the scalar field perturbations together with the metric perturbations, and also the linearized energy-momentum tensor. We use the conformal Newtonian gauge and assume for simplicity (though the result is valid in the general case) that the potentials Φ and Ψ are related by (2.73). Hence, the non-vanishing components of metric are (see (2.69))

$$\begin{aligned} g_{00} &= a^2(1 + 2\Phi), & g_{ij} &= -a^2 \delta_{ij}(1 - 2\Phi); \\ g^{00} &= a^{-2}(1 - 2\Phi), & g^{ij} &= -a^{-2} \delta_{ij}(1 + 2\Phi). \end{aligned} \quad (2.95)$$

By inserting (2.89) and (2.95) into the field equation (2.88) we obtain the equation for the field perturbation. In momentum representation

$$\varphi'' + 2\frac{a'}{a}\varphi' + k^2\varphi + m^2a^2\varphi - 4\phi'_c\Phi' + 2a^2m^2\phi_c\Phi = 0. \quad (2.96)$$

The energy-momentum for perturbations is obtained by inserting (2.89) and (2.95) into (2.87). One finds

$$\delta T_0^0 \equiv \delta\rho_\varphi = -\frac{1}{a^2}(\phi'_c)^2\Phi + \frac{1}{a^2}\phi'_c\varphi' + m^2\phi_c\varphi \quad (2.97)$$

$$\delta T_i^0 = \frac{1}{a^2}\phi'_c\partial_i\varphi \quad (2.98)$$

$$\delta T_j^i = -\delta_j^i\delta p_\varphi,$$

where

$$\delta p_\varphi = -\frac{1}{a^2}(\phi'_c)^2\Phi + \frac{1}{a^2}\phi'_c\varphi' - m^2\phi_c\varphi. \quad (2.99)$$

Comparing these expressions with (2.39), we see that the energy-momentum tensor has the standard form, and the role of the quantity $-(\rho + p)v$ entering Eq. (2.75) is played by

$$[-(\rho + p)v]_\varphi = \frac{1}{a^2}\phi'_c\varphi.$$

The idea of the analysis is to write the formal expansion of all quantities in m^{-1} and show that to the zeroth order, pressure δp_φ , and hence δT_i^j vanish upon averaging over the period of oscillations, while other components do not vanish. A subtlety is that the differentiation of rapidly oscillating functions with respect to η adds one power of the large parameter m ; for example, we have for the background field

$$\phi'_c = -mau \sin mt + \dots,$$

where dots denote corrections suppressed by m^{-1} .

To begin with, we notice that the background field ϕ_c (i.e., the amplitude u) is formally of order $O(m^{-1})$. Indeed, it is with this convention that its energy density

$$\rho_{\phi_c} = \frac{1}{2a^2}(\phi'_c)^2 + \frac{m^2}{2}\phi_c^2 \quad (2.100)$$

is of order $O(m^0)$. We decompose metric perturbations into slowly varying and oscillating parts $\bar{\Phi}$ and Φ_{osc} ,

$$\Phi = \bar{\Phi} + \Phi_{osc}, \quad (2.101)$$

then $\bar{\Phi}$ is of order $O(m^0)$, while we will see that Φ_{osc} should be considered as the quantity of order $O(m^{-1})$. Indeed, $\bar{\Phi}$ receives contributions from all components of cosmic medium. On the contrary, according to the assumptions made in the

beginning of this Section, the oscillating part Φ_{osc} is entirely due to the scalar field perturbations. Thus, we set

$$\bar{\Phi} = O(m^0), \quad \Phi_{osc} = O(m^{-1}). \quad (2.102)$$

Note that the time derivatives of these parts are of one and the same order,

$$\bar{\Phi}' = O(m^0), \quad \Phi'_{osc} = O(m^0). \quad (2.103)$$

Finally, to the leading and subleading orders, the oscillating field perturbation has frequency m only (at higher orders, the frequencies $2m, 3m, \dots$ are generated), so we can write

$$\varphi = B(\eta, \mathbf{x}) \sin mt + A(\eta, \mathbf{x}) \cos mt, \quad (2.104)$$

where A and B are slowly varying functions of conformal time.

To calculate the energy density and pressure to the leading (zeroth) order in m^{-1} we have to take into account the fact that the approximate background solution (2.93) is valid to the leading order in m^{-1} only. Including the subleading order, we have

$$\phi_c = u \cos mt + w \sin mt, \quad (2.105)$$

and according to the above discussion

$$u = O(m^{-1}), \quad w = O(m^{-2}). \quad (2.106)$$

By inserting (2.101) and (2.104) into (2.97)–(2.99), and making use of (2.106), we find the energy-momentum tensor up to order $O(m^0)$ (since A is small, see (2.110), the terms containing muA and wA , are of higher order, so we do not write them),

$$\begin{aligned} \delta\rho_\varphi &= -m^2 u^2 \bar{\Phi} \sin^2 mt + \frac{m}{a} u' B \cos^2 mt - \frac{m}{a} u B' \sin^2 mt \\ &\quad + m^2 w B + m^2 u A \end{aligned} \quad (2.107)$$

$$-[(\rho + p)v]_\varphi = -\frac{m}{a} u B \sin^2 mt \quad (2.108)$$

$$\begin{aligned} \delta p_\varphi &= -m^2 u^2 \bar{\Phi} \sin^2 mt + \frac{m}{a} u' B \cos^2 mt - \frac{m}{a} u B' \sin^2 mt \\ &\quad + m^2 w B \cos 2mt - m^2 u B \sin 2mt - m^2 u A \cos 2mt. \end{aligned} \quad (2.109)$$

These components, together with possible contributions of other types of matter, enter the right hand sides of Eqs. (2.74), (2.75) and (2.76). Making use of (2.103), we find that the left hand side of Eq. (2.74) is of order $O(m^0)$. Since $u = O(m^{-1})$, the last term in (2.107) is of order $O(m^0)$ (and not higher) for $A = O(m^{-1})$. On the other hand, other terms are of order $O(m^0)$ provided that $B = O(m^0)$. We conclude that the self-consistency of the expansion in m^{-1} requires that

$$B = O(m^0), \quad A = O(m^{-1}). \quad (2.110)$$

These assignments ensure also that other components of the energy-momentum tensor have correct properties¹³ with respect to the expansion in m^{-1} .

Let us now turn to the field equation (2.96). Its leading order is $O(m^1)$. To this order it reads

$$B' + \frac{3a'}{2a}B + mau\bar{\Phi} = 0. \quad (2.111)$$

It is now straightforward to see that the pressure perturbation δp_φ averaged over the period vanishes up to and including the order $O(m^0)$. Indeed, we average (2.109) over the period and obtain

$$\overline{\delta p}_\varphi = -\frac{1}{2}\frac{mu}{a}\left(B' - \frac{u'}{u}B + mau\bar{\Phi}\right). \quad (2.112)$$

It follows from (2.94) that this expression coincides with the left hand side of (2.111) up to a multiplicative factor. Thus,

$$\overline{\delta p}_\varphi = O(m^{-1}),$$

as promised. Other components of the energy-momentum tensor do not vanish to the order $O(m^0)$ upon averaging over the period. In particular,

$$\overline{[(\rho + p)v]}_\varphi = \frac{1}{2}\frac{m}{a}uB.$$

Hence, the equation of state for perturbations of the scalar field is, on average, $p = 0$, up to corrections suppressed by m^{-1} . Since the knowledge of the equation of state is sufficient for studying the evolution of perturbations, the scalar field condensate is equivalent to non-relativistic matter from the viewpoint of the cosmological perturbations. We conclude that at time scales exceeding m^{-1} one can forget about the field oscillations and describe slightly inhomogeneous scalar field in the same way as cold dark matter; the corresponding formulas are the same as in Section 2.3.3 with $p = \delta p = 0$.

A non-trivial check of the entire approach is the fulfillment of the covariant conservation equations for the energy-momentum tensor of the scalar field perturbations, *averaged over the period of oscillations*, with the background field and metric also averaged over the period. The covariant conservation of energy-momentum tensor *before averaging* is guaranteed by the field equations; this property, however, is no longer obvious after averaging. The point is that both the background field ϕ_c and background metric oscillate themselves, and that the average of a product of oscillating quantities is not equal to the product of their average values. Let us show that to the leading order in m^{-1} , the covariant conservation equations (2.77), (2.78) are indeed satisfied for average quantities. This will be the final proof that the scalar field is equivalent to fluid with $p = 0$ to the leading order in m^{-1} (this equivalence does not hold at higher orders).

To check that the covariant conservation equations for the average energy-momentum tensor are indeed satisfied, we need the expressions for relevant quantities to both the

¹³Note that the left hand side of Eq. (2.76) has the oscillating part Φ''_{osc} which is of order $O(m^1)$. The term of the same order exists also in δp_φ : this is the last term in (2.109).

leading and subleading orders in m^{-1} . Also, we will encounter products of fast oscillating functions like $\sin mt \cdot \sin 2mt$. It turns out to be sufficient to consider the zeroth and first harmonics, i.e., functions that either do not oscillate at all or oscillate with frequency m . In this regard, we will use, without explicitly indicating that, the following replacements,

$$\begin{aligned}\sin mt \cdot \sin 2mt &\rightarrow \frac{1}{2} \cos mt, & \cos mt \cdot \sin 2mt &\rightarrow \frac{1}{2} \sin mt, \\ \sin mt \cdot \cos 2mt &\rightarrow -\frac{1}{2} \sin mt, & \cos mt \cdot \cos 2mt &\rightarrow \frac{1}{2} \cos mt.\end{aligned}$$

Higher harmonics are relevant at the next-to-subleading order only, and we are not going to use them.

In the first place, the scale factor is an oscillating function at the subleading order (due to the oscillating background field ϕ_c). So, we write

$$a(\eta) = \bar{a}(\eta) + a_{osc}(\eta), \quad (2.113)$$

where \bar{a} does not oscillate and is of order $O(m^0)$, while a_{osc} is of order $O(m^{-2})$. The latter property follows from the fact that the oscillating part of the background energy density (2.100) is, to the leading non-trivial order,

$$\rho_{osc} = -\frac{1}{2\bar{a}} muu' \sin 2mt, \quad (2.114)$$

i.e., it is of order $O(m^{-1})$. Thus, $a'_{osc} = O(m^{-1})$, which gives $a_{osc} = O(m^{-2})$. We insert (2.113) and (2.114) into the Friedmann equation (2.4) and obtain, again to the leading non-trivial order,

$$a'_{osc} = \pi G m \bar{a}^2 u^2 \sin 2mt,$$

where we made use of (2.94). Hence, we have

$$a_{osc} = -\frac{\pi}{2} G \bar{a} u^2 \cos 2mt.$$

It is also convenient to write the equation for the function $w(\eta)$ determining the subleading part of the background scalar field (2.105),

$$w' + \frac{3\bar{a}'}{2\bar{a}} w + \frac{1}{2m\bar{a}} \left(u'' + \frac{2\bar{a}'}{\bar{a}} u' \right) - \frac{3\pi}{4} G m \bar{a} u^3 = 0. \quad (2.115)$$

This equation is obtained from (2.92) by extracting the terms of order $O(m^{-1})$. We do not write Eq. (2.92) to the higher orders, $O(m^1)$ and $O(m^0)$; we know that its solution is given by (2.93).

We cannot neglect also the oscillating term in the Newtonian potential (2.101). To the leading non-trivial order it is related to the scalar field perturbations by the Einstein equation (2.75). We extract the oscillating terms in the latter and obtain, to the leading order,

$$\Phi'_{osc} = 2\pi G m \bar{a} u B \cos 2mt,$$

which gives

$$\Phi_{osc} = \pi G u B \sin 2mt. \quad (2.116)$$

Hence, Φ_{osc} is indeed of order $O(m^{-1})$. Equation (2.74) gives nothing new at the orders we consider, see problem below.

Finally, we write the Klein–Gordon equation (2.96) to the leading and subleading orders in m^{-1} . To the leading order $O(m^1)$ it reduces, as we mentioned already, to Eq. (2.111)

for slowly varying quantities. To the subleading order $O(m^0)$ we find, by making use of the above formulas,

$$B'' + \frac{2\bar{a}'}{\bar{a}}B' + k^2B - 2m\bar{a}A' - 3m\bar{a}'A + 4m\bar{a}u\bar{\Phi}' + 2m^2\bar{a}^2w\bar{\Phi} - \frac{3\pi}{2}Gm^2\bar{a}^2u^2B = 0. \quad (2.117)$$

Problem 2.11. Show that to the leading non-trivial order in m^{-1} , the oscillating part of Eq. (2.74) reduces to Eqs. (2.116) and (2.111).

We are now ready to consider the covariant conservation equations for the average energy-momentum tensor of perturbations. To the leading order, the average quantities are obtained by averaging (2.107), (2.108) and are given by

$$\begin{aligned}\overline{\delta\rho_\varphi} &= m^2wB + m^2uA, \\ \overline{[(p+\rho)v]_\varphi} &= \frac{m}{2\bar{a}}uB, \\ \overline{\delta p_\varphi} &= 0,\end{aligned}$$

where we used (2.111). Our purpose is to check that they obey Eqs. (2.77), (2.78), in which the scale factor, Newtonian potential and background energy density are \bar{a} , $\bar{\Phi}$ and $\bar{\rho}_{\phi_c}$, while the background pressure of the scalar field is zero.

We begin with Eq. (2.78). To the leading order, we have

$$\overline{(\rho+p)}_{\phi_c} = \frac{1}{2}m^2u^2.$$

Equation (2.78) for average quantities is then cast into the form

$$\left(\frac{u}{\bar{a}}B\right)' + 4\frac{\bar{a}'u}{\bar{a}^2}B + mu^2\bar{\Phi} = 0. \quad (2.118)$$

Making use of the fact (see Eq. (2.94)) that

$$\frac{u'}{u} = -\frac{3\bar{a}'}{2\bar{a}}, \quad (2.119)$$

we find that Eq. (2.118) indeed reduces to Eq. (2.111), so it is indeed a consequence of the Klein–Gordon equation.

Making use of (2.115), (2.111) and above formulas, we cast equation (2.77) for the average quantities into the following form,

$$2m^2\bar{a}^2w\bar{\Phi} + \left(\frac{u''}{u} + 2\frac{\bar{a}'u'}{\bar{a}u}\right)B - \frac{3\pi}{2}Gm^2\bar{a}^2u^2B - 2m\bar{a}A' - 3m\bar{a}'A + k^2B + 3m\bar{a}u\bar{\Phi}' = 0.$$

This is equivalent to Eq. (2.117). A simple way to see that is to differentiate (2.111) with respect to η and use the resulting equations to eliminate B' and B'' from (2.117). One should also recall the relation (2.119) and its derivative with respect to η .

Thus, the Klein–Gordon equation for the scalar field and its perturbations gives the same covariant conservation equations for average energy-momentum tensor as those characteristic to pressureless matter. We conclude that the phenomenological description of slightly inhomogeneous massive scalar field as non-relativistic matter is exact to the leading order in m^{-1} .

The above analysis does not go through if the inequality (2.91) does not hold. So, let us consider the case

$$\frac{q^2}{H} \equiv \frac{k^2}{a^2H} \gtrsim m. \quad (2.120)$$

This situation occurs for tiny mass of the scalar field. As an example, the relation (2.120) is valid at the equality epoch for perturbations of the present wavelength $q_0^{-1} = 1$ Mpc, provided that

$$m \lesssim \frac{q_0^2 z_{eq}^2}{H_{eq}} \sim 0.3 \text{ pc}^{-1} \sim 2 \cdot 10^{-24} \text{ eV}, \quad (2.121)$$

where the numerical estimate is obtained with $z_{eq} = 3.2 \cdot 10^3$, $H_{eq}^{-1} = 2t_{eq} = 3.2 \cdot 10^{12} \text{ s} = 0.97 \cdot 10^{23} \text{ cm}$. Still, the fields of extremely small mass are discussed in literature as dark matter candidates [4–6]. The reason, as we see below, is that the density perturbations have interesting properties in this case.

The Hubble parameter is of order $O(m^0)$, so we see from (2.120) that we should assign conformal momentum the order $O(m^{1/2})$, i.e.,

$$k^2 = O(m).$$

We again set $u = O(m^{-1})$, $w = O(m^{-2})$, $\Phi = O(m^0)$. Then Eq. (2.74) requires that, formally, $\delta\rho = O(m)$. This property holds for the scalar field perturbations if one assigns the same order to the functions A and B entering (2.104), i.e., instead of (2.110) one sets

$$B = O(m^0), \quad A = O(m^0). \quad (2.122)$$

Then the energy density perturbation is, to the leading order,

$$\delta\rho_\varphi = m^2 u A, \quad (2.123)$$

while the averaged over the period pressure perturbation is still given by (2.112). However, the expression in the right hand side is no longer equal to zero. Indeed, we use (2.122) and find that the field equation (2.96) now gives two equations

$$\begin{aligned} B' + \frac{3a'}{2a}B + mau\bar{\Phi} + \frac{k^2}{2ma}A &= 0, \\ A' + \frac{3a'}{2a}A - \frac{k^2}{2ma}B &= 0. \end{aligned}$$

Making use of the first of these equations, we cast (2.112) into the form

$$\overline{\delta p_\varphi} = \frac{k^2}{4a^2} u A.$$

This expression is of order $O(m^0)$. Hence, generally speaking, pressure is no longer negligible. In the case under study, oscillating scalar field is *not* equivalent to non-relativistic matter.

The latter expression together with (2.123) gives for the sound speed

$$u_s^2 \equiv \frac{\delta p_\varphi}{\delta\rho_\varphi} = \frac{k^2}{4a^2 m^2} = \frac{q^2(\eta)}{4m^2}.$$

It is non-zero, albeit small. Let us illustrate the importance of this observation by the following example. Let us consider matter domination and assume that the

oscillating field ϕ makes all of dark matter. Then, as we know from Chapter 1, the system exhibits the Jeans instability. Since the sound speed does not vanish, the Jeans momentum is non-zero [4, 5],

$$q_J = (16\pi G \rho_\phi m^2)^{1/4}. \quad (2.124)$$

We note that

$$\frac{q_J^2}{H} \sim m,$$

i.e., the condition (2.91) ceases to hold precisely at the Jeans scale. Perturbations of momenta exceeding the Jeans momentum *do not grow*.

Thus, superlight oscillating scalar field is *warm* dark matter candidate: the spectrum of its perturbations is effectively cut off at short wavelengths. The order of magnitude of the cut-off scale is determined by the Jeans momentum (2.124) taken at radiation-matter equality (we will see in subsequent Chapters that the rapid growth of perturbations occurs at matter domination). The present momentum is estimated by

$$q_0 \sim z_{eq}^{-1} q_J(t_{eq}) \sim z_{eq}^{-1} (H_{eq} m)^{1/2}.$$

The estimate (2.121) shows that scalar field of mass $m \ll 10^{-23}$ eV is unacceptable as dark matter candidate: in such a model, galaxies do not form. On the other hand, scalar field of mass $m \sim 10^{-22} - 10^{-23}$ eV is interesting: it would form objects of galaxy scale in the same way as cold dark matter, while dwarf galaxies and objects of sub-galactic size would be suppressed. These are precisely the properties of warm dark matter, see Section 8.3.1.

To end this Section we make a comment relevant to all models with dark matter made of oscillating scalar field. The consistency with observational data requires that the primordial perturbations of this field are correlated with the primordial perturbations in hot cosmic plasma. More precisely, it is required that the adiabatic mode is dominant, see Section 5.1. As we will see in Section 14.4, this property does hold in the theory of axion in a certain range of its parameters. On the other hand, the adiabaticity requirement may be highly non-trivial in other scalar field theories.

This page is intentionally left blank

Chapter 3

Evolution of Vector and Tensor Perturbations

We begin with the analysis of the evolution of vector and tensor perturbations, since it is fairly simple. We will confirm the result of Section 1.2 that vector modes, most likely, are of no cosmological significance. This is because they remain small at all cosmological epochs (and decay in time at matter domination). Also, vector modes are not generated in simple inflationary models. On the other hand, tensor modes are of great interest, so we study them in detail. In the course of our study we encounter several properties which are inherent also in scalar perturbations, so our discussion of the simple case of tensor modes is useful from the general perspective as well. We emphasize, however, that neither vector nor tensor perturbations have been detected so far.

3.1 Vector Modes

Let us recall (see Section 1.2) that vector perturbations correspond to rotational excitations of cosmic medium that carry angular momentum. Energy density and pressure perturbations vanish for vector modes, see (2.63), while 3-velocity is normal to the wave vector \mathbf{k} . We also recall that V_i^T entering (2.64) and (2.65), are physical, rather than conformal velocities.

It follows from Eq. (2.64) that

$$(\rho + p)V_i^T = \frac{\text{const}}{a^4} \quad (3.1)$$

(hereafter const denotes a time-independent quantity which may be different in different equations). For barotropic fluid with $p \propto \rho$, the latter relation is interpreted as angular momentum conservation in the expanding Universe, see (1.22). In the case of multi-component medium, the relation (3.1) is valid for each of the components. Relativistic matter has $p \propto \rho \propto a^{-4}$, so that

$$V_i^T = \text{const.}$$

For non-relativistic matter, one has $p = 0$, $\rho \propto a^{-3}$, hence

$$V_i^T = \frac{\text{const}}{a}.$$

In either case, the velocity does not increase, and it decreases in the most interesting case.

We see from (2.65) that

$$\partial_\eta W_i^T = \frac{\text{const}}{a^2}.$$

Making use of (2.7), (2.8), (2.9) we find that metric perturbations decay as follows,

$$\begin{aligned} \text{RD epoch: } & W_i^T = \frac{\text{const}}{a} \\ \text{MD epoch: } & W_i^T = \frac{\text{const}}{a^{3/2}} \\ \text{AD epoch: } & W_i^T = \frac{\text{const}}{a^3} \end{aligned}$$

Hence, the metric perturbation always decreases in time, and the vector modes do not play a role in the standard cosmology. This completes the analysis of vector modes.

3.2 Tensor Modes: Relic Gravity Waves

The evolution of tensor modes is governed by Eq. (2.58). Let us write it in the momentum representation and omit the polarization index,

$$h'' + 2\frac{a'}{a}h' + k^2h = 0. \quad (3.2)$$

Super- and subhorizon perturbations obey $k \ll a'/a$ and $k \gg a'/a$ (or $k\eta \ll 1$ and $k\eta \gg 1$), respectively, see (2.83). It is clear from Eq. (3.2) that the superhorizon and subhorizon perturbations evolve in very different ways. In the former case, the last term in the left hand side is negligible, while in the latter case this term dominates.

3.2.1 Superhorizon perturbations: constant and decaying modes

Considering the superhorizon regime, we neglect the last term in Eq. (3.2) and write

$$h'' + 2\frac{a'}{a}h' = 0.$$

One of the solutions to this equation is independent of time,

$$h = h_{(i)} = \text{const.} \quad (3.3)$$

This is the constant mode; subscript (i) refers to the initial value of the amplitude. The second solution is the decaying mode,¹

$$h(\eta) = \text{const} \cdot \int \frac{d\eta}{a^2(\eta)}. \quad (3.4)$$

The decaying mode decreases in the same way as the metric perturbation in the vector mode W_i^T , namely, $h \propto a^{-1}, a^{-3/2}$ and a^{-3} at RD-, MD- and Λ D-epochs, respectively.

If the decaying mode was not exceedingly large very early at the Hot Big Bang epoch, it soon becomes negligibly small. Conversely, if this mode is substantial at the horizon entry, it was large at early times, i.e., the Universe was strongly inhomogeneous and anisotropic early on.² *We always assume in what follows that decaying modes are negligible at the time of the horizon entry.* Thus, we adopt the viewpoint that the Universe was homogeneous and isotropic at the earliest stages of the Hot Big Bang epoch, and that the primordial perturbations were built in already at that time. Our assumption concerns both tensor and scalar perturbations. To summarize, we totally ignore the decaying modes.

Our assumption belongs to the class of initial data serving as the input for the Hot Big Bang theory: there is no *a priori* reason for the Universe to be very homogeneous and isotropic at the beginning of its evolution. We discuss the initial data in Section 11, and here we emphasize that the question of whether the decaying modes are indeed negligible is to be answered by the cosmological observations. The existing observational data (albeit on scalar perturbations only) are consistent with our assumption. Briefly, the situation is as follows. First, those decaying modes which enter the horizon after recombination are indeed small at the horizon entry, otherwise they would be large at recombination, in contradiction to the CMB data. Second, the absence of the decaying modes of perturbations entering the horizon before recombination leads to well-defined phases of these perturbations; this property is confirmed by the data on the CMB temperature anisotropy and polarization. We come back to these issues later on.

Problem 3.1. Consider perturbations that enter the horizon after recombination. Let the amplitude of the decaying mode at recombination be of order $h \sim 10^{-6}$, which is not ruled out by observations. Find the amplitude of the decaying mode at the horizon entry for the present momentum $q_0 = (700 \text{ Mpc})^{-1}$.

¹This solution is sometimes called growing mode, since $h(\eta)$ grows as the scale factor *decreases* towards the singularity. We use the term “decaying mode” for superhorizon solutions which decay as the scale factor grows.

²This reasoning does not apply if the perturbations are generated shortly before the horizon entry or later than that. This is the case, e.g., for the topological defect mechanisms of the generation of the density perturbations, see Sections I.11.3 and I.11.5. Indeed, the properties of the perturbations generated by these mechanisms in a sense mimic the strong admixture of the decaying modes at the horizon entry. The observational data are inconsistent with such an admixture, and hence show that these and similar mechanisms either were not at work at all or were subdominant.

Problem 3.2. Let the amplitude of the decaying tensor mode of the present momentum $(100 \text{ Mpc})^{-1}$ be of order 10^{-6} at the horizon entry. Consider the evolution of this mode backwards in time and estimate at what temperature the Universe was strongly inhomogeneous and anisotropic.

3.2.2 Subhorizon perturbations. Matching to constant mode

To study the subhorizon regime, $k \gg a'/a$, we make use of the trick employed for analyzing the scalar field in Section I.2.3. We introduce a new unknown function

$$f(\eta) = a(\eta) \cdot h(\eta)$$

and write Eq. (3.2) as

$$f'' + k^2 f - \frac{a''}{a} f = 0. \quad (3.5)$$

This equation no longer involves the first derivative of the unknown function. The scale factor has power-law dependence on η , so $a''/a = \text{const}/\eta^2$ with constant of order 1. Since $k\eta \gg 1$, the last term in Eq. (3.5) can be neglected, and we come to the oscillator equation. Its general solution is

$$h(\eta) = \frac{C}{a(\eta)} \cos(k\eta + \alpha), \quad (3.6)$$

where C and α are time-independent amplitude and phase. Hence, after the horizon entry, the tensor perturbation describes gravity wave whose amplitude decays as $a^{-1}(\eta)$. The parameters C and α are to be found by matching to the constant mode (3.3). To begin with, let us perform this matching very crudely. The horizon entry occurs at the time $\eta_{\times} \simeq k^{-1}$. The function (3.6) is estimated at that time as $h \sim Ca^{-1}(\eta_{\times})$; equating it to the constant mode of the amplitude $h_{(i)}$ we find that after the horizon entry

$$h(\eta) = h_{(i)} \frac{a(\eta_{\times})}{a(\eta)} \cos(k\eta + \alpha).$$

If $h_{(i)}$ are the same for all wavelengths, then the dependence of the amplitude h on momentum at given time is determined by the time-dependence of the scale factor, $a(\eta_{\times}) \simeq a(k^{-1})$. For RD-entering modes (obeying $k \gg k^{(eq)}$, see Section 2.4)

$$h(k) \propto h_{(i)} k^{-1}, \quad k \gg k^{(eq)}, \quad (3.7)$$

while for MD-entering modes

$$h(k) \propto h_{(i)} k^{-2}, \quad k \ll k^{(eq)}. \quad (3.8)$$

To refine these estimates and find the phase α , we have to solve Eq. (3.2) at $\eta \sim \eta_{\times}$. For RD-entering modes we have at that time $a'/a = \eta^{-1}$, so Eq. (3.2) is

essentially the Bessel equation. Its solution that tends to a constant $h_{(i)}$ as $\eta \rightarrow 0$ (early times, superhorizon regime) is

$$h(\eta) = h_{(i)} \frac{\sin k\eta}{k\eta}. \quad (3.9)$$

Note that the phase in this solution is uniquely defined; of course, this is a consequence of our assumption of the absence of the decaying mode.

Problem 3.3. *Find the second solution to Eq. (3.2) for RD-entering mode. Show that this solution corresponds to the decaying mode (3.4).*

The solution (3.9) is valid at $\eta \sim \eta_\times$ only. To obtain $h(\eta)$ at later times, we write (3.9) in the following form,

$$h(\eta) = h_{(i)} \frac{a(\eta_\times)}{\eta_\times} \frac{\sin k\eta}{ka(\eta)}. \quad (3.10)$$

Here we made use of the fact that the ratio $a(\eta)/\eta$ is time-independent at $\eta \sim \eta_\times$. We find this ratio from (2.13b):

$$\frac{a(\eta_\times)}{\eta_\times} = a_0^2 H_0 \sqrt{\Omega_{rad}} \left(\frac{g_{*,0}}{g_*(\eta_\times)} \right)^{1/6}.$$

Since we know that the amplitude decays as a^{-1} at all times after horizon entry, the formula (3.10) gives the final expression for subhorizon RD-entering mode. As an example, at the present time

$$h(\eta_0) = h_{(i)} \frac{H_0 \sqrt{\Omega_{rad}}}{q_0} \left(\frac{g_{*,0}}{g_*(\eta_\times)} \right)^{1/6} \sin k\eta_0. \quad (3.11)$$

MD-entering modes enter the horizon when $a(\eta) \propto \eta^2$, so Eq. (3.2) near $\eta = \eta_\times$ has the form

$$h'' + \frac{4}{\eta} h' + k^2 h = 0 \quad (3.12)$$

This equation also reduces to the Bessel equation, and its solution that tends to $h_{(i)} = \text{const}$ as $\eta \rightarrow 0$ is

$$h(\eta) = h_{(i)} \cdot 3 \sqrt{\frac{\pi}{2}} \frac{1}{(k\eta)^{3/2}} J_{3/2}(k\eta) = -3h_{(i)} \cdot \frac{1}{(k\eta)^2} \left(\cos k\eta - \frac{\sin k\eta}{k\eta} \right).$$

At $k\eta \gg 1$ this gives

$$h(\eta) = -3h_{(i)} \frac{1}{(k\eta)^2} \cos k\eta, \quad (3.13)$$

which is again in accordance with (3.6), and also with (3.8). The phase is again fixed; it differs by $\pi/2$ from that of RD-entering modes.

Problem 3.4. *The same as in the previous problem, but for MD-entering mode.*

Problem 3.5. *Find the expression for $h(\eta)$ of MD-entering mode at all times, including the present epoch of accelerated expansion. Hint: Make use of the relation (2.14) valid for matter domination.*

Problem 3.6. *Show that the amplitudes of waves calculated according to (3.10) and (3.13) are of the same order of magnitude in the limit $t_x \rightarrow t_{eq}$. Hence, these formulas give one and the same order-of-magnitude estimate for the amplitudes of gravity waves entering the horizon at the equality epoch.*

Chapter 4

Scalar Perturbations: Single-Component Fluids

Cosmological medium in the real Universe is composed of several components: baryons, photons, neutrinos, dark matter, dark energy. At relatively late epoch, there is no interaction, except for gravitational, between different components. Before recombination, independent components are baryon-electron-photon plasma, dark matter and neutrinos; photons and baryons become two separate components after recombination. The study of perturbations in such a medium is rather involved, so we begin with simpler cases of single-component fluids. These cases are actually of direct relevance, as they describe perturbations in the component that dominates at a given cosmological epoch. As an example, perturbations in relativistic component play the major role at radiation domination;¹ likewise, perturbations of dark matter are most important at matter domination.

Perturbations in single-component fluid are described by Eqs. (2.74) and (2.76) with the label “tot” irrelevant. Equation (2.75) may be viewed as the relation determining the velocity potential v in terms of the gravitational potential Φ found by solving Eqs. (2.74) and (2.76), while Eqs. (2.77), (2.78) are automatically satisfied for solutions to Eqs. (2.74)–(2.76), if there is one component only.

4.1 Master Equation

Let us write Eqs. (2.74), (2.76) for single-component fluid, using momentum representation,

$$k^2\Phi + 3\frac{a'}{a}\Phi' + 3\frac{a'^2}{a^2}\Phi = -4\pi Ga^2 \delta\rho \quad (4.1)$$

$$\Phi'' + 3\frac{a'}{a}\Phi' + \left(2\frac{a''}{a} - \frac{a'^2}{a^2}\right)\Phi = 4\pi Ga^2 \delta p \quad (4.2)$$

In general, the equation of state for perturbations is

$$\delta p = u_s^2 \delta\rho. \quad (4.3)$$

¹Isocurvature perturbations discussed in Sections 6.4 and 6.5 are special in this regard.

where u_s is the sound speed which we consider arbitrary for the time being. By adding Eq. (4.1), multiplied by u_s^2 , to Eq. (4.2), we obtain the master equation for the gravitational potential,

$$\Phi'' + 3 \frac{a'}{a} (1 + u_s^2) \Phi' + \left[2 \frac{a''}{a} - \frac{a'^2}{a^2} (1 - 3u_s^2) \right] \Phi + u_s^2 k^2 \Phi = 0. \quad (4.4)$$

The behavior of its solutions depends, like in the case of tensor perturbations, on the relationship between the physical momentum $q(\eta) = k/a(\eta)$ and Hubble parameter $H(\eta)$. There is a qualification, however: an important role here is played by the combination $u_s q$. Namely, the last term in (4.4) is negligible at

$$u_s q = \frac{u_s k}{a} \ll H. \quad (4.5)$$

Conversely, the last term in (4.4) is of primary importance for $u_s q \gg H$. The inequality (4.5) can be written as

$$\lambda \gg u_s H^{-1},$$

where λ is the physical wavelength. The combination $u_s H^{-1}(\eta)$ is called *sound (acoustic) horizon*; it is of the order of the distance traveled by sound wave emitted right after the Big Bang singularity.

By making use of Eqs. (2.4) and (2.5), one expresses the quantity in square bracket in (4.4) in terms of the background pressure and energy density,

$$2 \frac{a''}{a} - \frac{a'^2}{a^2} (1 - 3u_s^2) = -8\pi G a^2 (p - u_s^2 \rho). \quad (4.6)$$

Let us assume that the cosmological expansion is dominated by the fluid under study (this is certainly true for the Universe filled with single-component fluid), and that the parameter w introduced in (2.79) is equal to u_s^2 , i.e.,

$$p = u_s^2 \rho,$$

and

$$\frac{\delta p}{p} = \frac{\delta \rho}{\rho}. \quad (4.7)$$

Note that the equality $w = u_s^2$ is valid in the real Universe throughout the Hot Big Bang epoch, except for the transition period from radiation to matter domination and recent accelerated expansion stage.

The quantity (4.6) vanishes for $w = u_s^2$, so Eq. (4.4) is simplified,

$$\Phi'' + 3 \frac{a'}{a} (1 + u_s^2) \Phi' + u_s^2 k^2 \Phi = 0. \quad (4.8)$$

Let us discuss perturbations of wavelengths exceeding the sound horizon, i.e., obeying (4.5). For negligible $u_s k$, Eq. (4.8) has constant solution

$$\Phi = \Phi_{(i)} = \text{const}, \quad (4.9)$$

and the solution that tends to infinity as $\eta \rightarrow 0$, the latter being the decaying mode. According to the assumption made in Section 3.2.1, the decaying mode vanishes, so the initial data for perturbations in the single-component fluid are given by (4.9). The energy density perturbation for large wavelength modes is determined by (4.1). Let us introduce the relative perturbation

$$\delta \equiv \frac{\delta\rho}{\rho}. \quad (4.10)$$

We find from (4.1), (4.9) and (2.4) that δ is time-independent for superhorizon modes,

$$\delta = \delta_{(i)} = -2\Phi_{(i)}. \quad (4.11)$$

Since the physical velocity is $\partial_i v$, we see from (2.75) that the velocity is small for these modes. We emphasize that the equality (4.11) is valid when $k \ll a'/a$, i.e., for wavelengths exceeding the true cosmological horizon, rather than the sound horizon. This qualification is particularly relevant for fluids with small sound speed (dark matter, baryons after recombination), whose sound horizon is much smaller than the true horizon.

Problem 4.1. *Find the relation between the velocity and gravitational potential $\Phi_{(i)}$ for modes of wavelengths larger than the sound horizon. Express this relation in terms of physical quantities. Show that the velocity is small compared to $\Phi_{(i)}$ only for $q \ll H$, i.e., when the wavelength exceeds the true horizon, rather than the sound horizon.*

The behavior of subhorizon modes depends strongly on the equation of state. We consider the two most relevant cases in the next Sections.

4.2 Relativistic Matter

The first important case is the perturbations of relativistic matter at radiation domination. In this case, $a \propto \eta$, $w = u_s^2 = 1/3$, so Eq. (4.8) reads

$$\Phi'' + \frac{4}{\eta}\Phi' + u_s^2 k^2 \Phi = 0.$$

We have already encountered this equation (see (3.12)); its solution obeying (4.9) as $\eta \rightarrow 0$ is

$$\begin{aligned} \Phi(\eta) &= \Phi_{(i)} \cdot 3\sqrt{\frac{\pi}{2}} \frac{1}{(u_s k \eta)^{3/2}} J_{3/2}(u_s k \eta) \\ &= -3\Phi_{(i)} \cdot \frac{1}{(u_s k \eta)^2} \left[\cos(u_s k \eta) - \frac{\sin(u_s k \eta)}{u_s k \eta} \right]. \end{aligned} \quad (4.12)$$

After entering the sound horizon, i.e., at $u_s k \eta \gg 1$, it describes the wave with decaying amplitude and definite phase,

$$\Phi(\eta) = -3\Phi_{(i)} \frac{1}{(u_s k \eta)^2} \cos(u_s k \eta). \quad (4.13)$$

When calculating the energy density perturbation, we note that the last term in the left hand side dominates in (4.1) and obtain

$$\delta\rho_{rad}(\eta) = -\frac{1}{4\pi G} \frac{k^2}{a^2} \Phi(\eta).$$

Let us recall the Friedmann equation (2.4), as well as the radiation domination relation

$$H^2 = \frac{1}{\eta^2 a^2}.$$

We also recall that $u_s^2 = 1/3$ and find from (4.13) that the relative perturbation (4.10) for subhorizon mode is

$$\delta_{rad}(\eta) = 6\Phi_{(i)} \cos(u_s k \eta) \quad (4.14)$$

$$= -3\delta_{rad,(i)} \cos(u_s k \eta), \quad (4.15)$$

where we made use of (4.11). We see that energy density perturbations undergo acoustic oscillations. Their amplitude does not decrease (but does not grow either) at radiation domination. The phenomenon analogous to the Jeans instability is absent at radiation domination because of the cosmological expansion.

By inserting the solution (4.12) into Eq. (2.75) we find the velocity perturbation for relativistic matter,

$$kv = \frac{3\Phi_{(i)}}{u_s} \left(\frac{\sin(u_s k \eta)}{(u_s k \eta)^2} - \frac{\cos(u_s k \eta)}{u_s k \eta} - \frac{1}{2} \sin(u_s k \eta) \right).$$

Recall (see Section 2.2.2), that $\mathbf{v} = kv$ is the physical velocity. As expected, velocity is small for subhorizon modes, $k\eta \ll 1$, namely,

$$kv = -\frac{1}{2}\Phi_{(i)} k \eta \ll \Phi_{(i)}.$$

On the other hand, subhorizon modes obey $k\eta \gg 1$, and we obtain

$$kv = -\frac{3\Phi_{(i)}}{2u_s} \sin(u_s k \eta). \quad (4.16)$$

Thus, the velocity undergoes acoustic oscillations whose phase is shifted by $\pi/2$ from the phase of the density oscillations, see (4.14). This is the well-known property of acoustic waves. The amplitude of the velocity perturbation is different by a factor of $1/(4u_s)$ from that of the density perturbation.

In fact, the oscillations of subhorizon modes can be seen from the continuity and Euler equations, Eqs. (2.77) and (2.78). For single-component fluid with $u_s^2 = w = 1/3$, these equations read

$$\begin{aligned}\delta' - \frac{4}{3}k^2v &= 4\Phi', \\ v' + \frac{1}{4}\delta &= -\Phi.\end{aligned}\tag{4.17}$$

The gravitational potential decays in the subhorizon regime, and the system (4.17) with $\Phi = 0$ reduces to the oscillator equation with the frequency $k/\sqrt{3}$; the relationship between the amplitudes of the velocity and density perturbations, $kv/\delta \sim \sqrt{3}/4$, and the phase shift by $\pi/2$ can also be seen from (4.17) with $\Phi = 0$.

4.3 Non-relativistic Matter

Let us study non-relativistic matter. Pressure is negligible, sound velocity is equal to zero, and the last term in Eq. (4.8) vanishes; in our terminology, the wavelength always exceeds the sound horizon. We still take constant mode as the initial condition and obtain that the gravitational potential does not depend on time at all,

$$\Phi(\eta) = \Phi = \text{const.}$$

It follows from (4.1) that the density perturbation is

$$\delta\rho = -\frac{1}{4\pi Ga^2} \left(k^2 + \frac{12}{\eta^2} \right) \Phi,\tag{4.18}$$

where we made use of the relation $a \propto \eta^2$, valid at matter domination. The second term in parenthesis dominates in the superhorizon regime, hence $\delta\rho \propto a^{-3}$. The background energy density is also proportional to a^{-3} , so the relative density perturbation is time-independent for superhorizon modes,

$$\delta = \delta_{(i)} = -2\Phi, \quad k\eta \ll 1.$$

The second equality here is obtained by making use of the matter domination relation $H = a'/a^2 = 2/(a\eta)$ together with the Friedmann equation. In the subhorizon regime, the first term in (4.18) dominates, and the density contrast grows in time,

$$\delta \propto a(\eta), \quad k\eta \gg 1.\tag{4.19}$$

This is the main mechanism of the perturbation growth that plays the role of the Jeans instability in the expanding Universe. We have already seen this behavior within the Newtonian theory in Section 1.2.

It follows from (4.18) and the Friedmann equation that subhorizon modes at matter domination obey

$$\delta = -\frac{2}{3} \frac{q^2(\eta)}{H^2(\eta)} \Phi = \frac{1}{3} \frac{q^2(\eta)}{H^2(\eta)} \delta_{(i)} = \frac{1}{3} \frac{a(\eta)}{a_\times} \delta_{(i)}, \quad k\eta \gg 1, \quad (4.20)$$

where the subscript \times refers to the horizon entry, when $q(\eta_\times) = H(\eta_\times)$. Clearly, long modes have smaller amplitudes at a given moment of time for momentum-independent $\delta_{(i)}$. This is of course due to the fact that they enter the horizon later.

Our analysis applies almost literally to dark matter perturbations which enter the horizon at matter domination.² The shortest of these perturbations have the present wavelengths of the order of several hundred Mpc, see (2.86). Formula (4.20) gives for their present amplitude,

$$\delta_{CDM,0} = \frac{1}{3} (1 + z_{eq}) \delta_{CDM,(i)}, \quad (4.21)$$

where $(1 + z_{eq}) \simeq 3000$ (in fact, the result (4.21) overestimates the amplitude, the reason being dark energy). The initial perturbations for these modes are of order $\delta_{CDM,(i)} \simeq 3 \cdot 10^{-5}$ (see Sections 5.4 and 6.1). Hence, the present density contrast at several hundred Mpc scale and beyond is small, $\delta \lesssim 0.03$. This is consistent with observations: the Universe is quite homogeneous at these scales.

With $\Phi(\eta) = \text{const}$, we find from (2.75) that the velocity is

$$kv = -\frac{\Phi}{3} k\eta. \quad (4.22)$$

As expected, the velocity is small in the superhorizon regime, $kv \ll \Phi_{(i)}$ for $k\eta \ll 1$. In subhorizon regime, velocity grows as $kv \propto \sqrt{a}$; this growth is milder than that of the density contrast (4.19).

4.4 Matter Perturbations at Λ Domination

The present composition of the Universe is the simplest example of multi-component fluid: the expansion rate is determined, to large extent, by dark energy, while perturbations exist in another component, non-relativistic matter. So, let us consider perturbations in non-relativistic matter at Λ domination, assuming that the dark energy density is constant in space and time (cosmological constant), $\delta\rho_\Lambda = \delta p_\Lambda = 0$. At Λ domination, the scale factor depends on conformal time according to (2.9).

²There are two effects that make this statement approximate only. First, the Universe undergoes dark energy driven accelerated expansion at the recent epoch. We consider the perturbations of non-relativistic matter at dark energy domination in Section 4.4. Second, non-relativistic component after recombination includes baryons. They are relatively important, since $\Omega_B/\Omega_{CDM} \approx 1/5$ is not negligibly small. The effect of baryons is studied in Section 7.1.

The relation (4.7) is no longer valid, so we solve directly Eq. (4.2). Making use of (2.9) we write Eq. (4.2) as follows,

$$\Phi'' - \frac{3}{\eta} \Phi' + \frac{3}{\eta^2} \Phi = 0 \quad (4.23)$$

Its solutions are

$$\Phi \propto \eta \propto \frac{1}{a} \quad \text{and} \quad \Phi \propto \eta^3 \propto \frac{1}{a^3}.$$

Let us now use Eq. (4.1) for subhorizon modes and recall that $\rho_M \propto a^{-3}$ for non-relativistic matter. Keeping the first solution for the gravitational potential (the second one rapidly decays) we find that density perturbations do not grow: we have for subhorizon modes

$$\delta_M \propto a^3 \frac{k^2}{a^2} \Phi = \text{const.}$$

Hence, the growth of perturbations terminates after the transition from matter to Λ domination. If the observed accelerated expansion is indeed the effect of the cosmological constant, structure formation process is close to its end and will never resume; structures of much larger sizes than existing today will never appear.

Problem 4.2. *Find the evolution of perturbations of non-relativistic matter in the unrealistic Universe whose expansion is dominated by negative spatial curvature.*

Problem 4.3. *Find the behavior of the matter velocity at Λ domination.*

Equation (4.23) is valid *deep* at Λ domination. This is not realistic for our Universe, since the energy density of matter is still considerable. Realistically, the scale factor depends on η according to (2.15). Gravitational potential still obeys Eq. (4.2) with $\delta p = 0$. Its solution can be found numerically. It is convenient to introduce the following notation,

$$\Phi(z, \mathbf{k}) \equiv g(z) \Phi_{MD}(\mathbf{k}), \quad (4.24)$$

where $\Phi_{MD}(\mathbf{k})$ is the gravitational potential deep at matter domination, when it is time-independent, and $g(z)$ is the suppression factor. Clearly, the function $g(z)$ obeys the same equation as $\Phi(z, \mathbf{k})$, i.e., Eq. (4.2). The initial data deep at matter domination are

$$g(z \rightarrow \infty) = 1, \quad \frac{dg}{dz}(z \rightarrow \infty) = 0. \quad (4.25)$$

The numerical solutions to Eq. (4.2) with $\delta p = 0$ and initial data (4.25) are shown in Fig. 4.1 for several values of Ω_Λ . Clearly, the function $g(z)$ is not very different from unity for realistic Ω_Λ . In particular, for $\Omega_\Lambda = 0.73$ we have

$$g(0) \approx 0.76. \quad (4.26)$$

To end this Section we point out that dark energy is not necessarily the cosmological constant; rather, it may have the equation of state $p_{DE} = w_{DE} \rho_{DE}$ with

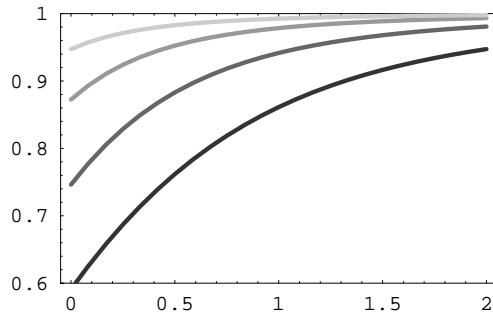


Fig. 4.1 The function $g(z)$ for spatially flat Universes with different values of the cosmological constant: $\Omega_\Lambda = 0.25, 0.5, 0.75, 0.9$; darker, lower lying curves correspond to larger Ω_Λ .

generally time-dependent $w_{DE} \neq -1$. If so, there are perturbations of dark energy itself, so the gravitational potential in the late Universe is different from that in the model with the cosmological constant. This effect is small, given the allowed range of w_{DE} , and we do not consider it in this book.

Chapter 5

Primordial Perturbations in Real Universe

In multi-component medium, gravitational interaction between the components affects perturbations in each of them. This is of great relevance for our Universe. The analysis of real scalar perturbations is, therefore, rather involved. We consider in this Chapter the issues related to the initial data for the evolution of scalar perturbations: what quantities are adequate for describing primordial scalar perturbations, and what types of primordial perturbations may exist in principle. As before, we assume that decaying modes are absent. The properties of the primordial perturbations known from the cosmological observations are summarized in Section 5.4. To orient the reader, we sketch the subsequent evolution of the adiabatic scalar perturbations in Section 5.5. More detailed analysis of this evolution is given in Chapters 6, 7, 8.

In this Chapter, the ideal fluid approximation is sufficient. Scalar perturbations in this approximation are described by the system of equations (2.74)–(2.78), which we write in the momentum representation for references:

- linearized Einstein equations,

$$k^2\Phi + 3\frac{a'}{a}\Phi' + 3\frac{a'^2}{a^2}\Phi = -4\pi Ga^2 \cdot \sum_{\lambda} \delta\rho_{\lambda}, \quad (5.1)$$

$$\Phi' + \frac{a'}{a}\Phi = -4\pi Ga^2 \cdot \sum_{\lambda} [(\rho + p)v]_{\lambda}, \quad (5.2)$$

$$\Phi'' + 3\frac{a'}{a}\Phi' + \left(2\frac{a''}{a} - \frac{a'^2}{a^2}\right)\Phi = 4\pi Ga^2 \cdot \sum_{\lambda} \delta p_{\lambda}; \quad (5.3)$$

- linearized covariant conservation of the energy-momentum tensor,

$$\delta\rho'_{\lambda} + 3\frac{a'}{a}(\delta\rho_{\lambda} + \delta p_{\lambda}) - (\rho_{\lambda} + p_{\lambda})(k^2v_{\lambda} + 3\Phi') = 0, \quad (5.4)$$

$$[(\rho_{\lambda} + p_{\lambda})v_{\lambda}]' + 4\frac{a'}{a}(\rho_{\lambda} + p_{\lambda})v_{\lambda} + \delta p_{\lambda} + (\rho_{\lambda} + p_{\lambda})\Phi = 0. \quad (5.5)$$

Here λ labels independent components of the cosmic fluid. We often use the fact that baryon-photon plasma before recombination is, to reasonable approximation, a single component; we sometimes use for it the notation $\lambda = B\gamma$. The notation $\lambda = CDM$ refers to dark matter. We recall (see Chapter 2) that not all equations of the above system are independent: for n -component fluid, this system has $2n + 3$ equations for $2n + 1$ unknowns.

It is convenient in what follows to use the relative perturbations of energy densities $\delta_\lambda = \delta\rho_\lambda/\rho_\lambda$ and write for each component

$$\delta\rho_\lambda = \rho_\lambda\delta_\lambda, \quad \delta p_\lambda = u_{s,\lambda}^2\delta\rho_\lambda = u_{s,\lambda}^2\rho_\lambda\delta_\lambda, \quad p_\lambda = w_\lambda\rho_\lambda.$$

Then the linearized conservation equations (5.4) and (5.5) take the form

$$\delta'_\lambda + 3\frac{a'}{a}(u_{s,\lambda}^2 - w_\lambda)\delta_\lambda - (1 + w_\lambda)k^2v_\lambda = 3(1 + w_\lambda)\Phi' \quad (5.6)$$

$$[(1 + w_\lambda)v_\lambda]' + \frac{a'}{a}(1 - 3w_\lambda)(1 + w_\lambda)v_\lambda + u_{s,\lambda}^2\delta_\lambda = -(1 + w_\lambda)\Phi \quad (5.7)$$

Here we used the relation (2.6) and accounted for the possibility that w_λ and $u_{s,\lambda}$ may depend on time, and that the sound speed squared $u_{s,\lambda}^2 = \delta p_\lambda/\delta\rho_\lambda$ may not coincide with the parameter $w_\lambda = p_\lambda/\rho_\lambda$.

5.1 Adiabatic and Isocurvature Modes

Initial data for scalar perturbations are defined deep at radiation domination. At that time, the modes of interest are superhorizon. They remain superhorizon at temperature, say, 1 MeV, see Section 2.4. So, we can always treat dark matter and baryons as non-relativistic,¹ and neutrino and dark matter as decoupled from baryon-electron-photon plasma. When studying superhorizon perturbations, we often consider formal limit $\eta \rightarrow 0$, although the relevant epoch is not really very early from the physics viewpoint.

Since the decaying mode is assumed to be absent, superhorizon modes depend on time weakly, if at all. We have already seen this, and will see again in what follows. Assuming that the cosmic medium does not contain exotic components, we can characterize it locally by four parameters:² temperature (or entropy density s), baryon number density n_B , number density of dark matter particles n_{CDM} and lepton number density n_L (lepton number at temperatures below m_e equals the excess of neutrinos over antineutrinos³). Accordingly, these four quantities as functions of spatial coordinates constitute the complete set of initial data; in the conformal Newtonian gauge, the gravitational potential is expressed through these quantities.

¹We assume that dark matter is cold. The case of warm dark matter is considered in Section 8.3.

²We simplify the situation here, neglecting possible mismatch between the velocities of neutrino and baryon-photon components, see the end of this Section.

³Due to neutrino oscillations, this excess is the same for all neutrino species, see Section I.7.2.

Possible initial data, and hence types (modes) of perturbations, are conveniently decomposed into linear independent parts: *adiabatic mode* and *isocurvature modes*. It is important here that we are dealing with linear theory: solutions to linear equations are linear in the initial data.

The adiabatic mode corresponds to the situation where relativistic matter, which is the dominant component at the early epoch, has non-vanishing energy density perturbations, i.e., temperature is inhomogeneous in space.⁴ Furthermore, the composition of cosmic medium is the same everywhere in space. This means that conserved quantities characterizing dark matter and baryons, n_B/s and n_{CDM}/s (see Section I.5.2) are independent of spatial coordinates. If there is lepton asymmetry, then n_L/s is also spatially homogeneous. Thus, the adiabatic mode in the superhorizon regime (formally, as $\eta \rightarrow 0$) obeys

$$\delta\left(\frac{n_B}{s}\right) = \delta\left(\frac{n_{CDM}}{s}\right) = \delta\left(\frac{n_L}{s}\right) = 0.$$

We will give somewhat more general definition of the adiabatic mode in Section 5.2.

On the contrary, the main property of baryon and CDM isocurvature modes is that there is no perturbation of the relativistic component deep at radiation domination, but the composition of the medium is spatially inhomogeneous (the definition of neutrino isocurvature modes needs qualification; we will briefly discuss this point at the end of this Section). Baryon isocurvature mode has inhomogeneous baryon number density, while temperature, dark matter density and lepton number densities are homogeneous in space. CDM isocurvature mode is characterized by inhomogeneous dark matter density and homogeneous other parameters, etc. Since the background values of n_B/s and n_{CDM}/s are homogeneous, the baryon and CDM isocurvature modes have the following properties in the formal limit $\eta \rightarrow 0$,

$$\begin{aligned} \text{Baryon isocurvature mode: } & \delta_B \neq 0, \quad \delta_{CDM} = 0, \quad \delta T = 0, \\ \text{CDM isocurvature mode: } & \delta_B = 0, \quad \delta_{CDM} \neq 0, \quad \delta T = 0. \end{aligned}$$

The term “isocurvature mode” reflects the fact that baryons and dark matter give small (vanishing in the limit $\eta \rightarrow 0$) contributions to the total energy density at radiation domination, so their inhomogeneities induce negligible gravitational potential Φ . This observation leads to the general definition of isocurvature modes in the conformal Newtonian gauge: the gravitational potential for these modes vanishes in the superhorizon regime deep at radiation domination. The latter definition is valid for neutrino isocurvature modes as well. We give the gauge-invariant definition in Section 5.2.

⁴This definition is not gauge invariant: one can choose hypersurfaces of constant time in such a way that temperature is spatially homogeneous at each moment of time. The latter gauge is different from conformal Newtonian, however. We introduce the gauge invariant characteristic of the adiabatic mode in Section 5.2.

Since the temperature is unperturbed at $\eta \rightarrow 0$ for baryon and CDM isocurvature modes, entropy density s is also homogeneous. The perturbations in the composition can be characterized by non-vanishing perturbations of entropy per baryon and entropy per dark matter particle $\delta(s/n_B)$ and $\delta(s/n_{CDM})$, respectively. Because of that, baryon and CDM isocurvature modes are also called *entropy modes*. Yet another term used for these perturbations is *isothermal modes*.

We will see in Section 6 that the adiabatic and isocurvature modes behave in very different ways, so that the cosmological observations are capable of discriminating between them. The data are consistent with the existence of the adiabatic mode only, while the bound on the admixture of the CDM isocurvature mode is at the level of a few percent, see Section 5.4; the bound on baryon isocurvature mode is weaker, since the baryon mass density is smaller than that of dark matter.

Adiabatic mode is very natural from the physics viewpoint. If dark matter and baryon asymmetry were generated in different regions of the Universe by the same physical mechanisms (e.g., by the mechanisms discussed in Chapters I.9 and I.11), one expects that the composition of the cosmic medium is the same in different parts of the Universe. This is precisely the property of the adiabatic mode. This observation is certainly valid if baryon asymmetry and dark matter were generated at the hot Big Bang epoch, and the cosmic plasma was in complete thermal equilibrium before their generation. Indeed, in that case the cosmic fluid was single-component at the early hot stage, so the only possible mode of perturbations was adiabatic. Since the laws of physics were the same everywhere, one and the same value of n_B/s was generated in all regions of the visible Universe, so the baryon isocurvature mode was not created. The same applies to dark matter. The generation of isocurvature modes needs something more exotic. Let us give an example by considering the Affleck–Dine mechanism of the baryon asymmetry generation, see Section I.11.6. The scalar field, which is at the core of this mechanism, may have different initial values in different places in the Universe; this inhomogeneity may be present due to the effects of inflation, see Chapter 14. If so, the baryon asymmetry n_B/s is spatially inhomogeneous too, which means the presence of the baryon isocurvature mode. We give other examples in Sections 14.2.2 and 14.4. Discovery of any of the isocurvature modes would be very important for cosmology and particle physics, since it would show that the generation of dark matter and/or matter-antimatter asymmetry has its roots at the stage preceding the hot stage of the cosmological evolution.

To end this Section, let us briefly discuss the possibility of neutrino isocurvature modes. In principle, there may be two of these modes [7]. One mode has neutrino-to-photon ratio inhomogeneous in space, with the perturbations in neutrino and photon components anticorrelating in such a way that there is no perturbation in the total energy density and hence the gravitational potential. This is *neutrino density isocurvature mode*. The second mode is called *neutrino velocity isocurvature mode*. This mode is characterized by non-trivial velocity potentials of neutrino and

photon components, which are different from each other and anticorrelate, summing up to zero total velocity of the cosmic medium.

Neutrinos were in thermal equilibrium with baryon-electron-photon plasma at temperatures above a few MeV, see Chapter I.7. Thus, if lepton asymmetry is not large, then the initial data for neutrino and photon components coincide, and the neutrino isocurvature modes are absent. Neutrino density isocurvature mode might in principle be sizeable for large lepton asymmetry of the Universe. However, under mild assumption that the maximum temperature exceeded 100 GeV, one expects that the lepton asymmetry is of the order of baryon asymmetry,

$$\frac{n_L}{s} \sim \frac{n_B}{s} \sim 10^{-10}. \quad (5.8)$$

The reason is that the lepton number can be converted into the baryon number at $T \gtrsim 100$ GeV, see Section I.11.2.1. We note, however, that (5.8) may not hold in some extensions of the Standard Model of particle physics, e.g., in models with sterile neutrino. In any case, fairly strong bound on lepton asymmetry follows from Big Bang Nucleosynthesis, see Section I.7.2. Hence, the neutrino density mode, if any, is most likely very small. It is even harder to generate the neutrino velocity mode. We do not study neutrino isocurvature modes in this book.

5.2 Adiabatic Mode in Superhorizon Regime

As we discussed in Section 2.4, the superhorizon regime intuitively corresponds to the picture of the Universe divided into independent homogeneous (and isotropic in the case of scalar perturbations) “universes” of superhorizon size that evolve independently and are characterized by somewhat different parameters. For the adiabatic mode, the composition of these “universes” is the same; they differ by energy density and hence scale factor. Heuristically, one can think of different superhorizon parts of the Universe starting their expansion at somewhat different time; some parts advance, and some lag behind the average cosmological evolution.⁵

This picture suggests that in the adiabatic mode, the perturbation of every component obeys

$$\delta\rho_\lambda = \rho'_\lambda \cdot \epsilon \quad (5.9)$$

$$\delta p_\lambda = p'_\lambda \cdot \epsilon, \quad (5.10)$$

where $\epsilon(\mathbf{x}, \eta)$ is the function common to all components, that depends on time and slowly varies in space. Heuristic, although not precise interpretation of ϵ is that it is the local advance time of the evolution of the “universe” centered at point \mathbf{x} . If there

⁵This picture is not gauge-invariant, but it is valid in the conformal Newtonian gauge at qualitative level.

is local thermal equilibrium, then temperature, and hence the energy density and pressure gradually change in space, and for homogeneous composition we have

$$\begin{aligned}\delta\rho_\lambda(\mathbf{x},\eta) &= \frac{\partial\rho_\lambda}{\partial T}\delta T(\mathbf{x},\eta) \\ \delta p_\lambda(\mathbf{x},\eta) &= \frac{\partial p_\lambda}{\partial T}\delta T(\mathbf{x},\eta).\end{aligned}$$

This is precisely the relation (5.9), (5.10) with

$$\epsilon(\mathbf{x},\eta) = \frac{\delta T(\mathbf{x},\eta)}{T'}. \quad (5.11)$$

We emphasize, however, that the relations (5.9), (5.10) are valid without the assumption of thermal equilibrium; they can be viewed as the *definition* of the adiabatic mode in the conformal Newtonian gauge.

It is not obvious, at first sight, that the Ansatz (5.9), (5.10) is consistent with the system of equations (5.1)–(5.5), especially in view of the requirement that Eqs. (5.4) and (5.5) must hold for every component of the cosmic medium separately. Indeed, we have only two unknowns ϵ and Φ , while the number of conservation equations may in principle be arbitrarily large. Let us see that the Ansatz (5.9), (5.10) does yield the solution to the system (5.1)–(5.5). In the first place, Eq. (5.5) can be viewed as the relation determining the velocity potential v_λ for each of the components. The velocity potential enters Eq. (5.4) in combination $k^2 v_\lambda$. The superhorizon regime is obtained in the formal limit $k \rightarrow 0$. Hence, the values of v_λ are irrelevant when solving Eqs. (5.1), (5.3) and (5.4). Let us now consider the conservation equation (5.4) that must hold for every component. Making use of (2.6) valid also for every component, we find that the conservation equation (5.4) reduces in the limit $k \rightarrow 0$ to one and the same equation for all λ ,

$$\Phi' = -\left(\frac{a'}{a}\epsilon\right)'.$$

This gives

$$\Phi = -\frac{a'}{a}\epsilon - \zeta, \quad (5.12)$$

where $\zeta(\mathbf{k})$ is independent of time. It follows from (5.9) and (2.6) that ζ is given by

$$\zeta = -\Phi + \frac{\delta\rho_{tot}}{3(\rho_{tot} + p_{tot})}. \quad (5.13)$$

Hence, the right hand side of this relation is independent of time for the superhorizon adiabatic mode. This is a very general fact independent of concrete properties of matter components [9]. It is worth noting that different terms in the sum in (5.13) may depend on time (this indeed happens at the transition from radiation to matter domination); the only time-independent combination is ζ itself.

Problem 5.1. Making use of Eqs. (5.1)–(5.5), express the time derivative $\dot{\zeta} \equiv \partial\zeta/\partial t$ through the physical momentum $q = k/a$, gravitational potential, Hubble parameter and their time derivatives. Hint: Make use of the relation

$$\frac{\delta p_{tot}}{\delta \rho_{tot}} = \frac{p'_{tot}}{\rho'_{tot}},$$

which is valid for superhorizon modes. Show that $\dot{\zeta}/\zeta$ is small compared to the expansion rate H for $q \ll H$.

Let us now see that the only time-independent quantity $\zeta(\mathbf{k})$ unambiguously determines the constant mode of the adiabatic perturbation in the superhorizon regime. To this end, we solve Eq. (5.1). Let us again take the formal limit $k \rightarrow 0$ and neglect the term with $k^2\Phi$ in that equation. Let us use (5.9) and (5.12) in Eq. (5.1). To simplify its right hand side we make use of (2.6), then (2.4) and (2.5), and express ρ'_{tot} in terms of the scale factor and its derivatives,

$$4\pi G a^2 \rho'_{tot} = 3 \frac{a'}{a} \left(\frac{a''}{a} - 2 \frac{a'^2}{a^2} \right).$$

As a result, Eq. (5.1) reduces to

$$\epsilon' + 2 \frac{a'}{a} \epsilon + \zeta = 0.$$

Its solution is

$$\epsilon(\eta) = -\zeta \cdot \frac{1}{a^2(\eta)} \int_0^\eta d\tilde{\eta} a^2(\tilde{\eta}). \quad (5.14)$$

The limit of integration here is chosen in such a way that $\epsilon(\eta)$ tends to zero as $\eta \rightarrow 0$, namely, $\epsilon \propto \eta$. This is necessary for finiteness of the potential (5.12) as $\eta \rightarrow 0$. The solution to the homogeneous equation, $\epsilon \propto a^{-2}$, corresponds to decaying mode which we assume to vanish. Equation (5.3) is then satisfied automatically.

Problem 5.2. Prove the last statement above.

Thus, Eqs. (5.1), (5.3) and (5.4) are indeed solved. As we have noticed, Eq. (5.5) determines then the velocity potentials. The latter are also related to ζ , and then Eq. (5.2) is satisfied automatically. We conclude that the function $\zeta(\mathbf{k})$ completely determines the superhorizon adiabatic mode.

It is useful to note that, in general, the quantity

$$\zeta_\lambda = -\Phi + \frac{\delta \rho_\lambda}{3(\rho_\lambda + p_\lambda)} \quad (5.15)$$

is time-independent for any component λ in the superhorizon regime. This follows from Eq. (5.4) in the limit $k \rightarrow 0$, once one recalls that $p'_\lambda = (\partial p_\lambda / \partial \rho_\lambda) \rho'_\lambda$, $\delta p_\lambda = (\partial p_\lambda / \partial \rho_\lambda) \delta \rho_\lambda$. The time-independence of ζ_λ in the superhorizon regime is the property of both adiabatic and isocurvature modes.

Instead of the parameter ζ defined by (5.13), one often uses the quantity

$$\mathcal{R} = -\Phi + \frac{a'}{a}v_{tot}, \quad (5.16)$$

where

$$v_{tot} = \frac{\sum(\rho_\lambda + p_\lambda)v_\lambda}{\sum(\rho_\lambda + p_\lambda)} \equiv \frac{[(\rho + p)v]_{tot}}{[\rho + p]_{tot}}. \quad (5.17)$$

These two are equal in the superhorizon regime, since Eqs. (5.1) and (5.2) give

$$\zeta - \mathcal{R} = \frac{\delta\rho_{tot}}{3[\rho + p]_{tot}} - \frac{a'}{a}v_{tot} = -\frac{1}{12\pi Ga^2[\rho + p]_{tot}}k^2\Phi.$$

The right hand side here is negligibly small for the superhorizon perturbations: it vanishes in the formal limit $k \rightarrow 0$.

It is now clear that the initial data for the adiabatic mode are adequately parameterized by $\zeta(\mathbf{k})$ or, equivalently, $\mathcal{R}(\mathbf{k})$. Note that we use these quantities for the modes in the superhorizon regime only. One of the major purposes of cosmological observations is to determine $\mathcal{R}(\mathbf{k})$ (and also the initial values of isocurvature perturbations, if any). Known to date properties of $\mathcal{R}(\mathbf{k})$ are summarized in Section 5.4.

Deep at radiation domination, one has $\delta\rho_{tot} = \delta\rho_{rad}$. In the limit $\eta \rightarrow 0$ the perturbations of relativistic matter obey (see (4.11))

$$\delta_{rad} = -2\Phi. \quad (5.18)$$

In the adiabatic mode, neutrino and photon components have the same density contrasts,

$$\delta_\gamma = \delta_\nu = -2\Phi. \quad (5.19)$$

It then follows from (5.13) that at radiation domination, when $p_{tot} = \rho_{tot}/3$, the gravitational potential is

$$\Phi = -\frac{2}{3}\zeta = -\frac{2}{3}\mathcal{R}. \quad (5.20)$$

The energy densities of non-relativistic and relativistic matter obey

$$\rho'_M = -3\frac{a'}{a}\rho_M, \quad \rho'_{rad} = -4\frac{a'}{a}\rho_{rad}.$$

Hence, the relation (5.9) gives the following equalities for the superhorizon adiabatic mode valid at any epoch,

$$\delta_{CDM} = \delta_B = \delta_M = \frac{3}{4}\delta_{rad} = \frac{3}{4}\delta_\gamma. \quad (5.21)$$

At radiation domination

$$\delta_{CDM} = \delta_B = \delta_M = \frac{3}{4}\delta_{rad} = \frac{3}{4}\delta_\gamma = -\frac{3}{2}\Phi = \mathcal{R}. \quad (5.22)$$

Hereafter the subscript M denotes the sum of contributions of baryons and cold dark matter, i.e., $\delta\rho_M = \delta\rho_B + \delta\rho_{CDM}$, etc. The latter relations determine the

initial data for the perturbations of each of the components in the adiabatic mode. We note that in terms of the parameters ζ_λ introduced in (5.15), the superhorizon adiabatic mode obeys

$$\zeta_{CDM} = \zeta_B = \zeta_\gamma = \zeta_\nu = \zeta.$$

These relations may be viewed as an alternative definition of the adiabatic mode.

On the contrary, (almost) gauge-invariant definition of the isocurvature modes is

$$\zeta_{tot} = \mathcal{R} = 0 \quad \text{as } \eta \rightarrow 0.$$

We discuss the initial data for these modes in more detail in Section 5.3.

Problem 5.3. *The relations (5.18), (5.20) and (5.22) are valid in the formal limit $\eta \rightarrow 0$. Assuming that the Universe contains only dark matter and radiation (relativistic matter), find δ_{rad} , δ_{CDM} , Φ and velocity potentials v_{rad} and v_{CDM} to the first non-trivial order in η . Express them in terms of $\mathcal{R}(\mathbf{k})$.*

Let us now discuss the physical meaning of ζ and \mathcal{R} . Their definitions in the gauge $h_{0i} = 0$ for arbitrary medium, not necessarily ideal fluid, are

$$\zeta = \Psi + \frac{\delta\rho_{tot}}{3(\rho_{tot} + p_{tot})}, \quad (5.23)$$

$$\mathcal{R} = \Psi + \frac{a'}{a}v_{tot}, \quad (5.24)$$

where v_{tot} is still defined by (5.17). In the ideal fluid approximation, we have $\Psi = -\Phi$, and get back to (5.13), (5.16).

Let us consider reference frame where the energy density is homogeneous in space at any moment of time, i.e., the reference frame whose equal-time hypersurfaces are hypersurfaces of constant energy density. We still keep $h_{0i} = 0$. The latter gauge choice means that modulo normalization, the proper time of matter elements is the same as the conformal time, $d\tau = f(\eta, \mathbf{x}) d\eta$. In other words, direction in space-time along the coordinate η coincides everywhere with the normal to the hypersurfaces of constant energy density. The coordinate transformation from the conformal Newtonian frame to the constant energy density frame is

$$\tilde{x}^\mu = x^\mu + \xi^\mu, \quad (5.25)$$

where $\xi^\mu(x^\nu)$ are small parameters obeying (2.33). Under this transformation, $h^{\mu\nu}(x)$ turns into $\tilde{h}^{\mu\nu}(x)$ given by (2.31) (recall that $h^{\mu\nu}$ and $\tilde{h}^{\mu\nu}$ are functions of one and the same argument, see details in Section I.A.6). It follows from the transformation law of the energy-momentum tensor that the linearized transformation of the energy density is

$$(\rho + \tilde{\delta}\rho)(\tilde{x}) = (\rho + \delta\rho)(x),$$

i.e.,

$$\tilde{\delta}\rho(x) = \delta\rho(x) - \rho' \xi^0. \quad (5.26)$$

We see that the transformation to the constant energy density frame, where $\tilde{\delta\rho}_{tot} = 0$, is given by

$$\xi^0 = \frac{\delta\rho_{tot}}{\rho'_{tot}}.$$

We now compare the transformation law for the metric (2.31) to the representation (2.48) and find that our gauge transformation induces non-zero $E(x)$ and, what is more important, transforms $\Psi(x)$ into

$$\begin{aligned}\tilde{\Psi} &= \Psi - \frac{a'}{a} \xi^0 \\ &= \Psi + \frac{\delta\rho_{tot}}{3(\rho + p)_{tot}},\end{aligned}\tag{5.27}$$

where we made use of the conservation law (2.6). We see that $\zeta(x)$ coincides with the gravitational potential $\tilde{\Psi}$ at hypersurfaces of constant energy density. Let us finally calculate the *3-dimensional* curvature scalar for spatial metric

$$dl^2 = a^2(\delta_{ij} - h_{ij}) dx^i dx^j = [\delta_{ij}(1 + 2\tilde{\Psi}) - \partial_i \partial_j E] dx^i dx^j$$

(the sign in front of h_{ij} follows from the definition (2.26)). To the linear order, the curvature scalar is (cf. (B.8))

$$R^{(3)} = -a^{-2} (\partial_i \partial_j h_{ij} - \Delta h) = -\frac{4}{a^2} \Delta \tilde{\Psi}.$$

Thus, the physical meaning of ζ is that it determines the spatial curvature of hypersurfaces of constant energy density,⁶

$$\text{Hypersurfaces of constant } \rho_{tot}: \quad R^{(3)} = -\frac{4}{a^2} \Delta \zeta.\tag{5.28}$$

There is an alternative interpretation of ζ . Making use of the gauge transformation (5.27), one can choose the gauge $\tilde{\Psi} = 0$. Spatial curvature $R^{(3)}$ vanishes in such a frame. Each of the quantities (5.15) is in this frame

$$\text{Hypersurfaces } R^{(3)} = 0: \quad \zeta_\lambda = \frac{\delta\rho_\lambda}{3(\rho_\lambda + p_\lambda)} = -\frac{\delta\rho_\lambda}{H\dot{\rho}_\lambda}.\tag{5.29}$$

In particular, for relativistic matter in thermal equilibrium

$$\text{Hypersurfaces } R^{(3)} = 0: \quad \zeta_{rad} = \frac{\delta s}{3s},$$

where $s \propto \rho_{rad}^{3/4}$ is the entropy density. So, $\zeta = \zeta_{rad}$ is the relative perturbation of the entropy density in the frame with spatially flat hypersurfaces at radiation domination. We note that the relation (5.29) is valid for all perturbations, including isocurvature.

⁶Note that this interpretation is valid irrespective of whether the mode is super- or subhorizon, if ζ is understood as the time-dependent quantity. We do not use $\zeta(\mathbf{k}, \eta)$ in the subhorizon regime.

The physical meaning of \mathcal{R} is that it determines the curvature of equal-time hypersurfaces of comoving reference frame, namely,

$$\text{Hypersurfaces } v_{tot} = 0: \quad R^{(3)} = -\frac{4}{a^2} \Delta \mathcal{R}. \quad (5.30)$$

This is also the general property of scalar perturbations.

The relations (5.28) and (5.30) no longer refer to the conformal Newtonian gauge. This suggests that there are gauge invariants that reduce in conformal Newtonian gauge to ζ and \mathcal{R} defined by (5.23) and (5.24). In fact, ζ is gauge-invariant by itself, while the gauge-invariant definition of \mathcal{R} is constructed in a fairly straightforward way.

Problem 5.4. Show by direct calculation that ζ defined by (5.23) is invariant under arbitrary gauge transformations generated by the small coordinate transformations (5.25).

Problem 5.5. Show that \mathcal{R} defined by (5.24) is invariant under gauge transformations consistent with the gauge condition $h_{0i} = 0$. Give the gauge-invariant definition of \mathcal{R} , such that (5.24) is restored for $h_{0i} = 0$. Hint: Recall that $u^i = a^{-1} v^i = a^{-1} \partial_i v$ are components of 4-velocity.

Problem 5.6. Consider comoving reference frame, where $v^i = 0$ and also $h_{0i} = 0$. Show that the formula (5.30) is indeed valid in this frame.

5.3 Initial Data for Isocurvature Modes

We have seen in Section 5.2 that the gauge-invariant quantity ζ_λ , defined by (5.15), is time-independent in the superhorizon regime for every component λ of the cosmic medium. Isocurvature modes are such that the contributions of different components to \mathcal{R} cancel each other, $\mathcal{R} = 0$. The same holds for ζ defined for the whole medium by (5.13). Thus, it is natural to consider pairwise differences of ζ_λ , rather than ζ_λ themselves. According to the tradition, one defines

$$\mathcal{S}_{\lambda, \lambda'} = 3(\zeta_\lambda - \zeta_{\lambda'}) = \frac{\delta_\lambda}{1 + w_\lambda} - \frac{\delta_{\lambda'}}{1 + w_{\lambda'}}.$$

Furthermore, one usually chooses photons as one of the components and works with $\mathcal{S}_{CDM, \gamma}$, $\mathcal{S}_{B, \gamma}$, etc. We omit the subscript γ and write $\mathcal{S}_\lambda \equiv \mathcal{S}_{\lambda, \gamma}$. Hence, baryon and CDM isocurvature modes we consider in this book are characterized by the following quantities (recall that baryons and dark matter particles are non-relativistic at temperatures of interest),

$$\mathcal{S}_\lambda = \delta_\lambda - \frac{3}{4} \delta_\gamma, \quad \lambda = B, \quad CDM.$$

These parameters are time-independent in the superhorizon regime and they fully determine the initial data for isocurvature modes.

Since the entropy density perturbations obey

$$\frac{\delta s}{s} = \frac{\delta n_\gamma}{n_\gamma} = \frac{3}{4} \delta_\gamma,$$

the parameter \mathcal{S}_λ at radiation domination is

$$\mathcal{S}_\lambda = \frac{\delta(n_\lambda/s)}{(n_\lambda/s)}, \quad \lambda = B, \text{ CDM}. \quad (5.31)$$

Perturbations of relativistic component vanish at early times in the conformal Newtonian gauge and in most other conventionally used gauges. Hence, there are perturbations of baryon or CDM number densities only, while the temperature is homogeneous in space. As we have already mentioned, these perturbations are also called entropy modes, which is fully consistent with (5.31).

5.4 Primordial Spectra: Results from Observations

In this Section we summarize the existing knowledge about primordial perturbations. The properties of *scalar* perturbations are fairly well-known from observations; we discuss some observational aspects in Chapters 7, 9 and 10. On the contrary, *tensor* perturbations have not been detected yet, so assumptions on their properties are based solely on theoretical prejudice. As we have seen in Section 3.1, vector modes rapidly decrease in time, and they most likely are irrelevant; we are not going to discuss them.

Let us begin with scalar perturbations. Observational data show that they are adiabatic, and the decaying mode is indeed absent. The admixture of isocurvature modes is rather strongly constrained, we quantify this point below. The observations are consistent with the property that the primordial field $\mathcal{R}(\mathbf{x})$ that specifies the initial data for the adiabatic mode, is the *Gaussian random field*; definition and properties of Gaussian random fields are given in Appendix C. This field is completely determined by the two-point correlation function

$$\langle \mathcal{R}(\mathbf{k}) \mathcal{R}(\mathbf{k}') \rangle = \frac{P_{\mathcal{R}}(k)}{(2\pi)^3} \delta(\mathbf{k} + \mathbf{k}'). \quad (5.32)$$

Hence, if \mathcal{R} is indeed the Gaussian random field, one of the major tasks for the observational cosmology is to determine the scalar function — the power spectrum $P_{\mathcal{R}}(k)$.

Let us pause here to discuss the averaging in the left hand side of (5.32). In the theory of random fields, one considers averages over an ensemble of systems, each system having a certain realization of the random field. In our case, the system is the entire Universe. Hence, the left hand side of (5.32) is the average over the ensemble of Universes like ours. Since our Universe is unique, the random variable like $\mathcal{R}(\mathbf{k})$ (or $\mathcal{R}(\mathbf{x})$) takes a definite value which is *a priori* unknown. Still, the properties of this random field can be studied, say, by measuring correlators $\langle \mathcal{R}(\mathbf{x}) \mathcal{R}(\mathbf{x}') \rangle$ (more

precisely, observables related to them) for many points \mathbf{x}, \mathbf{x}' at a given distance from each other. Procedures of this sort are indeed used in the observational cosmology. One should keep in mind that they are plagued, at least in principle and sometimes in practice, by irreducible statistical uncertainty — *cosmic variance* — which is due to the fact that we observe one Universe, and hence deal with one realization of random field. We encounter an example of cosmic variance in Section 9.

Let us continue with the discussion of the scalar adiabatic perturbations. It is convenient to introduce the quantity

$$\mathcal{P}_{\mathcal{R}}(k) = \frac{k^3}{2\pi^2} P_{\mathcal{R}}(k), \quad (5.33)$$

which is also called power spectrum. It follows from (5.32) (see also Section C.2), that the fluctuation of the random field $\mathcal{R}(\mathbf{x})$ has a simple form,

$$\langle \mathcal{R}^2(\mathbf{x}) \rangle = \int_0^\infty \frac{dk}{k} \mathcal{P}_{\mathcal{R}}(k). \quad (5.34)$$

Hence, $\mathcal{P}_{\mathcal{R}}(k)$ is the contribution to the fluctuation coming from a decimal interval of momenta around k . The spectrum for which this contribution is independent of k , i.e., $\mathcal{P}_{\mathcal{R}}(k) = \text{const}$, is called flat or Harrison–Zeldovich spectrum [10, 11]. We note that one often uses the notation

$$\Delta_{\mathcal{R}}^2(k) \equiv \mathcal{P}_{\mathcal{R}}(k).$$

The parameter $\Delta_{\mathcal{R}}(k)$ characterizes the *amplitude* of perturbations of momentum k , whereas $\mathcal{P}_{\mathcal{R}}$ gives the amplitude squared.

It follows from observations that the scalar spectrum is nearly flat. The natural parameterization in this situation is power-law, which is traditionally written as

$$\mathcal{P}_{\mathcal{R}}(k) = A_{\mathcal{R}} \left(\frac{k}{k_*} \right)^{n_s - 1}. \quad (5.35)$$

Here k_* is some conveniently chosen fiducial momentum (e.g., the team of WMAP experiment has chosen $k_*/a_0 = 0.002 \text{ Mpc}^{-1}$), the parameter $A_{\mathcal{R}} = \mathcal{P}_{\mathcal{R}}(k_*)$ is power at $k = k_*$, and $(n_s - 1)$ is the spectral tilt. Flat spectrum has $n_s = 1$; spectra with $n_s < 1$ and $n_s > 1$ are “red” and “blue” (enhanced at large and small wavelengths), respectively.

Generally, the power spectrum need not have simple power-law behavior. To account for this possibility, one introduces yet another parameter, *running index*, $\frac{dn_s}{d \log k} \equiv \frac{dn_s}{d \log k}(k_*)$ by writing

$$\mathcal{P}_{\mathcal{R}}(k) = A_{\mathcal{R}} \left(\frac{k}{k_*} \right)^{n_s - 1 + \frac{dn_s}{d \log k} \log \frac{k}{k_*}}. \quad (5.36)$$

Running index is the rate of the spectral index variation near $k = k_*$. The current values⁷ obtained by fitting the entire set of cosmological data *under the assumption of no tensor perturbations and no momentum-dependence of the tilt* are [2]

$$A_{\mathcal{R}} = (2.46 \pm 0.09) \cdot 10^{-9}, \quad n_s = 0.960 \pm 0.014 \quad (5.37)$$

for $k_*/a_0 = 0.002 \text{ Mpc}^{-1}$. Hence, the primordial scalar amplitude is

$$\Delta_{\mathcal{R}} \simeq 5 \cdot 10^{-5}.$$

The result (5.37) suggests that the spectrum may not be exactly flat. This evidence, however, becomes weaker if one allows for tensor perturbations. If the assumption of momentum-independence of the tilt is relaxed, the bound on this dependence is [2]

$$-0.07 < \frac{dn_s}{d \log k} < 0.01, \quad \text{C.L. 95\%.} \quad (5.38)$$

The interval of allowed values of $n_s(k_*)$ is extended in that case.

The final point about the adiabatic scalar perturbations is that the initial values of all other quantities (density contrasts, gravitational potentials and velocity potentials) are linearly related to \mathcal{R} in the conformal Newtonian gauge. Hence, these initial values are also Gaussian random fields. In particular, the definition of the power spectrum of the initial gravitational potential $P_\Phi(\mathbf{k})$ is analogous to (5.32),

$$\langle \Phi_{(i)}(\mathbf{k}) \Phi_{(i)}(\mathbf{k}') \rangle = \frac{P_\Phi(k)}{(2\pi)^3} \delta(\mathbf{k} + \mathbf{k}'). \quad (5.39)$$

The quantity P_Φ is defined in complete analogy with (5.33). The relation between the gravitational potential and \mathcal{R} at radiation domination is given by (5.20). Therefore, the primordial power spectrum for the gravitational potential is, in the power-law approximation,

$$P_\Phi = A_\Phi \left(\frac{k}{k_*} \right)^{n_s - 1}, \quad (5.40)$$

where

$$A_\Phi = \frac{4}{9} A_{\mathcal{R}}.$$

It follows from (5.37), that the power of gravitational potential at momentum scale $k_*/a_0 = 0.002 \text{ Mpc}^{-1}$ is

$$A_\Phi = P_\Phi(k_*) = 1.1 \cdot 10^{-9}. \quad (5.41)$$

It is independent of momenta for flat spectrum.⁸

⁷Unless stated otherwise, the error intervals correspond to 68% confidence level.

⁸As a curiosity, the power $A_{\mathcal{R}}$ is of the same order as the baryon-to-photon ratio η_B . All known mechanisms of the generation of density perturbations, on the one side, and baryon asymmetry, on the other, are completely unrelated to each other. Hence, the above fact is probably yet another coincidence between cosmological parameters of different origin.

Let us turn to isocurvature modes. Their initial data are parameterized by the quantities \mathcal{S}_{CDM} and \mathcal{S}_B defined by (5.31) (recall that we do not consider neutrino modes in this book). By analogy to (5.35) one may expect that if these modes exist, their spectrum is nearly flat too. This is indeed the case in models we discuss in Sections 14.2 and 14.4. With this assumption, these modes have a certain power $A_{\mathcal{S}}$ and tilt. The isocurvature and adiabatic modes may be totally uncorrelated, or, alternatively, there may be correlation or anticorrelation between them. The total correlation between adiabatic and, say, CDM mode would mean that CDM-to-entropy ratio is larger in precisely the same places where the temperature is higher. The anticorrelation would mean the opposite.⁹ The models exhibiting correlation and anticorrelation between the adiabatic and CDM modes is given in Section 14.2. Here we introduce the parameter β characterizing the correlation properties,¹⁰

$$\beta = \frac{\mathcal{P}_{RS}}{\sqrt{\mathcal{P}_R \mathcal{P}_S}}$$

where \mathcal{P}_{RS} is defined in analogy to (5.34),

$$\langle \mathcal{R}(\mathbf{x}) \mathcal{S}(\mathbf{x}) \rangle = \int_0^\infty \frac{dk}{k} \mathcal{P}_{RS}(k).$$

Observational limits on the admixture of isocurvature modes depend on β , i.e., on the correlation properties of perturbations. For $\beta = 0$ (no correlation) and $\beta = 1$ (complete correlation) the bounds on the CDM mode are [2]

$$\frac{\mathcal{P}_{S_{CDM}}}{\mathcal{P}_R} < 0.07, \quad \beta = 0, \quad \frac{\mathcal{P}_{S_{CDM}}}{\mathcal{P}_R} < 0.004, \quad \beta = 1, \quad (5.42)$$

at 95% confidence level. As we already noted, bounds on baryon isocurvature mode expressed in terms of $\mathcal{P}_{S_B}/\mathcal{P}_R$ are weaker, because the baryon mass density is smaller than that of CDM. These bounds are obtained from (5.42) by multiplying the right hand sides by $(\Omega_{CDM}/\Omega_B)^2 \sim 20$.

Finally, let us consider tensor perturbations. According to (2.57), they are given by the sum over polarizations

$$h_{ij}^{TT}(\eta, \mathbf{k}) = \sum_{A=+, \times} e_{ij}^{(A)} h^{(A)}(\eta, \mathbf{k}). \quad (5.43)$$

In analogy to the scalar case one assumes that $h_{(i)}^{(A)}$ are the independent Gaussian random fields whose two-point correlation functions are

$$\left\langle h_{(i)}^{(A)}(\mathbf{k}) h_{(i)}^{(B)}(\mathbf{k}') \right\rangle = \frac{1}{2} \delta_{AB} \cdot \frac{P_T(k)}{(2\pi)^3} \delta(\mathbf{k} + \mathbf{k}'), \quad (5.44)$$

where the left hand side is again the average over the ensemble of Universes. Unlike for scalar perturbations, the assumption that $h_{(i)}^{(A)}$ are the Gaussian fields and

⁹Some papers use the opposite terminology: what we call correlation is dubbed anticorrelation, and vice versa.

¹⁰In view of the previous footnote, some papers use β defined with the opposite sign.

the expectation concerning the shape of their spectrum are based on theoretical prejudice. Most inflationary models predict Gaussianity, and yield the power-law spectrum

$$\mathcal{P}_T = A_T \left(\frac{k}{k_*} \right)^{n_T}, \quad (5.45)$$

with $|n_T| \ll 1$. Here the power spectrum $\mathcal{P}_T \equiv \Delta_T^2$ is defined in complete analogy to (5.33), so that

$$\sum_A \langle (h_{(i)}^{(A)}(\mathbf{x}))^2 \rangle = \int_0^\infty \frac{dk}{k} \mathcal{P}_T(k).$$

To characterize the relative power of the tensor and scalar perturbations, one introduces the ratio

$$r = \frac{A_T}{A_R}.$$

The observational bound on this ratio is [2]

$$r < 0.2, \quad 95\% \text{ C.L.} \quad (5.46)$$

It is expected that the sensitivity of future observations to tensor perturbations will increase substantially as instruments designed specifically to study CMB polarization are put into operation.

We leave aside the important issue of possible deviations from Gaussianity. These deviations may in principle be of very different types; there is no unambiguous way to parameterize them. One of the numerous possibilities is that the primordial gravitational potential $\Phi_{(i)}$ of the adiabatic perturbations is *local in space*, but non-linear function of the Gaussian field Φ_L . Restricting to non-derivative relation between $\Phi_{(i)}$ and Φ_L , one writes to the lowest non-trivial order in Φ_L [12]

$$\Phi_{(i)}(\mathbf{x}) = \Phi_L(\mathbf{x}) + f_{NL} \Phi_L(\mathbf{x}) \Phi_L(\mathbf{x}), \quad (5.47)$$

where f_{NL} is a dimensionless parameter. The non-Gaussianity of roughly this type emerges in the inflaton (Sections 13.1, 13.2) and curvaton (Section 14.2.1) mechanisms of the generation of scalar perturbations. In these cases, the non-linear term in (5.47) is due to both non-linearity of the equation for the relevant scalar field and non-linear relation between scalar field fluctuations and gravitational potential. The inflaton mechanism yields small non-Gaussianity, $f_{NL} \lesssim 1$; this is, generally speaking, not the case in curvaton models.

Since the amplitude of scalar perturbations is small, $\Phi \sim 10^{-5}$, *relative* contribution of the second term in (5.47) is of order $10^{-5} f_{NL}$. This means that the effect is very small for $f_{NL} \sim 1$. The current limit on the non-Gaussianity of the form (5.47) is [2]

$$-9 < f_{NL} < 111, \quad 95\% \text{ C.L.}$$

Detailed discussion of non-Gaussianity of cosmological perturbations is given in Ref. [8].

5.5 Evolution of Adiabatic Perturbations: A Preview

Before going to the detailed study, let us sketch the evolution of the adiabatic scalar perturbations in the real Universe and related phenomena.

From the viewpoint of structure formation, *dark matter* perturbations are of primary interest. They produce gravitational wells into which baryons fall. Soon after recombination, dark and conventional matter evolve together, and in the end form galaxies, clusters of galaxies, etc.

The evolution of dark matter perturbations can be understood at the qualitative level by making use of the results of Sections 1.2 and 4.3. As we know from Section 4.3, large wavelength perturbations that enter the horizon at matter domination have small amplitudes at present and evolve in the linear regime. The same holds for modes entering the horizon shortly before radiation-matter equality. Hence, the Universe is homogeneous at the scales of about 100 Mpc and larger; there are no gravitationally bound structures of masses exceeding $10^{16} M_\odot$ (the relationship between the distance and mass scales is given in Section 1.3). Dark matter perturbations of shorter wavelengths enter the horizon at radiation domination. They grow logarithmically at that epoch, $\delta_{CDM}(\eta) \propto \log \eta/\eta_\times$, where $\eta_\times \simeq k^{-1}$ is the horizon entry time, see Section 1.2. Later on, at matter domination, the growth becomes linear in the scale factor, see Section 4.3, and at that epoch

$$\delta_{CDM}(\mathbf{k}, z) \simeq C\Phi_{(i)}(\mathbf{k}) \frac{1 + z_{eq}}{1 + z} \log(0.2k\eta_{eq}), \quad (5.48)$$

where $\Phi_{(i)}$ is the initial gravitational potential that completely determines the initial data for adiabatic perturbation in the conformal Newtonian gauge (the numerical factor in the argument of logarithm is obtained in Section 6.2.1). The numerical constant C here turns out to be fairly large, $|C| \simeq 27/2$. The latter property is due to the enhancement of perturbation at both horizon crossing and radiation-matter equality, see Section 6.2.1. Making use of this fact, one finds that the perturbations with $k/a_0 \gtrsim (0.2a_0\eta_{eq})^{-1} \sim (20 \text{ Mpc})^{-1}$ are not small at the present epoch. Indeed, the amplitude of the primordial gravitational potential is $\Delta_\Phi = \sqrt{\mathcal{P}_\Phi} \sim 3 \cdot 10^{-5}$, see (5.41); setting $\log(0.2k\eta_{eq}) \sim 1$ for an estimate, we obtain $\delta_{CDM,0} \sim 1$ for $k/a_0 \sim (20 \text{ Mpc})^{-1}$. Somewhat simplifying the situation, we conclude that the modes with $k/a_0 \sim (10 \text{ Mpc})^{-1}$ enter the non-linear regime at the present epoch; the corresponding length and mass scales are $\lambda_0 \simeq a_0\pi/k \sim 30 \text{ Mpc}$ and $10^{15} M_\odot$, see Section 1.3. Due to the logarithmic factor in (5.48), modes of even shorter wavelengths entered the non-linear regime and collapsed earlier; as an example, intense¹¹ formation of galaxies with $M \sim (10^{11} - 10^{12}) M_\odot$ began at $z \sim 4$. Thus, the formula (5.48) suggests that structures were formed in the hierarchical manner: smaller structures were formed earlier, while larger structures have been formed in the process of clustering of smaller ones. This general picture, as well as the

¹¹Since the field of perturbations is random, some galaxies were formed even earlier, see Chapter 7.

quantitative theory of structure formation, are in agreement with the observational data.

From the viewpoint of CMB properties, of great interest are perturbations of baryons and photons at recombination (more precisely, at photon last scattering epoch). Inhomogeneities of the baryon-photon component, its motion and gravitational potentials existing at that epoch give rise to CMB temperature anisotropy and polarization. One (but not the only) effect is that regions of higher density of baryons and photons have higher temperature,

$$\frac{\delta T}{T}(\mathbf{x}) \propto \delta_{B\gamma}(\mathbf{x}).$$

A wave of perturbation of conformal momentum k , that existed at recombination epoch with $\eta = \eta_r$, is seen today at angle $\Delta\theta \propto (k\eta_0)^{-1}$. Indeed, $2\pi/k$ is the coordinate wavelength, and $(\eta_0 - \eta_r) \approx \eta_0$ is the coordinate distance traveled by photon since last scattering; their ratio is precisely the angle $\Delta\theta$. In the decomposition of CMB temperature in spherical harmonics (see Sections I.1.2.5 and 9.1), the angular scale $\Delta\theta$ corresponds to angular harmonic $l \propto (\Delta\theta)^{-1}$. Hence, crudely speaking, there is an approximate correspondence

$$C_l \longleftrightarrow \delta_{B\gamma}^2(k, \eta_r), \quad k \simeq \frac{l}{\eta_0}, \quad (5.49)$$

where C_l are CMB temperature multipoles (recall that C_l are quadratic in $\delta T/T$, see formulas in Sections I.1.2.5 and 9.1).

Large wavelength perturbations enter the horizon after recombination. They are time-independent up to last scattering epoch, and proportional to the initial gravitational potential (cf. Section 5.2),

$$\delta_{B\gamma}(\mathbf{k}) \propto \Phi(\mathbf{k}) \propto \Phi_{(i)}(\mathbf{k}).$$

For flat primordial spectrum, they induce almost flat CMB anisotropy spectrum at $l \lesssim \eta_0/\eta_r \simeq 50$: C_l depend on l rather weakly. On the other hand, we know from Section 4.2 that smaller wavelength perturbations oscillate at radiation domination,

$$\delta_{B\gamma}(\eta) = \text{const} \cdot \cos \left(k \int_0^\eta u_s d\eta' \right), \quad (5.50)$$

where $u_s \simeq 1/\sqrt{3}$ is the sound speed (it slightly decreases towards recombination due to increasing baryon contribution to the energy density). We emphasize that the phase of these acoustic oscillations is well-defined; this is both because the perturbations are adiabatic, and due to the absence of the decaying mode. The acoustic oscillations continue at matter domination until recombination. It follows from (5.50) that at last scattering, perturbations in the baryon-photon component contain the contribution that oscillates as a function of momentum,

$$\delta_{B\gamma}(\eta_r) = \text{const} \cdot \cos(kr_s), \quad (5.51)$$

where $r_s = \int_0^{\eta_r} u_s d\eta$ is the coordinate size of the sound horizon, $r_s \approx \eta_r/\sqrt{3}$. According to (5.49) this leads to oscillations of C_l as function of l . The period of the oscillations of C_l (distance between the peaks) is determined by half-period of oscillations of $\delta_{B\gamma}$ in k and is estimated as

$$\Delta l \simeq \frac{\pi}{r_s} \eta_0 \simeq \sqrt{3}\pi \frac{\eta_0}{\eta_r} \simeq 300.$$

This picture is qualitatively consistent with the CMB data.

Oscillations in the baryon-photon component at recombination, given by (5.51), are present in the initial data for the subsequent evolution of baryons. They leave their trace in the distribution of structure in the late Universe (it is important here that Ω_B/Ω_{CDM} is not negligibly small). This effect is called *baryon acoustic oscillations*. As we point out in Section 7.1.2, they are very similar to the Sakharov oscillations in the cold Universe model which was popular before the discovery of CMB.

The neutrino component is special. Neutrinos have high velocities during considerable part of the cosmological evolution. They escape gravitational potentials and hence do not participate in structure formation. At the same time, neutrinos give sizeable contribution to the energy density. These two properties lead to the suppression of dark matter perturbations (as well as baryon perturbations after recombination), especially at relatively small scales. This effect depends on neutrino masses; it is used for establishing the cosmological bounds on these masses.

We give rather detailed treatment of the above aspects of the cosmological perturbations in subsequent Chapters. Here we simply note that quantitative results on structures and CMB depend both on the properties of primordial perturbations and on the cosmological parameters like h , Ω_B , Ω_{CDM} , Ω_Λ , Ω_{curv} , dark energy equation of state, etc. Hence, the observational data on structures and CMB are the major source of information about our Universe.

This page is intentionally left blank

Chapter 6

Scalar Perturbations before Recombination

This Chapter is devoted to the evolution of scalar perturbations before recombination. Of particular relevance here are perturbations in dark matter and baryon-photon components, as well as gravitational potentials. The detailed study of the period before recombination has several implications. First, recombination (more precisely, photon last scattering epoch) is the period at which CMB photons decouple from baryonic matter, so perturbations at that period are directly probed by the CMB observations. Second, perturbations of dark matter and baryons at recombination serve as initial data for their further evolution that leads to structure formation. Finally, peculiarities of perturbations in baryons at recombination leave small but observable trace in the distribution of matter in the Universe. This trace — baryon acoustic oscillations — has been detected.

We leave the study of the evolution of dark matter and baryons after recombination for Chapter 7; properties of neutrino component are discussed in Section 8.4.

As we have already seen, superhorizon perturbations can be studied analytically. In other words, the initial data for the evolution are understood in analytical terms. On the other hand, the evolution itself is fairly complex, and its precise analysis is based on numerical methods. Although the equations governing the evolution well before structure formation are linear, their complete analytical solution is impossible. Furthermore, the effects beyond the ideal fluid approximation are described by fairly cumbersome system of the Boltzmann equations. All these difficulties can be overcome by numerical analysis. In this book we restrict ourselves to approximate analytical study which is useful for understanding the physics behind various phenomena. We quote the results of numerical studies wherever appropriate.

The evolution of perturbation strongly depends on its wavelength (more precisely, on its conformal momentum). According to Section 2.4, we consider separately perturbations that enter the horizon at radiation and matter domination (RD- and MD-entering). The intermediate wavelengths are important from the viewpoint of CMB properties at angular scales of a few degrees. However, the

analytical study of these wavelengths is difficult, so we discuss them at the qualitative level only. Since recombination and radiation-matter equality are not very much separated in conformal time, $\eta_r/\eta_{eq} \approx 2.35$, we are unable to study quantitatively the modes that enter the horizon between these two epochs. Finally, we cannot give quantitative account of the baryon-photon perturbations which enter the *sound* horizon between η_{eq} and η_r . We recall in this regard that the sound speed in this component is of order $1/\sqrt{3}$ (we refine this estimate later on), so our analysis is at qualitative level for modes with momenta in the interval $\sqrt{3}/\eta_{eq} \gtrsim k \gtrsim 1/\eta_r$.

As we have already noted, the cosmic medium is not an ideal fluid. This property is particularly important for baryon-electron-photon and neutrino components; some of the effects are considered in Chapter 8. Dark matter particles may also have considerable velocities of chaotic motion at the epoch of interest; in that case we are dealing with warm or hot dark matter, which is also not ideal fluid. We consider the latter possibility in Section 8.3.1, and in this Chapter we study cold dark matter. For obtaining simple formulas, we sometimes have to make the approximation of strong dark matter domination over baryons,

$$\rho_{CDM} \gg \rho_B. \quad (6.1)$$

This is a reasonable but not at all perfect approximation, since $\Omega_B \simeq 0.04$, $\Omega_{CDM} \simeq 0.2$.

Everywhere in this Chapter we make use of the generalization of the ideal fluid approximation for the baryon-electron-photon component before photon last scattering. It is called *tight coupling*; the results concerning deviations from the tight coupling approximation are simply quoted in appropriate places. We define what is meant by the tight coupling approximation later in the Chapter, and here we note that its applicability is due to the fact that the interaction between baryons, electrons and photons is strong enough. This approximation has its limit of applicability which is discussed in Section 8.5. Here we quote that it is indeed valid for sufficiently long waves, $k \lesssim 30/\eta_r \simeq 1500/\eta_0$.

It is useful to note that at recombination and earlier, baryon-photon component is relativistic to good accuracy: the ratio of baryon energy (mass) density to that of photons at recombination and equality is (see Section 6.2 for details)

$$\frac{\rho_B}{\rho_\gamma}(t_r) \simeq 0.85 \quad \text{and} \quad \frac{\rho_B}{\rho_\gamma}(t_{eq}) = \frac{1 + z_r}{1 + z_{eq}} \frac{\rho_B}{\rho_\gamma}(t_r) \simeq 0.30,$$

respectively. Hence, the dominant contribution to the energy of the baryon-photon component comes from photons. We continue to label this component by subscript $B\gamma$ and sometimes use the relativistic approximation, i.e., set $w_{B\gamma} = u_{s,B\gamma}^2 = 1/3$. However, the deviation of the equation of state from that of relativistic matter is important in some places; we give more accurate analysis wherever necessary (see Section 6.2).

6.1 Adiabatic Modes of Large Wavelengths

The behavior of the adiabatic perturbations entering the horizon at matter domination is rather simple. Recall that the present momenta of these modes are smaller than $q_0^{(eq)}$, see (2.85). The value of ζ (and \mathcal{R}) for such a mode remains constant at the initial stage of matter domination, when the mode is still superhorizon. At that time, the non-relativistic matter perturbations already dominate. They obey (see (4.11))

$$\delta_M = -2\Phi. \quad (6.2)$$

It follows from (5.13) and $\rho_{tot} = \rho_M$, $p_{tot} = p_M = 0$, that the relation between the gravitational potential and ζ (and \mathcal{R}) for superhorizon mode at matter domination is

$$\Phi = -\frac{3}{5}\zeta = -\frac{3}{5}\mathcal{R} = \frac{9}{10}\Phi_{(i)}. \quad (6.3)$$

This gives $\delta_M = 6\mathcal{R}/5$. Notice that the gravitational potential for these modes changes at radiation-matter equality: Φ decreases by a factor of 9/10. When the mode is still superhorizon, perturbation in photon energy density obeys (5.21), i.e., at matter domination

$$\delta_\gamma = \frac{8}{5}\mathcal{R} = -\frac{12}{5}\Phi_{(i)}. \quad (6.4)$$

We will see in Chapter 9 that formulas (6.3) and (6.4) determine the (dominant) contribution of adiabatic scalar perturbations to CMB temperature anisotropy at large angular scales.

At matter domination, the evolution of modes studied here proceeds in accordance with Section 4.3: relativistic matter gives small contribution to the energy density, and corrections due to its presence in the Universe are small. After the horizon entry, the density contrast δ_M grows linearly with the scale factor, but the large wavelength modes never become non-linear.

It follows from (5.4), (5.5) and from time-independence of the gravitational potential that the velocity of superhorizon modes is negligible at matter domination, $v \simeq 0$.

6.2 Adiabatic Modes Entering the Sound Horizon at Radiation Domination

In this Section, we study in detail the modes that enter the *sound* horizon at radiation domination. These modes obey $k \lesssim \sqrt{3}/\eta_{eq}$. As we have already pointed out, the intermediate range of momenta $\sqrt{3}/\eta_{eq} \gtrsim k \gtrsim 1/\eta_r$ is difficult to study analytically, so we discuss it at the qualitative level. This is done in Section 6.3.

The evolution of perturbations in the relativistic component at radiation domination proceeds as described in Section 4.2. The gravitational potential for mode of conformal momentum k is given by (4.12) where the initial value is (see (5.20))

$$\Phi_{(i)} = -\frac{2}{3}\zeta = -\frac{2}{3}\mathcal{R}. \quad (6.5)$$

In what follows, we express the results in terms of the initial value $\Phi_{(i)}$ of the gravitational potential in the conformal Newtonian gauge, with understanding that it is related to gauge-invariant initial data by Eq. (6.5).

The new phenomenon is that this potential induces dark matter perturbation after the horizon entry, which grows logarithmically at radiation domination. This effect is crucial for formation of galaxies and other gravitationally bound systems in our Universe: as we have seen in Chapter 4, perturbations in relativistic matter do not grow at radiation domination, and with $\Delta_{\mathcal{R}} = 5 \cdot 10^{-5}$, the growth of perturbations at matter domination is by itself insufficient for producing $\delta_M \sim 1$. The results for modes under study are illustrated in Fig. 6.1.

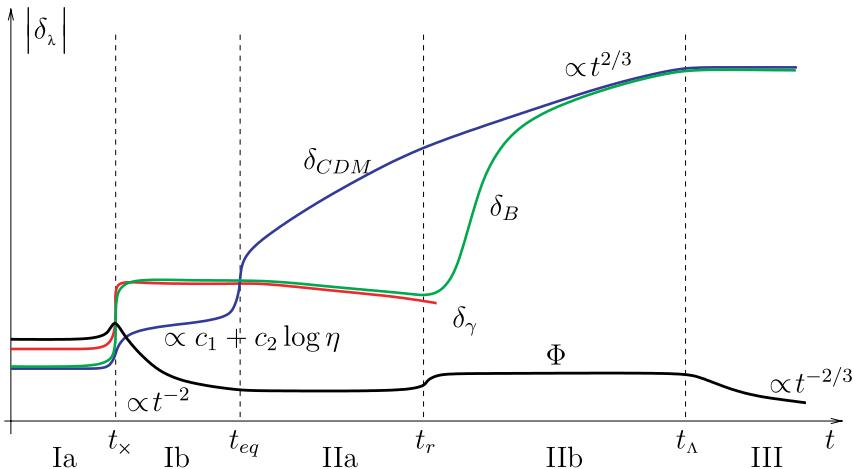


Fig. 6.1 Schematic plot of the evolution of adiabatic scalar perturbation of a given conformal momentum k that enters the sound horizon at radiation domination. Oscillations are not shown. t_x is the time of the sound horizon entry. The behavior near the times ($t_x, t_{eq}, t_r, t_\Lambda$) is shown qualitatively. The scales of the axes are arbitrary. I. Radiation domination. Perturbation of relativistic component dominates; it behaves as described in Section 4.2; Ia: before the horizon entry, perturbation does not evolve, its properties are given by Eq. (5.22); Ib: near the horizon entry, perturbation in relativistic component induces logarithmically growing dark matter perturbation, Eq. (6.11). II. Matter domination. CDM perturbation dominates and evolves according to Section 4.3. Initial data for this evolution is given by (6.11); IIa: perturbation in baryon-photon component before recombination — sound wave — oscillates according to (6.36); CDM perturbation induces additional perturbation in the baryon-photon component, formula (6.35); IIb: baryons are decoupled, their perturbation catch up with CDM perturbation, formulas (7.19), (7.18). III. Λ -domination; growth of perturbations terminates, Section 4.4.

Perturbations in the baryon-photon component continue to oscillate at matter domination until recombination. This leads to observable consequences, see Section 7.1.2 and Chapters 9, 10. There is yet another effect relevant for CMB. Namely, the gravitational potential produced by CDM perturbation induces, in turn, extra perturbation in baryon-photon component. This effect is important between radiation-matter equality and recombination. It is studied in Section 6.2.2.

The two cases in which perturbations in the dominant component induce perturbations in the subdominant one are described by making use of the covariant conservation equations (5.6) and (5.7), applied to the subdominant component. The unknown functions $\delta_\lambda(\eta)$ and $v_\lambda(\eta)$ entering these equations refer to the subdominant component, while the potential Φ and the cosmological expansion rate are determined by the dominant one and obtained by solving the independent set of equations for the latter.

6.2.1 Dark matter perturbations

Let us now give the detailed analysis. We consider first the adiabatic mode at radiation domination. The behavior of the dominant relativistic component is given by formulas of Section 4.2. Let us study dark matter perturbations indicated by the potential (4.12). We set $w = u_s^2 = 0$ in Eqs. (5.6) and (5.7) and cast them in a simple form,

$$\delta'_{CDM} - k^2 v_{CDM} = 3\Phi' \quad (6.6)$$

$$v'_{CDM} + \frac{1}{\eta} v_{CDM} = -\Phi, \quad (6.7)$$

where we have used the fact that $a \propto \eta$ at radiation domination, see (2.7). The potential (4.12) decreases rather rapidly at late times, so at $u_s k \eta \gg 1$ (i.e., deep inside the sound horizon; hereafter u_s denotes the sound speed in the baryon-photon component unless stated otherwise), Eqs. (6.6), (6.7) become homogeneous,

$$\delta'_{CDM} - k^2 v_{CDM} = 0 \quad (6.8)$$

$$v'_{CDM} + \frac{1}{\eta} v_{CDM} = 0. \quad (6.9)$$

The solution to Eq. (6.9) is

$$v_{CDM} = \frac{c_1}{k^2 \eta},$$

where c_1 is dimensionless constant. Then Eq. (6.8) gives (cf. (1.27))

$$\delta_{CDM} = c_1 \log k \eta + c_2,$$

Hence, in accordance with Section 1.2, CDM perturbations, generally speaking, grow logarithmically at radiation domination.

We note, however, that the logarithmically growing part in δ_{CDM} would be absent, if not for the sources in the right hand sides of Eqs. (6.6), (6.7). Indeed, this part would diverge as $\eta \rightarrow 0$, i.e., it would be the decaying mode of primordial perturbations; we assume that the decaying mode vanishes. In the presence of the gravitational potential produced by perturbation in relativistic matter, the logarithmic part does exist after the sound horizon crossing. To find the constants c_1 and c_2 , let us first solve Eq. (6.7) with the gravitational potential given by (4.12). The solution, finite as $\eta \rightarrow 0$, is

$$v_{CDM} = -\frac{1}{\eta} \int_0^\eta d\tilde{\eta} \tilde{\eta} \Phi(\tilde{\eta}),$$

so that $v_{CDM} \rightarrow 0$ as $\eta \rightarrow 0$. Then we obtain the solution to Eq. (6.6),

$$\delta_{CDM}(\eta) = 3\Phi(\eta) + c - k^2 \int_0^\eta \frac{d\hat{\eta}}{\hat{\eta}} \int_0^{\hat{\eta}} d\tilde{\eta} \tilde{\eta} \Phi(\tilde{\eta}).$$

where c is a constant. We express the latter in terms of $\delta_{CDM,(i)}$ and $\Phi_{(i)}$ and change the order of integration. The result is

$$\delta_{CDM}(\eta) = \delta_{CDM,(i)} + 3(\Phi(\eta) - \Phi_{(i)}) - k^2 \int_0^\eta d\tilde{\eta} \tilde{\eta} \Phi(\tilde{\eta}) \log \frac{\eta}{\tilde{\eta}}.$$

The potential (4.12) rapidly decreases as $u_s k \eta \gg 1$, the integral converges, and we obtain in this regime

$$\delta_{CDM}(\eta) = \delta_{CDM,(i)} - 9\Phi_{(i)} \cdot \left[\log(u_s k \eta) + \mathbf{C} - \frac{2}{3} \right], \quad (6.10)$$

where $\mathbf{C} = 0.577\dots$ is the Euler constant. Here we recalled that $u_s^2 = 1/3$ and used the values of the integrals

$$\sqrt{\frac{\pi}{2}} \int_0^\infty J_{3/2}(z) \frac{dz}{\sqrt{z}} = 1, \quad \sqrt{\frac{\pi}{2}} \int_0^\infty J_{3/2}(z) \log z \frac{dz}{\sqrt{z}} = 1 - \mathbf{C}.$$

It is clear from (6.10) that the main effect at large $u_s k \eta$ is indeed the logarithmic growth of dark matter perturbations; their initial value $\delta_{CDM,(i)}$ is not particularly important. Making use of (5.22) and (5.20), we finally obtain for the subhorizon adiabatic mode at radiation domination

$$\delta_{CDM}(\eta) = -9\Phi_{(i)} \left(\log \frac{k\eta}{\sqrt{3}} + \mathbf{C} - \frac{1}{2} \right). \quad (6.11)$$

CDM perturbation gets enhanced considerably as compared to its initial value $\delta_{CDM,(i)} = -(3/2)\Phi_{(i)}$.

In turn, dark matter perturbation generates extra contribution to the gravitational potential. At large $k\eta$ it is determined by (5.1), where time-derivatives can be neglected for $u_s k \eta \gg 1$. We denote this extra contribution by Φ_{CDM} , since it is

due to dark matter perturbation. We find

$$\Phi_{CDM}(\eta) = -\frac{a^2(\eta)}{k^2} 4\pi G \rho_{CDM} \delta_{CDM}(\eta). \quad (6.12)$$

This contribution is small at radiation domination as compared to (4.12) because ρ_{CDM} is small. However, the potential produced by dark matter becomes *dominant after radiation-matter equality*.

To estimate dark matter density perturbation at matter domination (i.e., when dark matter itself is the dominant component), we recall the results of Section 4.3. Namely, the gravitational potential gets frozen in, while the density perturbation grows as $a(\eta)$. Roughly speaking, this growth begins from the value given by (6.11) with $\eta = \eta_{eq}$. More precise estimate is (we give the derivation later on; see problem 6.1 for the estimate of the numerical coefficient in the argument of logarithm, this estimate is refined in (6.26))

$$\delta_{CDM}(\eta) = -\frac{27}{2} \frac{a(\eta)}{a_{eq}} \Phi_{(i)} \log(0.2k\eta_{eq}) \quad (6.13)$$

Note that this result shows the enhancement by a factor of $3/2$ as compared to the naive estimate obtained by simple matching with (6.11) at $\eta = \eta_{eq}$. It follows from (6.12) that the gravitational potential for this perturbation is

$$\begin{aligned} \Phi_{CDM} &= \frac{27}{2} \Phi_{(i)} 4\pi G \rho_{CDM} \frac{a^2}{k^2} \frac{a}{a(\eta_{eq})} \log(0.2k\eta_{eq}) \\ &= \frac{81}{4} \Phi_{(i)} \frac{H_0^2 a_0^2}{k^2} \Omega_{CDM} (1 + z_{eq}) \log(0.2k\eta_{eq}). \end{aligned} \quad (6.14)$$

Another form for the latter expression, convenient for some applications, is obtained by making use of (2.16),

$$\Phi_{CDM} = \Phi_{(i)} \cdot 81 \cdot I^2 \cdot \frac{\Omega_{CDM}}{\Omega_M} (1 + z_{eq}) \frac{\log(0.2k\eta_{eq})}{(k\eta_0)^2}. \quad (6.15)$$

The gravitational potential (6.14) is time-independent, as it should. Note that the approximation used for obtaining (6.14) is valid for $u_s k \eta_{eq} \gg 1$ only, and the argument of logarithm in (6.14) must exceed 1. Hence, the sign of Φ_{CDM} coincides with the sign of $\Phi_{(i)}$ at both radiation and matter domination.

We see that the above mechanism of the generation of dark matter perturbation at radiation domination is quite efficient. It leads to both logarithmic enhancement and extra numerical factor as compared to perturbations that enter the horizon at matter domination.

Let us note that formulas (6.13) and (6.14) are only approximate at recombination. This is due to the fact that recombination occurs relatively soon after radiation-matter equality, so the regime $\Phi_{CDM} = \text{const}$, $\delta_{CDM} \propto a(\eta)$ has not yet been established by recombination. Nevertheless, the results (6.13) and (6.14) are reasonably accurate and we use them in further estimates.

We also note that the results (6.13), (6.14) are valid in the approximation of negligible baryon contribution to both average energy density and matter perturbations, i.e., in the formal limit $\Omega_B/\Omega_{CDM} \rightarrow 0$. We briefly consider the effect of baryons before recombination later in this Section, and study the evolution of baryons and dark matter after recombination in Section 7.1.

To obtain the formula (6.13) we note that the potential (6.12) evolves fairly slowly near $\eta = \eta_{eq}$. Perturbation in the relativistic component induced by this potential is small: it follows from (5.5) that this contribution to $\delta_{B\gamma}$ is of order Φ_{CDM} , while Eq. (5.1) gives

$$\delta_{tot} \sim \frac{k^2}{a^2 H^2} \Phi_{CDM} \gg \Phi_{CDM}.$$

Hence, we can neglect perturbation in the relativistic component when calculating δ_{CDM} near $\eta \sim \eta_{eq}$. Then dark matter perturbation is governed by Eqs. (5.6), (5.7). Taking into account the gravitational potential, they read

$$\delta'_{CDM} - k^2 v_{CDM} = 3\Phi'_{CDM}, \quad (6.16a)$$

$$v'_{CDM} + \frac{a'}{a} v_{CDM} = -\Phi_{CDM}. \quad (6.16b)$$

We recall that $k^2 \Phi_{CDM} \gg \Phi''_{CDM}$ for modes we study, and make use of (6.12). This gives the equation for δ_{CDM} alone,

$$\delta''_{CDM} + \frac{a'}{a} \delta'_{CDM} - 4\pi G \rho_{CDM} a^2 \delta_{CDM} = 0. \quad (6.17)$$

this equation is valid at both $\eta \sim \eta_{eq}$ and $\eta \gg \eta_{eq}$. Its solution should match (6.11) at $\eta \ll \eta_{eq}$.

It is convenient to solve (6.17) by introducing, instead of η , the variable

$$x = \frac{a}{a_{eq}}.$$

Once the baryon contribution to the energy density is neglected, one has

$$\frac{8\pi}{3} G \rho_{CDM}(\eta_{eq}) = \frac{8\pi}{3} G \rho_{rad}(\eta_{eq}) = \frac{1}{2} H^2(\eta_{eq}) \equiv \frac{1}{2} H_{eq}^2.$$

Hence, at all relevant times

$$4\pi G \rho_{CDM} = \frac{3}{4} H_{eq}^2 \frac{1}{x^3}, \quad 4\pi G \rho_{rad} = \frac{3}{4} H_{eq}^2 \frac{1}{x^4}.$$

Equations (2.4) and (2.5) can now be written as

$$\frac{a'^2}{a^2} = \frac{1}{2} H_{eq}^2 a_{eq}^2 \left(\frac{1}{x} + \frac{1}{x^2} \right), \quad \frac{a''}{a} = \frac{1}{4} H_{eq}^2 a_{eq}^2 \frac{1}{x}.$$

Making use of these relations, we cast Eq. (6.17) into the form

$$x(1+x)\frac{d^2\delta_{CDM}}{dx^2} + \left(1 + \frac{3}{2}x\right)\frac{d\delta_{CDM}}{dx} - \frac{3}{2}\delta_{CDM} = 0. \quad (6.18)$$

One of the solutions to this equation is

$$\delta_{CDM}^{(1)} = C_1 \left(1 + \frac{3}{2}x\right); \quad (6.19)$$

another is then straightforwardly obtained,

$$\begin{aligned} \delta_{CDM}^{(2)} &= C_2 \left(1 + \frac{3}{2}x\right) \int_{\infty}^x \frac{d\tilde{x}}{\tilde{x}\sqrt{(1+\tilde{x})(1+3\tilde{x}/2)^2}} \\ &= C_2 \left(1 + \frac{3}{2}x\right) \left(\log \frac{\sqrt{1+x}-1}{\sqrt{1+x}+1} + 6\frac{\sqrt{1+x}}{2+3x}\right). \end{aligned} \quad (6.20)$$

These expressions give at small x , with logarithmic accuracy,

$$\delta_{CDM} = C_1 + C_2 \log x = C_1 + C_2 \log \frac{\eta}{\eta_{eq}}.$$

We compare this with (6.11) and obtain, also with logarithmic accuracy,

$$C_2 = -9\Phi_{(i)}, \quad C_1 = -9\Phi_{(i)} \log k\eta_{eq}.$$

The part proportional to C_2 is small at large x (late times), and δ_{CDM} is

$$\delta_{CDM} = C_1 \cdot \frac{3}{2}x.$$

This gives (6.13) with logarithmic accuracy.

Problem 6.1. Show that modulo corrections of order $(k\eta_{eq})^{-1}$, a_{eq}/a and Ω_B/Ω_{CDM} , the dark matter density contrast at matter domination is given by

$$\begin{aligned} \delta_{CDM}(\eta) &= -\frac{27}{2} \frac{a(\eta)}{a_{eq}} \Phi_{(i)} \left[\log(k\eta_{eq}) + \mathbf{C} - \frac{7}{2} + \log \frac{2}{\sqrt{6} - \sqrt{3}} \right] \\ &= -\frac{27}{2} \frac{a(\eta)}{a_{eq}} \Phi_{(i)} \log(0.15k\eta_{eq}). \end{aligned} \quad (6.21)$$

Hint: Use the results of Section 2.1.2. Corrections due to finite Ω_B/Ω_M modify the numerical coefficient in the argument of logarithm, see below; this modification is accounted for in (6.13).

The results (6.13) and (6.14) have been obtained in the approximation of negligible baryon contribution to the average energy density at matter domination. Let us show that this contribution yields slight suppression of the dark matter density contrast and gravitational potential in comparison with (6.13) (6.14). We modify the calculation after

(6.17) and obtain, instead of (6.18),

$$x(1+x)\frac{d^2\delta_{CDM}}{dx^2} + \left(1 + \frac{3}{2}x\right)\frac{d\delta_{CDM}}{dx} - \frac{3}{2}\frac{\Omega_{CDM}}{\Omega_M}\delta_{CDM} = 0, \quad (6.22)$$

where $x = a/a_{eq}$. The latter equation differs from (6.18) by the factor Ω_{CDM}/Ω_M in the last term. This factor originates from the fact that when expressing ρ_{CDM} in (6.17) by making use of the Friedmann equation we did not neglect the baryon contribution to total energy density. Equation (6.22) admits analytical solution. In terms of the variable $y = -x$ it is reduced to the hypergeometric equation, see Ref. [13]. The general solution to (6.22) is a linear combination of two solutions

$$\begin{aligned} \delta_{CDM}^{(i)} &= \frac{1}{(1+x)^{r_i}} \cdot {}_2F_1\left(r_i, r_i + \frac{1}{2}, 2r_i + \frac{1}{2}, \frac{1}{1+x}\right), \quad i = 1, 2, \\ r_{1,2} &= \frac{1}{4} \cdot \left(1 \mp \sqrt{1 + \frac{24\Omega_{CDM}}{\Omega_M}}\right). \end{aligned} \quad (6.23)$$

One can see, by making use of the properties of the hypergeometric functions (see [13]), that at $\Omega_B = 0$ these two solutions coincide with (6.19) and (6.20), respectively. For finite Ω_B , the function ${}_2F_1$ is equal to 1 in the late time limit $x \rightarrow \infty$, hence

$$\delta_{CDM}^{(1,2)} \propto a^{\pm\frac{1}{4} \cdot \left(\sqrt{1 + \frac{24\Omega_{CDM}}{\Omega_M}} \mp 1\right)}.$$

This means that the absence of growing baryon perturbation before recombination yields the suppression of the growth of dark matter perturbation. At small Ω_B/Ω_M the growth factor is

$$\delta_{CDM} \propto a^{1 - \frac{3\Omega_B}{5\Omega_M}}. \quad (6.24)$$

for actual values of Ω_B and Ω_{CDM} given in (2.10), the correction in the exponent is about 10%. We note that the asymptotics we have obtained is not reached in the real Universe, since recombination occurs rather soon after radiation-matter equality. Still, some suppression does exist. We also note that similar effect on matter perturbations occurs due to massive neutrinos, see Section 8.4.3.

To obtain the complete formula for perturbation at late times, we match the solution (6.23) with (6.11) in the limit $x \rightarrow 0$. The hypergeometric function of real argument has the following behavior as $z \rightarrow 0$ [13]

$${}_2F_1(c, b, c+b; z \rightarrow 0) = \frac{\Gamma(c+b)}{\Gamma(c)\Gamma(b)} \cdot [2\psi(1) - \psi(c) - \psi(b) - \log(1-z)],$$

where ψ is the derivative of the logarithm of Γ -function. Hence, in the limit $x \rightarrow 0$ we have

$$\delta_{CDM}^{(i)} \rightarrow \frac{\Gamma(2r_i + \frac{1}{2})}{\Gamma(r_i)\Gamma(r_i + \frac{1}{2})} \left[-\log x + 2\psi(1) - \psi(r_i) - \psi\left(r_i + \frac{1}{2}\right) \right].$$

Now, the ratio of the scale factors at early times is given by

$$x = \frac{a(\eta)}{a_{eq}} = 2(\sqrt{2} - 1) \frac{\eta}{\eta_{eq}}.$$

Matching the solutions at small x then gives

$$\delta_{CDM}(a, \mathbf{k}) \simeq -9 \cdot \Phi_{(i)}(\mathbf{k}) \cdot \left(D_1 \cdot \delta_{CDM}^{(1)} + D_2 \cdot \delta_{CDM}^{(2)}\right), \quad (6.25)$$

where

$$D_1 = -\frac{\Gamma(r_1)\Gamma(r_1 + \frac{1}{2})}{\Gamma(2r_1 + \frac{1}{2})} \cdot \frac{\log \frac{0.6a_{eq}}{a_\times} + 2\psi(1) - \psi(r_2) - \psi(r_2 + \frac{1}{2})}{\psi(r_1) - \psi(r_2) + \psi(r_1 + \frac{1}{2}) - \psi(r_2 + \frac{1}{2})},$$

$$D_2 = D_1(r_1 \leftrightarrow r_2, B_1 \leftrightarrow B_2).$$

Here a_\times is the scale factor at the horizon entry,

$$a_\times = \frac{a_{eq}^2 H_{eq}}{\sqrt{2}k} = \frac{2(\sqrt{2}-1)a_{eq}}{k\eta_{eq}}.$$

The solution that increases at late times is $\delta_{CDM}^{(1)}$. Making use of the above formulas, we obtain at late times, to the linear order in Ω_B/Ω_M ,

$$\begin{aligned} \delta_{CDM}(a) &\approx -\frac{27}{2} \frac{a}{a_{eq}} \Phi_{(i)} \\ &\times \left[\left(1 - \frac{0.6\Omega_B}{\Omega_M}\right) \log(k\eta_{eq}) + \left(1 - \frac{\Omega_B}{\Omega_M}\right) \left(\mathbf{C} - \frac{7}{2} + \log \frac{2}{\sqrt{6} - \sqrt{3}}\right) \right]. \end{aligned} \quad (6.26)$$

We see that the solution (6.19) is restored for $\Omega_B = 0$. In the real Universe, the effect of baryons yields the suppression of the momentum-dependent part of dark matter perturbations by about 12%. The momentum-independent part increases by about 25%. The latter property leads to the modification of the numerical coefficient in the argument of logarithm in (6.21): 0.15 changes to 0.2.

Problem 6.2. Find the analytical solution to Eq. (6.22) in elementary functions in the unrealistic case $\Omega_B = 2\Omega_{CDM}$. Find the explicit expression for the mode that grows at late times. Compare it with the solution (6.26) obtained to the linear order in Ω_B/Ω_{CDM} .

6.2.2 Perturbations in baryon-photon component

Let us now consider baryon-photon component at the period from radiation-matter equality to recombination, $\eta_{eq} \lesssim \eta \leq \eta_r$. This analysis is important, in particular, from the viewpoint of CMB and baryon acoustic oscillations. Because of that, we are interested in relative perturbations in photon and baryon components, $\delta_\gamma = \delta\rho_\gamma/\rho_\gamma$ and $\delta_B = \delta\rho_B/\rho_B$, separately. It is important for the analysis that at the epoch we consider, ρ_B and ρ_γ are of the same order of magnitude, see below.

Due to intense scattering of photons off electrons and Coulomb interaction between electrons and baryons, the baryon-photon plasma is a single medium, in the sense that *velocities* of baryon and photon components coincide. The equality

$$v_\gamma = v_B \equiv v_{B\gamma} \quad (6.27)$$

is the main relation of the tight coupling approximation. We note here that this relation is not exact: it is valid only before recombination and only for sufficiently long waves. We discuss its limit of applicability in Section 8.5, and here we simply use it.

In the tight coupling approximation, the relation (6.27) is valid for all modes of perturbations. For adiabatic mode, there are two other relations, $\delta_B = 3\delta T/T$, $\delta_\gamma = 4\delta T/T$, so that

$$\delta_B = \frac{3}{4}\delta_\gamma. \quad (6.28)$$

The baryon-photon component is subdominant at matter domination, so we describe it by making use of the covariant conservation equations, now written for baryon-photon component. Of course, the sum of photon and baryon energy-momentum tensors obeys the covariant conservation equations (5.4) and (5.5) where $\delta\rho_\lambda = \delta\rho_\gamma + \delta\rho_B$, $\delta p_\lambda = \delta p_\gamma = (1/3)\delta\rho_\gamma$, and $v_\lambda \equiv v_{B\gamma}$ is the common velocity potential. It is useful to obtain separate equations for δ_B and δ_γ , without using the relation (6.28) for the time being. These separate equations will be instrumental for studying baryon isocurvature mode in Section 6.5. Since baryons are non-relativistic, the *energy transfer* from photons to baryons is negligible. Hence, the covariant energy conservation equation (5.4) is satisfied by baryons and photons separately. This can be seen also by considering the conservation of the baryon number.

Problem 6.3. Consider the baryonic current j_B^μ . In the locally Lorentz frame where baryons are at rest, it is equal to $j_B^\mu = (n_B, 0, 0, 0)$. In an arbitrary frame

$$j_B^\mu = n_B u^\mu,$$

where u^μ is the 4-velocity. The current obeys the covariant conservation equation

$$\nabla_\mu j_B^\mu = 0. \quad (6.29)$$

- (1) Show that the covariant conservation law (6.29) implies the conservation of the baryon number (the integral of n_B over infinite space-like hypersurface is independent of the choice of this hypersurface).
- (2) Show, making use of the relations $\rho_B = m_p n_B$, $\delta\rho_B = m_p \delta n_B$ and $p_B = \delta p_B = 0$, that the linearized conservation equation (6.29) is equivalent to Eq. (5.4) written for the baryon component.

The covariant conservation equation (5.4) written for the whole baryon-photon component, together with (6.27) and obvious relations like $\rho_{B\gamma} = \rho_B + \rho_\gamma$, etc., imply then that Eq. (5.4) is valid for the photon component as well.

Thus, Eq. (5.4) written for baryons and photons separately, gives

$$\delta'_B - k^2 v_{B\gamma} = 3\Phi' \quad (6.30)$$

$$\delta'_\gamma - \frac{4}{3}k^2 v_{B\gamma} = 4\Phi'. \quad (6.31)$$

Equation (5.5) is valid for baryon-photon component as a whole. It is conveniently written in terms of the functions δ_γ and

$$R_B = \frac{3\rho_B}{4\rho_\gamma}.$$

Let us recall that

$$\rho'_B = -3\frac{a'}{a}\rho_B, \quad \rho'_\gamma = -4\frac{a'}{a}\rho_\gamma$$

and $p = p_\gamma = \rho_\gamma/3$. Then Eq. (5.5) becomes

$$v'_{B\gamma} + \frac{a'}{a} \frac{R_B}{1+R_B} v_{B\gamma} + \frac{3}{4} u_s^2 \delta_\gamma + \Phi = 0, \quad (6.32)$$

where

$$u_s^2 = \frac{\delta p}{\delta \rho} = \frac{1}{3} \frac{\delta \rho_\gamma}{\delta \rho_\gamma + \delta \rho_B} = \frac{1}{3(1+R_B)} \quad (6.33)$$

is the sound speed squared for the baryon-photon plasma. For the subhorizon modes at matter domination, one has $\Phi' \ll k\Phi$. Hence, Eq. (6.31) can be used for expressing the velocity potential in terms of δ'_γ . We insert this expression into Eq. (6.32) and obtain the equation for δ_γ ,

$$\delta''_\gamma + 2\frac{a'}{a} \gamma_{B\gamma} \delta'_\gamma + k^2 u_s^2 \delta_\gamma = -\frac{4}{3} k^2 \Phi, \quad (6.34)$$

where

$$\gamma_{B\gamma} = \frac{R_B}{2(1+R_B)}.$$

Equations (6.30), (6.31) and (6.34) are valid for both adiabatic and isocurvature modes of perturbations.

Let us continue with the adiabatic mode. We see from (6.30) and (6.31) that the relation (6.28) indeed holds for baryon-photon perturbations. Now, the source in the right hand side of Eq. (6.34) at matter domination is the potential (6.15). To the leading order in k , particular solution to Eq. (6.34) is

$$\delta_\gamma = -\frac{4}{3u_s^2} \Phi_{CDM} = -4(1+R_B) \Phi_{CDM}. \quad (6.35)$$

This shows that the perturbations in the baryon-photon component are indeed induced by the gravitational potential produced by dark matter. This effect is almost time-independent, since the gravitational potential (6.15) does not depend on time at all.

Problem 6.4. Estimate the accuracy at which the formula (6.35) is valid for particular solution to Eq. (6.34) at $\eta_{eq} < \eta < \eta_r$.

The solution to the homogeneous equation describes the acoustic oscillation of the baryon-photon component. This is the continuation of the oscillations (4.14). We find the latter solution, as usual, by getting rid of the term δ'_γ (cf. problem 6 in Section I.2.4) and neglecting then the time-derivatives of slowly varying parameters (this is essentially the WKB approximation),

$$\delta_\gamma(\eta) = A \frac{1}{(3u_s^2)^{1/4}} \exp \left(- \int_0^\eta d\tilde{\eta} \gamma_{B\gamma} \frac{a'}{a} \right) \cos \left(k \int_0^\eta d\tilde{\eta} u_s \right), \quad (6.36)$$

where A is a constant. This solution is valid at radiation domination as well, and at that epoch baryons are negligible, i.e., $\delta_\gamma = \delta_{B\gamma}$ and $u_s^2 = 1/3$. By matching this solution to (4.14) at $k^{-1} \ll \eta \ll \eta_{eq}$, we find

$$A = 6\Phi_{(i)}, \quad (6.37)$$

and the lower limit of integration in the argument of cosine is equal to zero. The total photon density contrast is given by the sum of (6.35) (6.36). Since $R_B \propto a$, the integral in the exponent in (6.36) is straightforwardly calculated,

$$\exp\left(-\int_0^\eta d\tilde{\eta} \gamma_{B\gamma} \frac{a'}{a}\right) = \frac{1}{\sqrt{1 + R_B(\eta)}}. \quad (6.38)$$

Making use of (6.15), (6.33), we obtain the final expression for photon perturbation in the tight coupling approximation,

$$\begin{aligned} \delta_\gamma = \Phi_{(i)} \cdot & \left[-324 \cdot (1 + R_B) I^2 \frac{\Omega_{CDM}}{\Omega_M} (1 + z_{eq}) \frac{\log(0.2k\eta_{eq})}{(k\eta_0)^2} \right. \\ & \left. + \frac{6}{(1 + R_B)^{1/4}} \cos\left(k \int_0^\eta d\tilde{\eta} u_s\right) \right], \end{aligned} \quad (6.39)$$

We emphasize again that this expression is valid for $k\eta_{eq} \gg 1$ only, and the argument of the logarithm exceeds 1.

The expression (6.39) contains the term which monotonically decreases with momentum, and the oscillating term with definite phase. This has direct relevance to the angular spectrum of CMB temperature anisotropy, as we show in Chapter 9. Another important property is that at matter domination, when gravitational potential is time-independent,¹ there is a simple relation between velocity and density contrast in the baryon-photon component,

$$kv_{B\gamma} = kv_\gamma = \frac{3}{4k} \delta'_\gamma. \quad (6.40)$$

Hence, like at radiation domination, the velocity potential oscillates with the same frequency as the density contrast, and phases of these oscillations differ by $\pi/2$.

It is useful for what follows to estimate R_B at the time of photon last scattering, $z_r = 1100$. We recall that about 25% of baryons and electrons are bound by that time in helium atoms, see Chapter I.8. Hence, the effective baryon number density in ionized baryon-photon medium is equal to $0.75n_B$, where n_B is the total number density of baryons. We also recall that the photon energy density at temperature T is (see Section I.5.1)

$$\rho_\gamma = \frac{\pi^2}{15} T^4.$$

¹In fact, since recombination occurs soon after the beginning of matter domination, there is weak dependence of the gravitational potential on time. This, however, has small effect on the velocity of the baryon-photon component. We neglect this subtlety in what follows.

The effective baryon mass density is then

$$\rho_B = 0.75 \cdot m_p \eta_B n_\gamma = 0.75 \cdot m_p \eta_B \frac{2\zeta(3)}{\pi^2} T^3,$$

where η_B is the baryon-to-photon ratio, m_p is the proton mass and $\zeta(3) = 1.20 \dots$. This gives

$$R_B = 0.75 \cdot \frac{3\rho_B}{4\rho_\gamma} = 0.75 \cdot \frac{45\zeta(3)}{2\pi^4} \eta_B \frac{m_p}{T}.$$

It is worth noting that the value of $R_B(\eta_r)$ is neither very small nor very large; with the proton mass $m_p = 0.938 \text{ GeV}$ and temperature at last scattering $T_r \simeq 0.26 \text{ eV}$ we obtain

$$R_B(\eta_r) \simeq 0.75 \cdot 10^9 \eta_B.$$

The measured value of η_B is $\eta_B \simeq 0.62 \cdot 10^{-9}$, so that

$$R_B(\eta_r) \simeq 0.47. \quad (6.41)$$

The fact that R_B turns out to be of order 1 is due to two properties which appear pure coincidences: comparable values of Ω_{CDM} and Ω_B and short time interval between radiation-matter equality and recombination. Due to these properties, the effect of baryons on density perturbations, and hence on CMB, is neither negligible, nor so strong that the oscillating part of these perturbations is completely damped. The suppression effect of baryons on the oscillating term in (6.39) is numerically about 0.9.

The dependence of the baryon-photon perturbations, and hence CMB properties, on η_B is used for determining the baryon-to-photon ratio from the CMB data. As we mentioned in the accompanying book [1], this determination is in good agreement with the BBN result.

6.3 Adiabatic Perturbations of Intermediate Momenta

The formulas of Sections 6.1 and 6.2 are valid for long and short waves, $k\eta_{eq} \ll 1$ and $kus\eta_{eq} \gg 1$, respectively. From the CMB viewpoint, the perturbations of intermediate wavelengths are also of great interest. Unfortunately, reliable analytical description of these perturbations is lacking. Nevertheless, one can get an idea of their properties by employing the approximation that radiation-matter equality and recombination are separated by long interval of conformal time, i.e., $\eta_r \gg \eta_{eq}$. Then one can study waves that enter the sound horizon between these epochs, $\eta_{eq} \ll \eta_x \ll \eta_r$.

In this case, the gravitational potential at $\eta \gg \eta_{eq}$ is given by (6.3), while the density contrast is described by Eq. (6.34). Since the velocity potential for the adiabatic perturbations is nearly zero at the horizon entry, the solution of that equation is again given by the sum of the particular solution (6.35) and the solution

(6.36) to the homogeneous equation. The constant A in (6.36) is determined by matching at the sound horizon. The matching condition in our case is (see (6.4)),

$$\delta_\gamma(\eta_{\times}, \mathbf{k}) = -\frac{12}{5}\Phi_{(i)}(\mathbf{k}).$$

It gives

$$\frac{A}{(1 + R_B(\eta_{\times}))^{1/4}} = \frac{6}{5}(1 + 3R_B(\eta_{\times}))\Phi_{(i)}.$$

Hence, in the approximation we use, the photon density contrast for perturbation that crosses the sound horizon at $\eta = \eta_{\times}$ is

$$\begin{aligned} \delta_\gamma(\eta, \mathbf{k}) &= \frac{18}{5}\Phi_{(i)}(\mathbf{k}) \left\{ -[1 + R_B(\eta)] + \left[\frac{1}{3} + R_B(\eta_{\times}) \right] \right. \\ &\quad \left. \times \left[\frac{1 + R_B(\eta_{\times})}{1 + R_B(\eta)} \right]^{1/4} \cos \left(k \int_0^\eta d\tilde{\eta} u_s \right) \right\}. \end{aligned} \quad (6.42)$$

We see that the properties of the intermediate modes are qualitatively similar to the properties of the modes that enter the horizon at radiation domination.

Clearly, the approximation we have used is unrealistic: the ratio $\eta_r/\eta_{eq} \simeq 2.3$ is not a large parameter. Nevertheless, our result, together with those of Section 6.2 shows that the following qualitative picture is valid for all modes that enter the sound horizon before recombination. The gravitational potential just before last scattering is

$$\Phi(\mathbf{k}, \eta) = F(k, \eta)\Phi_{(i)}(\mathbf{k}), \quad (6.43)$$

where $F(k, \eta)$ is a function that slowly varies in time. The photon energy density contrast has a part that also slowly varies in time, and an oscillating part. The slowly varying part obeys (6.35), while the phase of the oscillating part is well defined. In other words,

$$\delta_\gamma(\mathbf{k}, \eta) = -4[1 + R_B(\eta)]\Phi(\mathbf{k}, \eta) + 4\Phi_{(i)}(\mathbf{k}) \cdot A(k, \eta) \cos \left(k \int_0^\eta u_s d\tilde{\eta} \right), \quad (6.44)$$

where $A(k, \eta)$ is another function that slowly varies in time (the numerical factor 4 is introduced for convenience). Both $F(k, \eta)$ and $A(k, \eta)$ are positive. These properties yield the oscillatory behavior of CMB angular spectrum, see Chapters 9 and 10. At given η , the function $F(k, \eta)$ strongly depends on momentum, and decays as k^{-2} at large k , see (6.14). The function $A(k, \eta)$ also strongly depends on k in the intermediate range of momenta, and tends to a constant at large k , see (6.39).

Finally, since the gravitational potential varies slowly in time, the relation (6.40) is approximately valid for the velocity in the baryon-photon perturbations of both intermediate and small wavelengths.

Problem 6.5. Making use of the results of this and previous Sections, estimate the variation of the function $A(k, \eta_r)$ in the interval of conformal momenta from $k \simeq 1/\eta_r$ to $k \sim 1/(u_s \eta_{eq})$. The same for the function $F(k, \eta_r)$.

6.4 CDM Isocurvature Perturbations

In the case of isocurvature modes, the cosmic medium can almost never be treated as single fluid. Hence, the results of Chapter 4 can almost never be used. This is in contrast to the adiabatic mode whose properties are often described by formulas of Chapter 4.

Let us first study the behavior of superhorizon CDM isocurvature perturbations. In this regime, the time-independent quantities ζ_λ are equal to

$$\zeta_{CDM} = -\Phi + \frac{1}{3}\delta_{CDM} = \frac{1}{3}\delta_{CDM,(i)} \quad (6.45a)$$

$$\zeta_B = -\Phi + \frac{1}{3}\delta_B = 0, \quad \zeta_\gamma = -\Phi + \frac{1}{4}\delta_\gamma = 0,$$

$$\zeta_{rad} = -\Phi + \frac{1}{4}\delta_{rad} = 0, \quad (6.45b)$$

where we recalled that gravitational potential and perturbations in photons, baryons and neutrinos vanish in the conformal Newtonian gauge as $\eta \rightarrow 0$. We note that (6.45) implies

$$\delta_{CDM,(i)} = \mathcal{S}_{CDM} = \text{const.} \quad (6.46)$$

The gravitational potential for superhorizon modes is found from (5.1) in the limit $k \rightarrow 0$. Making use of (6.45) we write Eq. (5.1) in the following form,

$$\begin{aligned} 3\frac{a'}{a}\Phi' + 3\frac{a'^2}{a^2}\Phi + 4\pi Ga^2\rho_{rad} \cdot 4\Phi + 4\pi Ga^2\rho_M \cdot 3\Phi \\ = -4\pi Ga^2\rho_{CDM}\delta_{CDM,(i)}. \end{aligned} \quad (6.47)$$

Let us find the solution to this equation at radiation domination. To this end, we neglect the last term in the left hand side and notice that the right hand side behaves as $a^2\rho_{CDM} \propto a^{-1} \propto \eta^{-1}$. Now, both remaining coefficients in front of Φ in the left hand side behave as $a^2\rho_{rad} \propto a'^2/a^2 \propto \eta^{-2}$. This implies that Φ grows linearly with η . Hence, $\Phi' = \Phi/\eta = (a'/a)\Phi$, and Eq. (6.47) becomes

$$\left(6\frac{a'^2}{a^2} + 16\pi Ga^2\rho_{rad}\right)\Phi = -4\pi Ga^2\rho_{CDM}\delta_{CDM,(i)}.$$

Finally, we use the Friedmann equation at radiation domination and obtain

$$\Phi = -\frac{1}{8}\frac{\rho_{CDM}}{\rho_{rad}}\delta_{CDM,(i)}, \quad \text{radiation domination.} \quad (6.48)$$

Similar calculation gives for the superhorizon mode at matter domination

$$\Phi = -\frac{1}{5} \frac{\rho_{CDM}}{\rho_M} \delta_{CDM,(i)}, \quad \text{matter domination.} \quad (6.49)$$

Perturbations that enter the horizon at matter domination have time-independent gravitational potential, and after horizon entry, δ_{CDM} grows as a until dark energy effect shows up, see Section 4.3. This is one of a few situations where the results of Chapter 4 apply to the isocurvature mode. It follows from (6.45b) that

$$\delta_\gamma = 4\Phi. \quad (6.50)$$

The relations (6.49) and (6.50) give the gravitational potential and photon density contrast for superhorizon modes at recombination. These formulas are of direct relevance for possible contribution of isocurvature modes into CMB anisotropy, see Chapter 9.

Shorter waves are of interest too. Like in Section 6.2, we consider here perturbations that enter the *sound* horizon at radiation domination. Our main purpose is to find Φ and δ_γ at recombination.

We begin with radiation domination epoch. Equation (6.48) suggests that the gravitational potential is small at this epoch. We firmly establish this fact later on. Therefore, the effect of the gravitational potential on δ_{CDM} is negligible, so we set $\Phi = 0$ in Eq. (6.16). It then follows from Eq. (6.16) that δ_{CDM} stays constant and equal to $\delta_{CDM,(i)}$. The potential Φ , albeit small, is non-zero. It generates perturbation in the baryon-photon component. To study this effect, we make use of Eqs. (5.1) and (5.3). We neglect the baryon contribution in the right hand side of Eq. (5.1), since the baryon energy density is small, $\rho_B \ll \rho_{B\gamma}$, while it follows from (6.45) that $\delta_B \sim \delta_{B\gamma}$. Then the right hand side of Eq. (5.1) contains the combination $(\rho_{B\gamma}\delta_{B\gamma} + \rho_{CDM}\delta_{CDM,(i)})$. The right hand side of Eq. (5.3) contains $\delta p = (1/3)\rho_{B\gamma}\delta_{B\gamma}$. By combining these two equations, we obtain the equation for the gravitational potential,

$$\Phi'' + \frac{3(1+u_s^2)}{\eta} \Phi' + u_s^2 k^2 \Phi = -\frac{4\pi}{3} G a^2 \rho_{CDM} \delta_{CDM,(i)}, \quad (6.51)$$

where we used the fact that $a \propto \eta$ at radiation domination. We set $u_s = 1/\sqrt{3}$ for estimates. The combination in the right hand side of (6.51) can be written as

$$a^2 \rho_{CDM} = \frac{C}{\eta}, \quad (6.52)$$

where C is a constant. The particular solution to the inhomogeneous equation (6.51) is

$$\Phi = -\frac{4\pi}{3} G \frac{C}{ku_s} \left[\frac{2}{(ku_s\eta)^3} + \frac{1}{ku_s\eta} \right] \cdot \delta_{CDM,(i)}, \quad (6.53)$$

while the general solution to the homogeneous equation is a linear combination of $(ku_s\eta)^{-3/2} J_{3/2}(ku_s\eta)$ and $(ku_s\eta)^{-3/2} N_{3/2}(ku_s\eta)$, where $J_{3/2}$ and $N_{3/2}$ are the

Bessel functions, cf. (4.12). We are interested in the solution to Eq. (6.51) with $\Phi \rightarrow 0$ as $\eta \rightarrow 0$. It is a linear combination of the solution (6.53) and the second solution to the homogeneous equation. Explicitly,

$$\begin{aligned} \Phi = & -\frac{4\pi}{3}G\frac{C}{ku_s}\left[\frac{2}{(ku_s\eta)^3} + \frac{1}{ku_s\eta} - \frac{2\cos(ku_s\eta)}{(ku_s\eta)^3}\right. \\ & \left.- \frac{2\sin(ku_s\eta)}{(ku_s\eta)^2}\right] \cdot \delta_{CDM,(i)}, \end{aligned} \quad (6.54)$$

where we used the explicit form of $N_{3/2}$. Note that this solution indeed coincides with (6.48) in the limit $\eta \rightarrow 0$.

After entering the sound horizon, but still at radiation domination, the perturbations are described by the long time asymptotics of the solution (6.54),

$$\begin{aligned} \Phi = & -\frac{4\pi}{3}G\frac{C}{ku_s}\left[\frac{1}{ku_s\eta} - \frac{2\sin(ku_s\eta)}{(ku_s\eta)^2}\right] \cdot \delta_{CDM,(i)} \\ = & -\frac{4\pi Ga^2\rho_{CDM}}{k^2}\left(1 - \frac{2\sin(ku_s\eta)}{ku_s\eta}\right) \cdot \delta_{CDM,(i)} \\ \equiv & \Phi_{CDM}(\eta) + \Phi^{(osc)}(\eta), \end{aligned}$$

where we recalled (6.52). The gravitational potential has monotonic part $\Phi_{CDM} \propto \eta^{-1}$ and oscillating part $\Phi^{(osc)} \propto \sin(ku_s\eta)$. The amplitude of oscillations decreases as η^{-2} . The perturbation $\delta_{B\gamma}$, and hence $\delta_\gamma = \delta_{B\gamma}$, also has monotonic and oscillating parts. We find the oscillating part from Eq. (5.3):

$$\delta_{B\gamma}^{(osc)} = \delta_\gamma^{(osc)} = -\frac{k^2\Phi^{(osc)}}{4\pi Ga^2\rho_{B\gamma}} = -2\frac{\rho_{CDM}}{\rho_{B\gamma}ku_s\eta}\sin(ku_s\eta)\delta_{CDM,(i)};$$

we do not need the expression for monotonic part here. We see that the amplitude of oscillations of δ_γ is independent of time, i.e., the oscillations behave in accordance with Eq. (6.34) with zero right hand side (and $R_B = 0$, since we study radiation domination). We write this amplitude in a convenient form by recalling that at radiation domination $1/\eta = a_0 H_0 \sqrt{\Omega_{rad}}(1+z)$ and that $\Omega_M/\Omega_{rad} = 1+z_{eq}$. Making use of (2.16), we obtain

$$\frac{\rho_{CDM}}{k\eta\rho_{rad}} = \frac{\Omega_{CDM}}{\Omega_M}\frac{2I}{k\eta_0}\sqrt{1+z_{eq}},$$

which gives

$$\delta_\gamma^{(osc)} = -2\sqrt{3}\frac{\Omega_{CDM}}{\Omega_M}\frac{2I}{k\eta_0}\sqrt{1+z_{eq}} \cdot \sin(ku_s\eta) \cdot \delta_{CDM,(i)}. \quad (6.55)$$

The transition to matter domination is studied in much the same way as in Section 6.2. After radiation-matter equality, the perturbation δ_{CDM} grows as a . Dark matter perturbation at matter domination is

$$\delta_{CDM}(\eta) = \frac{3}{2}\frac{a(\eta)}{a_{eq}}\delta_{CDM,(i)}. \quad (6.56)$$

The numerical factor here is obtained as follows. Near η_{eq} and later, the solution δ_{CDM} is given by (6.19). The solution (6.20) which tends to $\log \eta$ as $\eta \rightarrow 0$, is absent for the isocurvature mode. The result (6.56) is obtained by matching (6.19) to $\delta_{CDM} = \delta_{CDM,(i)}$ at small η .

The monotonic part of the gravitational potential at matter domination is found from (6.12),

$$\begin{aligned}\Phi_{CDM} &= -\frac{3}{2} \frac{a^2}{k^2} \frac{a}{a_{eq}} 4\pi G \rho_{CDM} \delta_{CDM,(i)} \\ &= -9 \cdot I^2 \cdot \frac{\Omega_{CDM}}{\Omega_M} \frac{1}{(k\eta_0)^2} (1 + z_{eq}) \cdot \delta_{CDM,(i)}.\end{aligned}\quad (6.57)$$

As always, Φ_{CDM} is independent of time. The oscillating part of the gravitational potential is small for large $ku_s\eta$, so we neglect it from now on. To find δ_γ , we use Eq. (6.34). We match the oscillating solution to (6.55) and again use the particular solution (6.35) to the inhomogeneous equation. The result is

$$\begin{aligned}\delta_\gamma &= \delta_{CDM,(i)} \cdot \left[36 \cdot (1 + R_B) I^2 \frac{\Omega_{CDM}}{\Omega_M} \frac{1}{(k\eta_0)^2} (1 + z_{eq}) \right. \\ &\quad \left. - 4I\sqrt{3} \frac{\Omega_{CDM}}{\Omega_M} \frac{1}{k\eta_0} \frac{\sqrt{1 + z_{eq}}}{(1 + R_B)^{1/4}} \sin \left(k \int_0^\eta d\tilde{\eta} u_s \right) \right]\end{aligned}\quad (6.58)$$

This expression is quite different from the result (6.39) valid for adiabatic perturbations. First, the phase of oscillations is shifted by $\pi/2$. Second, the oscillation amplitude as function of momentum decays as k^{-1} . Finally, monotonic part does not contain the logarithmic factor.

Similarly for adiabatic mode, the intermediate wavelength perturbations do not admit analytical treatment. It follows from (6.34) that photon perturbations of these wavelengths contain oscillating and monotonic parts. Unlike in the case of the adiabatic perturbations, the phase of oscillations depends on momentum; instead of (6.44) we now have

$$\begin{aligned}\delta_\gamma(\mathbf{k}, \eta) &= -4[1 + R_B(\eta)]\Phi(\mathbf{k}, \eta) \\ &\quad - 4\delta_{CDM,(i)}(\mathbf{k}) \cdot A(k, \eta) \cdot \sin \left[k \int_0^\eta u_s d\tilde{\eta} + \varphi(k) \right].\end{aligned}\quad (6.59)$$

The non-trivial dependence of the phase on momentum is clear from the result of the following problem.

Problem 6.6. Consider the case $\eta_r \gg \eta_{eq}$. Find $\delta_\gamma(\mathbf{k}, \eta)$ for perturbations that enter the sound horizon between matter-radiation equality and recombination. Show that the phase $\varphi(k)$ in (6.59) equals $(-\pi/2)$ in this regime.

6.5 Baryon Isocurvature Perturbations

When studying baryon isocurvature perturbations, we can no longer treat baryon-electron-photon plasma as a single fluid. The perturbations in the energy densities of baryons and photons, $\delta\rho_B$ and $\delta\rho_\gamma$, are to be considered separately. The initial data is determined by²

$$\mathcal{S}_B = \frac{\delta(n_B/s)}{(n_B/s)} = \delta_B - \frac{3}{4}\delta_\gamma. \quad (6.60)$$

At early times, $\eta \rightarrow 0$, we find in the conformal Newtonian gauge

$$\delta_{B,(i)} = \mathcal{S}_B \neq 0,$$

while $\delta_{\gamma,(i)} = \delta_{CDM,(i)} = 0$.

The analysis of the baryon isocurvature mode at radiation domination, as well as at matter domination but before horizon entry, coincides word by word with the analysis of CDM isocurvature mode given in Section 6.4. It is sufficient to change the notation, $CDM \longrightarrow B$, in appropriate places in Section 6.4. In particular, the result (6.55) is modified by the replacement

$$\delta_{CDM,(i)} \longrightarrow \frac{\Omega_B}{\Omega_{CDM}}\delta_{B,(i)}. \quad (6.61)$$

The analysis of the transition to matter domination must, however, be modified for modes that enter the horizon at radiation domination. The point is that unlike CDM perturbations, baryon perturbations do not grow before recombination, since baryons are tightly coupled to photons, and the baryon-photon plasma has high pressure. Indeed, we find from Eqs. (6.30), (6.31) that $\delta'_B = 3\delta'_\gamma/4$, and instead of (6.56) we obtain that up to recombination, $\delta_B = \delta_{B,(i)}$ since δ_γ is small. Still, the gravitational potential induced by baryons generates CDM perturbations. At radiation-matter equality these are of order $\delta_{CDM} \sim (\Omega_B/\Omega_{CDM})\delta_{B,(i)}$, and then they grow linearly with the scale factor. As a result, we find at matter domination but before recombination

$$\delta_{CDM} = \frac{3}{2} \frac{a(\eta)}{a_{eq}} \frac{\Omega_B}{\Omega_{CDM}} \delta_{B,(i)}. \quad (6.62)$$

To obtain this formula, we make use of Eq. (6.16) and the fact that Φ is determined by baryon and dark matter perturbations. Recalling that the baryon density contrast is time-independent, we obtain, instead of (6.17), the inhomogeneous equation,

$$\delta''_{CDM} + \frac{a'}{a}\delta'_{CDM} - 4\pi G\rho_{CDM} a^2 \delta_{CDM} = 4\pi G\rho_B a^2 \delta_{B,(i)}.$$

²The cosmic medium is electrically neutral locally. This means that the baryon and electron number densities are equal, $n_B = n_e$, $\delta n_B = \delta n_e$. Therefore, the electron contribution to the energy density is proportional to that of baryons, $\delta\rho_e = (m_e/m_p)\delta\rho_B$; we need not consider this contribution separately, and simply include it into $\delta\rho_B$.

Since ρ_{CDM} and ρ_B depend on time in the same way, the obvious particular solution to this equation is

$$\delta_{CDM} = -\frac{\Omega_B}{\Omega_{CDM}} \delta_{B,(i)}. \quad (6.63)$$

We, however, need the solution that tends to zero as $\eta \rightarrow 0$. It is obtained by adding to (6.63) the solution (6.19) with $C_1 = (\Omega_B/\Omega_{CDM}) \cdot \delta_{B,(i)}$. This gives precisely the result (6.62).

Hence, dark matter perturbations behave after radiation matter equality in the same way as in the case of CDM isocurvature mode, see (6.56). We see that within our approximation, *all* results for the baryon isocurvature mode are the same as for the CDM isocurvature mode, modulo the replacement (6.61). In fact, this result is very general: at least in the ideal fluid approximation, the baryon and CDM isocurvature modes behave exactly in the same way. There is a simple reason behind this property. Since the complete system of equations for perturbations is linear, we can consider isocurvature mode in which both $\delta_{CDM,(i)}$ and $\delta_{B,(i)}$ are non-zero and obey

$$\delta_{CDM,(i)} = -\frac{\Omega_B}{\Omega_{CDM}} \delta_{B,(i)}. \quad (6.64)$$

In other words, the initial perturbation is such that $\delta\rho_M \equiv \delta\rho_{CDM} + \delta\rho_B = 0$. Initially, perturbations in other components, photons and neutrinos, are absent. In the superhorizon regime, velocity perturbations vanish as well. Since the total energy density is unperturbed, gravitational potentials are not generated, and in their absence no perturbations in relativistic components and no velocities are generated in the course of the evolution. As a result, perturbations with the above initial data do not evolve at all. This is consistent with the system of equations (5.1)–(5.5): it has a solution with time-independent $\delta_{CDM} = \delta_{CDM,(i)}$, $\delta_B = \delta_{B,(i)}$ and vanishing other perturbations Φ , v_λ , δ_γ etc., provided the initial values obey (6.64). This proves that the baryon and CDM isocurvature modes are described by one and the same solution to the equations for perturbations modulo the replacement (6.61).

Chapter 7

Structure Formation

In this Chapter, we discuss the theory of structure formation based on the *linearized* Einstein and covariant conservation equations. The growth of perturbations leads to the formation of *protostructures*, regions with relative matter perturbations $\delta_M \sim 1$. These protostructures are progenitors of all gravitationally bound systems in the Universe, including first stars and quasars, galaxies and clusters of galaxies.¹ The cosmological expansion becomes irrelevant for protostructures with $\delta_M \gtrsim 1$, and their further collapse proceeds in the non-linear regime. Some aspects of the collapse are considered in Appendix A. The detailed study of structure formation in the non-linear regime is beyond the scope of this book.

It is remarkable that the linearized theory of the cosmological perturbations is sufficient for obtaining semi-quantitative answers to a number of questions concerning structures in the Universe. In particular, one can estimate the mass spectrum of structures (the average number of objects of a given mass per unit volume). This mass spectrum is different at different redshifts, since the structure formation process began relatively late and continues at the present epoch.

Observational data on structure distribution, combined with other data, are important for determining the properties of primordial perturbations and the values of the cosmological parameters like Ω_M , Ω_Λ and Ω_{curv} , uncovering the dark matter properties and estimating the neutrino masses. Concrete examples are given in this Chapter and in Chapter 8.

From the viewpoint of structure formation, the major players are the dark matter and baryon perturbations. We have studied their evolution before recombination in Chapter 6. We begin this Chapter by obtaining the spectrum of matter perturbations after recombination.

¹Underdense regions develop into voids, some of which have sizes of tens Mpc.

7.1 Matter Perturbations after Recombination: Linear Regime

7.1.1 Power spectrum $P(k, z)$

We continue the analysis of Sections 6.1 and 6.2, and study the linear evolution of scalar perturbations after recombination. We consider the adiabatic perturbations and concentrate on perturbations in dark matter and baryons. In the linear theory, the only relevant interaction of dark matter and baryons is gravitational.

We begin with modes that enter the horizon after recombination. Before the horizon entry, relative perturbations of baryons and dark matter are equal to each other in the adiabatic mode; they remain equal for large wavelength modes after horizon entry as well. The gravitational potential after recombination is given by (4.24), and the gravitational potential at matter domination is related to its initial value by (6.3). Hence, we have

$$\Phi(z) = \frac{9}{10} g(z) \Phi_{(i)}.$$

The evolution of dark matter and baryon perturbations is given by (4.18), so the density contrast after the horizon entry is

$$\delta(z) = -\frac{k^2}{4\pi G a^2 \rho_M} \Phi(z) = -\frac{3}{5} \frac{k^2}{a_0^2 H_0^2 \Omega_M} \frac{g(z)}{1+z} \Phi_{(i)} \quad (7.1)$$

(we denoted the density perturbation in dark matter and baryons together by δ_M in the previous Chapters; in this Chapter we omit the subscript M). Here we used the relation $4\pi G \rho_M(z) = 4\pi G \rho_{M,0}(1+z)^3 = (3/2) H_0^2 \Omega_M (1+z)^3$. As we pointed out several times, the matter density perturbations grow as $a(z) \propto (1+z)^{-1}$ after the horizon entry, modulo the factor $g(z)$ accounting for the dark energy effect at $z \lesssim 1$, see Section 4.4.

Matter perturbations are characterized by their power spectrum $P(k, z)$. It is defined in the standard way,

$$\langle \delta(\mathbf{k}, z) \delta(\mathbf{k}', z) \rangle = \frac{P(k, z)}{(2\pi)^3} \delta(\mathbf{k} + \mathbf{k}') = \frac{k^3 \mathcal{P}(k, z)}{4\pi} \delta(\mathbf{k} + \mathbf{k}'). \quad (7.2)$$

In this Chapter we study the power spectrum at relatively low z for modes which entered the horizon by that time. For small k , there is a simple relation between this spectrum and the spectrum of primordial perturbations $P_\Phi(k)$. This relation follows directly from (7.1). We recall that $P_\Phi = 2\pi^2 \mathcal{P}_\Phi / k^3$, see (5.33), and write

$$P(k, z) = \frac{18\pi^2}{25} \frac{k}{a_0^4 H_0^4 \Omega_M^2} \frac{g^2(z)}{(1+z)^2} \mathcal{P}_\Phi, \quad k\eta_{eq} \ll 1. \quad (7.3)$$

If the primordial spectrum is flat, i.e., \mathcal{P}_Φ is independent of k , the power spectrum $P(k)$ increases linearly with k .

This behavior is inherent for momenta below η_{eq}^{-1} . For $k \sim \eta_{eq}^{-1}$, the power spectrum is still roughly estimated by the formula (7.3). Let us estimate the present power at these momenta. With the values of the cosmological parameters given in

Section 2.1.2 and the value of A_Φ quoted in (5.41) we obtain for flat primordial spectrum

$$P(k \sim \eta_{eq}^{-1}, z = 0) \sim 6 \cdot 10^4 \left(\frac{\text{Mpc}}{h} \right)^3, \quad (7.4)$$

where we temporarily set $a_0 = 1$.

Let us turn to modes that enter the horizon deep at radiation domination. In the first place, we have to determine the initial data for the epoch we consider here, i.e., the density contrasts and velocities of baryons and dark matter right after recombination, at $\eta = \eta_r$. We use the tight coupling approximation for baryons as well as the approximation of their instantaneous decoupling from photons at the last scattering epoch. We find from (6.44) that the density contrast and velocity of baryon perturbations right after last scattering are

$$\delta_B(\eta_r) = \frac{3}{4} \delta_\gamma(\eta_r) = -\frac{1}{u_s^2(\eta_r)} \Phi(\eta_r) + 3\Phi_{(i)} \cdot A(k, \eta_r) \cdot \cos(kr_s), \quad (7.5a)$$

$$kv_B(\eta_r) = \frac{3}{4k} \delta'_\gamma(\eta_r) = -3\Phi_{(i)} \cdot u_s(\eta_r) \cdot A(k, \eta_r) \cdot \sin(kr_s), \quad (7.5b)$$

where $\Phi(\eta_r) \equiv \Phi_{CDM}(\eta_r)$ is given by (6.14),

$$u_s(\eta_r) = \frac{1}{\sqrt{3[1 + R_B(\eta_r)]}}$$

is the sound velocity in the baryon-photon plasma just before recombination, and

$$r_s = \int_0^{\eta_r} u_s d\eta \quad (7.6)$$

is the coordinate size of the sound horizon at that epoch. Here we used the relation (6.28) between the baryon and photon perturbations and formula (6.40) for the velocity potential. In the approximation used in Section 6.2 the function $A(k, \eta_r)$ entering (7.5) is given by

$$A(k, \eta_r) = \frac{3}{2(1 + R_B)^{1/4}}, \quad (7.7)$$

see (6.39).

The expressions (7.5a) and (7.5b) give the initial data for the evolution of baryon perturbations after recombination. The initial value of the dark matter density contrast is

$$\delta_{CDM}(\eta_r) = -\frac{k^2}{4\pi G \rho_{CDM}(\eta_r) a^2(\eta_r)} \cdot \Phi(\eta_r), \quad (7.8)$$

see (6.12). We determine the dark matter velocity potential at recombination later on. In the approximation of Section 6.2, the explicit expression for $\delta_{CDM}(\eta_r)$ is

$$\delta_{CDM}(\eta_r) = -\frac{27}{2} \left(1 - \frac{0.6\Omega_B}{\Omega_M} \right) \left(\frac{a_r}{a_{eq}} \right)^{1 - \frac{0.6\Omega_B}{\Omega_M}} \Phi_{(i)} \log(0.2k\eta_{eq}), \quad (7.9)$$

where we included corrections of order Ω_B/Ω_M to the formula (6.13); these corrections have been obtained at the end of Section 6.2. The expressions for these corrections are approximate only, since they have been found in the limit $\eta_r/\eta_{eq} \rightarrow \infty$; we write them here to illustrate the slight suppression due to baryons.

Let us neglect energy densities of photons and neutrinos and consider the matter domination epoch; the effect due to dark matter is, as usual, accounted for by introducing the suppression factor $g(z)$. The evolution of the baryon and dark matter perturbations is determined by the covariant conservation equations (5.6), (5.7) for free baryons after recombination,

$$\delta'_B - k^2 v_B = 3\Phi', \quad (7.10a)$$

$$v'_B + \frac{2}{\eta} v_B = -\Phi, \quad (7.10b)$$

and similar equations for dark matter,

$$\delta'_{CDM} - k^2 v_{CDM} = 3\Phi', \quad (7.11a)$$

$$v'_{CDM} + \frac{2}{\eta} v_{CDM} = -\Phi, \quad (7.11b)$$

where we used the fact that $a \propto \eta^2$ at matter domination. We use the following form of Eq. (5.1) determining the gravitational potential,

$$k^2 \Phi = -\frac{6}{\eta^2} \left(\frac{\Omega_{CDM}}{\Omega_M} \delta_{CDM} + \frac{\Omega_B}{\Omega_M} \delta_B \right). \quad (7.12)$$

It is obtained by noticing that the term proportional to k^2 dominates in the left hand side of (5.1) for subhorizon modes and by invoking the Friedmann equation. It is straightforward now to solve the system of Eqs. (7.10)–(7.12). We first make use of the fact that the time derivatives of the gravitational potential are negligible in (7.10a) and (7.11a). Hence, the baryon and dark matter velocities are expressed through the density perturbations,

$$k^2 v_B = \delta'_B, \quad k^2 v_{CDM} = \delta'_{CDM}. \quad (7.13)$$

The second of these equations is valid at recombination as well, so the initial (for the evolution after recombination) value of the dark matter velocity is

$$k^2 v_{CDM}(\eta_r) = \delta'_{CDM}(\eta_r) = \left(\frac{a'}{a} \delta_{CDM} \right)(\eta_r) = \frac{2}{\eta_r} \delta_{CDM}(\eta_r). \quad (7.14)$$

Here we recalled that the dark matter density contrast grows as $a(\eta)$ after radiation-matter equality.²

²We use the approximation $\eta_r \gg \eta_{eq}$ and neglect the effect of baryons and photons on the dark matter velocity at $\eta_{eq} < \eta < \eta_r$. The corresponding corrections to the above formulas are small.

Now, combining Eqs. (7.10) and (7.11), we obtain the homogeneous equation for the function

$$\Delta = \delta_{CDM} - \delta_B. \quad (7.15)$$

This equation is

$$\Delta'' + \frac{2}{\eta}\Delta' = 0.$$

With the above initial data, its solution is given by

$$\Delta(\eta) = C_1 + \frac{C_2}{\eta}, \quad (7.16)$$

where

$$\begin{aligned} C_1 &= \delta_{CDM}(\eta_r) - \delta_B(\eta_r) + k^2\eta_r[v_{CDM}(\eta_r) - v_B(\eta_r)], \\ C_2 &= -k^2\eta_r^2[v_{CDM}(\eta_r) - v_B(\eta_r)]. \end{aligned}$$

We see from (7.16) that the function $(\delta_{CDM} - \delta_B)$ tends to a constant at late times, while the density contrasts δ_{CDM} and δ_B themselves grow as $a(\eta)$. So, the density contrasts of baryons and dark matter become equal to each other in several Hubble times after recombination, and then baryons and dark matter evolve together. Figuratively speaking, baryons fall into gravitational potential wells prepared by dark matter.

Let us now find the dark matter density contrast at late times; the baryon density contrast is the same in this asymptotics. We use Eq. (7.11b) and the second of Eq. (7.13) and write the following equation,

$$\delta_{CDM}'' + \frac{2}{\eta}\delta_{CDM}' = -k^2\Phi.$$

Making use of the definition (7.15) we write Eq. (7.12) as

$$k^2\Phi = -\frac{6}{\eta^2}\delta_{CDM} + \frac{6}{\eta^2}\frac{\Omega_B}{\Omega_M}\Delta.$$

These two equations yield the final equation for δ_{CDM} ,

$$\delta_{CDM}'' + \frac{2}{\eta}\delta_{CDM}' - \frac{6}{\eta^2}\delta_{CDM} = -\frac{6}{\eta^2}\Delta.$$

Inserting here the solution (7.16) for $\Delta(\eta)$, we obtain the general solution to the latter equation,

$$\delta_{CDM}(\eta) = \frac{\Omega_B}{\Omega_M}\Delta(\eta) + \alpha\eta^2 + \frac{\beta}{\eta^3}, \quad (7.17)$$

where the constants α and β are determined by the initial data at $\eta = \eta_r$. The constant α is most relevant for the late time asymptotics. The expression (7.17)

must be equal to the known value $\delta_{CDM}(\eta_r)$ at $\eta = \eta_r$, and its derivative must be equal to $k^2 v_{CDM}(\eta_r)$. With these initial data we get

$$\alpha = \frac{\Omega_{CDM}}{\Omega_M} \left(\frac{3}{5\eta_r^2} \delta_{CDM}(\eta_r) + \frac{k^2}{5\eta_r} v_{CDM}(\eta_r) \right) + \frac{\Omega_B}{\Omega_M} \left(\frac{3}{5\eta_r^2} \delta_B(\eta_r) + \frac{k^2}{5\eta_r} v_B(\eta_r) \right).$$

Finally, we recall (7.14) and obtain

$$\alpha = \frac{\Omega_{CDM}}{\Omega_M} \frac{1}{\eta_r^2} \delta_{CDM}(\eta_r) + \frac{\Omega_B}{\Omega_M} \frac{1}{\eta_r^2} \left(\frac{3}{5} \delta_B(\eta_r) + \frac{k\eta_r}{5} k v_B(\eta_r) \right), \quad (7.18)$$

where $\delta_{CDM}(\eta_r)$, $\delta_B(\eta_r)$ and $v_B(\eta_r)$ are given by (7.9), (7.5a) and (7.5b), respectively. Hence, the baryon and dark matter density contrasts are given at late time by

$$\delta_B = \delta_{CDM} \equiv \delta = \alpha \eta^2. \quad (7.19)$$

Needless to say, they grow as $a(\eta)$.

According to (7.5) and (7.9), the function $\alpha(k)$, and hence the late-time density contrasts, contain the monotonic part and the part that oscillates as function of momentum. The oscillating part is suppressed by Ω_B/Ω_{CDM} . We discuss this part in Section 7.1.2, and here we concentrate on the dominant monotonic part. We find from (7.8) that $\Phi(\eta_r) \sim (a_r^2 H_r^2/k^2) \delta_{CDM}(\eta_r) \ll \delta_{CDM}(\eta_r)$. We then use (7.5a) and see that the monotonic part is determined solely by dark matter perturbation at recombination:

$$\delta_{mon} = \frac{\Omega_{CDM}}{\Omega_M} \frac{1+z_r}{1+z} g(z) \cdot \delta_{CDM}(\eta_r), \quad (7.20)$$

where we restored the suppression factor $g(z)$ encoding the dark energy effect. For large momenta, Eq. (7.9) gives, to the leading and subleading orders in Ω_B/Ω_M ,

$$\delta_{mon} = -F \left(z, \frac{\Omega_B}{\Omega_M} \right) \Phi_{(i)} \log(0.2k\eta_{eq}) \quad (7.21)$$

where

$$F \left(z, \frac{\Omega_B}{\Omega_M} \right) = \frac{27}{2} \left[1 - \left(1.6 + 0.6 \log \frac{1+z_{eq}}{1+z_r} \right) \frac{\Omega_B}{\Omega_M} \right] \frac{1+z_{eq}}{1+z} g(z) \quad (7.22)$$

We see that baryons reduce the monotonic part considerably; the quantity in square brackets in (7.22) is about 0.7 instead of 1.

The short wavelength part of the power spectrum is obtained from (7.21),

$$P(k, z) = \frac{2\pi^2}{k^3} \mathcal{P}_\Phi F^2(z) \log^2(0.2k\eta_{eq}), \quad k\eta_{eq} \gg 1. \quad (7.23)$$

We see that the power spectrum is nearly flat at $k\eta_{eq} \gg 1$ for flat primordial spectrum: modulo logarithmic factor, it behaves as

$$P(k) \propto P_\Phi(k) \propto k^{-3}, \quad k\eta_{eq} \gg 1.$$

Since the power spectrum grows with k at $k\eta_{eq} \ll 1$, see (7.3), the function $P(k)$ has a maximum at $k/a_0 \sim 1/(\eta_{eq} a_0) \sim 0.01 \text{ Mpc}^{-1}$ (the numerical value of $a_0\eta_{eq}$

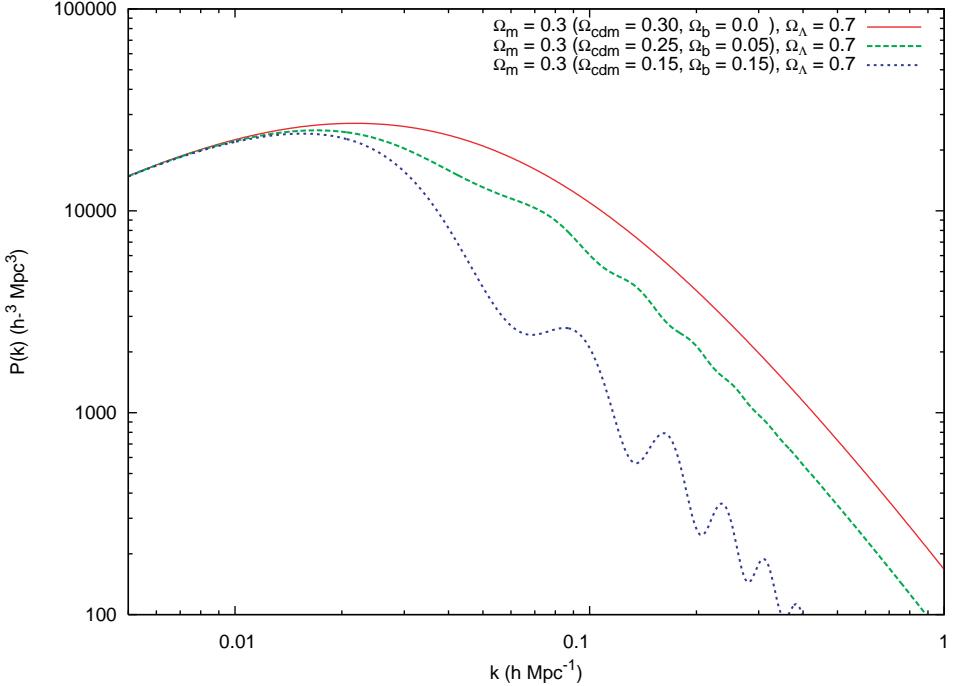


Fig. 7.1 The present power spectrum of matter perturbations $P(k)$ obtained in the linear theory for various cosmological parameters [14]. The scale factor normalization is $a_0 = 1$. The scale in the horizontal axis is $h \text{ Mpc}^{-1}$; with this choice, 0.1 corresponds to momentum of about 0.07 Mpc^{-1} . The momentum η_{eq}^{-1} is about $0.011 (h \text{ Mpc}^{-1})$, see Section 2.1.2. The spectrum grows as $P(k) \propto k$ at $k\eta_{eq} \ll 1$ and decreases as $P(k) \propto k^{-3}$ at $k\eta_{eq} \gg 1$. The power decreases in the high momentum region as Ω_B/Ω_M grows. The power at low momenta depends only on $\Omega_M = \Omega_{CDM} + \Omega_B$, rather than on Ω_{CDM} and Ω_B separately, in accordance with (7.3). The amplitude of oscillations becomes considerably smaller as the baryon fraction decreases.

is given in Section 2.1.2), and its present value at maximum is estimated in (7.4). This is confirmed by numerical calculations and agrees with observational data, see Figs. 7.1 and 7.2.

Let us make a comment concerning Figs. 7.1 and 7.2. At sufficiently high momenta, $k/a_0 \gtrsim 0.1 \text{ Mpc}^{-1}$, non-linear effects are sizeable; moreover, perturbations of momenta $k/a_0 \gtrsim 1 \text{ Mpc}^{-1}$ are entirely non-linear. On the other hand, the spectra shown in Figs. 7.1 and 7.2 are the present spectra which *would exist if non-linear effects were completely absent*. In other words, these are the spectra of *linear* theory. Hence, they do not reflect the properties of real matter inhomogeneities at relatively small spatial scales. In particular, they neglect collapse of matter perturbations which begins at $\delta \sim 1$ and leads to formation of structures, they do not account for the destruction of smaller objects when larger ones are formed, etc. Still, this representation is very useful. From the theoretical viewpoint, it enables one to discard complex processes occurring in the non-linear regime. Furthermore,

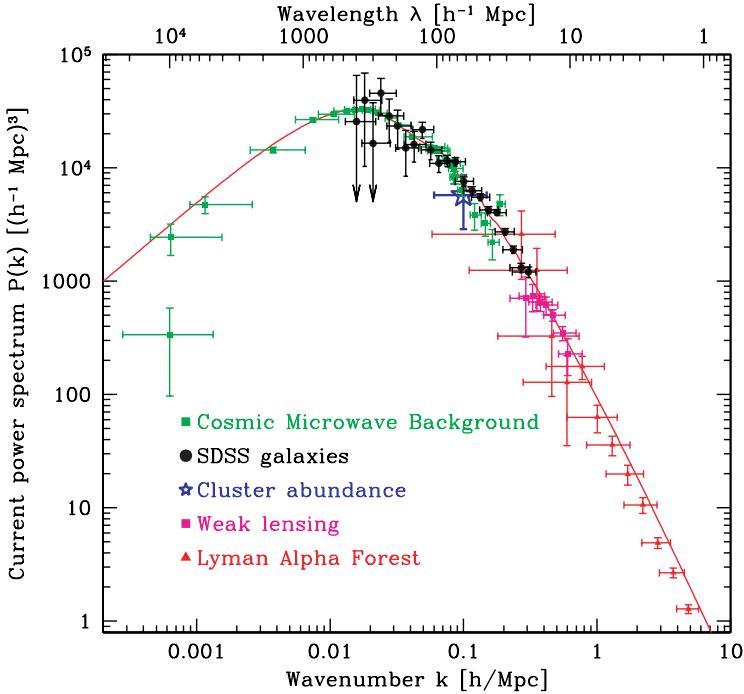


Fig. 7.2 Observational data on the present linear power spectrum of matter perturbations [15]. See Fig. 17.2 for color version.

the perturbations $\delta(\mathbf{k})$ are Gaussian in the linear theory only; the non-linearity leads to the non-Gaussianity, so one cannot use the notion of the power spectrum in the non-linear theory. From the viewpoint of the interpretation of observational data, it is important that the results obtained for different redshifts can be presented in a standardized way. Of course, one cannot get around the difficulties due to the non-linearity, since one has to relate real observational data to the power spectrum of the formal linear theory. These difficulties, however, are inevitable anyway once one compares observations with theory.

The power spectrum $P(k, z)$ determines the spatial correlation function of matter perturbations in the linear theory,

$$\xi(\mathbf{y}) = \langle \delta(\mathbf{x}) \delta(\mathbf{x} + \mathbf{y}) \rangle, \quad (7.24)$$

where we do not show the dependence on redshift. For the homogeneous random field $\delta(\mathbf{x})$, this correlation function depends on \mathbf{y} only. We write, by definition,

$$\xi(\mathbf{y}) = \int d^3k d^3k' \langle \delta(\mathbf{k}) \delta(\mathbf{k}') \rangle e^{i\mathbf{k}' \cdot \mathbf{x}} e^{i\mathbf{k} \cdot (\mathbf{x} + \mathbf{y})} = \int \frac{d^3k}{(2\pi)^3} P(k) e^{i\mathbf{k} \cdot \mathbf{y}}.$$

Since the random field of perturbations is isotropic, the power spectrum is independent of the direction of vector \mathbf{k} , and the correlation function is independent of

the direction of vector \mathbf{y} . We integrate over angles and obtain

$$\xi(y) = \int_0^\infty \frac{k^2 dk}{2\pi^2} \frac{\sin(ky)}{ky} P(k). \quad (7.25)$$

This integral is saturated³ at $ky \lesssim 1$, so the dependence of ξ on y is determined by the properties of the power spectrum at $k \lesssim y^{-1}$.

Let us find the correlation function at large distances, $y \gg \eta_{eq}$. In the realistic case of flat primordial spectrum, one has $P(k) \propto k$ at low momenta, see (7.3), while $P(k)$ decreases with k at high momenta, see (7.23). Let us integrate (7.25) by parts three times and obtain

$$\xi(y) = -\frac{1}{2\pi^2 y^4} \left\{ \frac{\partial^2}{\partial k^2} [k P(k)] \Big|_{k=0} + \int_0^\infty dk \cos(ky) \frac{\partial^3}{\partial k^3} [k P(k)] \right\}.$$

The integral here behaves as y^{-2} ; this can be seen by integrating by parts once again. Hence, to the leading order in y^{-1} we obtain

$$\begin{aligned} \xi(y) &= -\frac{1}{2\pi^2 y^4} \frac{\partial^2}{\partial k^2} [k P(k)] \Big|_{k=0} \\ &= -\frac{18}{25} \frac{1}{a_0^4 y^4} \frac{1}{H_0^4 \Omega_M^2} \frac{g^2(z)}{(1+z)^2} \mathcal{P}_\Phi, \quad y \gg \eta_{eq}, \end{aligned} \quad (7.26)$$

where \mathcal{P}_Φ is the primordial spectrum at low k . Note that $\xi(y)$ is negative at large y , i.e., the matter perturbations *anticorrelate* at large distances. The correlation function decreases as y grows, $|\xi| \propto y^{-4}$.

Let us now estimate the correlation function at short distances, $y \ll \eta_{eq}$. The relevant formula here is (7.23). The integral (7.25) is cut off at $k \sim y^{-1}$, and we obtain for nearly flat primordial spectrum

$$\xi(y) = \frac{1}{3} \mathcal{P}_\Phi F^2(z) \log^3 \left(\frac{0.2 \eta_{eq}}{y} \right), \quad y \ll \eta_{eq}, \quad (7.27)$$

where $\mathcal{P}_\Phi = \mathcal{P}_\Phi(k \sim y^{-1})$. This expression is valid in the leading logarithmic approximation, the corrections are suppressed by $\log^{-1}(0.2 \eta_{eq}/y)$. Intermediate distances $y \sim \eta_{eq}$ are hard to study; in fact, this region is quite large, $10 \text{ Mpc} \lesssim a_0 y \lesssim 100 \text{ Mpc}$.

The correlation function (7.27) decreases as y increases. It is positive at short distances. The fact that the correlation function (7.24) is sign-indefinite, is general; it does not rely upon the mechanism of generation or evolution of perturbations. It is sufficient that the field of perturbations $\delta(\mathbf{x})$ is random and homogeneous. Indeed, the integral of the matter density $\rho(\mathbf{x})$ over the whole visible Universe is equal, by definition, to the average density multiplied by the volume. Therefore

$$\int d^3y \delta(\mathbf{x} + \mathbf{y}) = 0.$$

³This is valid for monotonic part of the spectrum $P(k)$ only, see Section 7.1.2.

This gives

$$\int d^3y \xi(\mathbf{y}) = 0, \quad (7.28)$$

so the function $\xi(\mathbf{y})$ must indeed be sign-indefinite.

Problem 7.1. *Estimating the logarithmic factor in (7.27) as a quantity of order 1, show that*

- (1) *the expressions (7.26) and (7.27) for the correlation function give the same order-of-magnitude estimates at $y = \eta_{eq}$;*
- (2) *the positive contribution to the integral in (7.28) calculated by using (7.27) is of the same order as the negative contribution obtained with (7.26).*

Let us come back to the power spectrum $P(k, z)$. Its determination at various momenta are made using different methods. Let us give an idea of some of these methods. The information on the lowest momenta (the largest wavelengths) is obtained mostly from CMB observations, see Chapters 9, 10 for details. Usually, there is a correspondence between the angular scale θ and conformal momentum scale k , namely, $k\eta_0 \sim \pi/\theta$. Presently, CMB anisotropy and polarization are measured at angular scale of a few arcminutes and larger; this corresponds to momenta $k/a_0 \sim 10^{-4} - 10^{-1} \text{ Mpc}^{-1}$, see Fig. 7.2. It is important that CMB anisotropy and polarization are generated by perturbations *well into the linear regime* of their evolution. So, their measurements give, at least in principle, very precise determination of the power spectrum at these scales.

The power spectrum at intermediate momenta, $k/a_0 \sim 0.01 - 0.3 \text{ Mpc}^{-1}$ is obtained from the analysis of deep galaxy and quasar surveys, which measure both positions in the sky and redshifts of these objects. One of the observables is the two-point correlation function at a given redshift. Namely, let $n_g(\mathbf{x})$ be the number density of galaxies at a given redshift, averaged over a ball of radius R centered at the point \mathbf{x} . The radius R must be large enough, so that there are many galaxies in this ball. Also, perturbations at the length scale R should be better in the linear regime. The latter two requirements imply $R \gtrsim 10 \text{ Mpc}$. Let us define the galaxy correlation function,

$$\xi_g(\mathbf{y}) = \langle n_g(\mathbf{x}) n_g(\mathbf{x} + \mathbf{y}) \rangle, \quad (7.29)$$

where the average is still over an ensemble of Universes. The property that ξ_g is independent of \mathbf{x} is due to the homogeneity of the random field $n_g(\mathbf{x})$. This correlation function is well-defined for large y only,

$$a_0 y \gg R. \quad (7.30)$$

Let us assume that the number density of galaxies is proportional to the matter density (also averaged over a large ball),

$$n_g(\mathbf{x}) = \text{const} \cdot \rho_M(\mathbf{x}) = \text{const} \cdot \bar{\rho}_M \cdot [1 + \delta(\mathbf{x})]. \quad (7.31)$$

where the proportionality coefficient depends on redshift; we temporarily use the notation $\bar{\rho}_M$ for the average density. The relation (7.31) is not exact, but it works reasonably well for relatively small z , when substantial part of matter resides in galaxies (see Section 7.2.2); we do not discuss corrections to this relation. It follows from (7.31) that

$$\xi_g(\mathbf{y}) = \bar{n}_g^2 \cdot [1 + \xi(\mathbf{y})], \quad (7.32)$$

where $\xi(\mathbf{y})$ is the correlation function of matter perturbations (7.25), and \bar{n}_g is the average number density of galaxies. Measurements of the galaxy correlation function (7.29) give, with account of (7.25), the information on the power spectrum $P(k, z)$. The momentum range $k \sim y^{-1}$ probed in this way is limited from above because of (7.30); the correlation function is very small at large y (and hence small k), see (7.26), so the errors in the determination of $\xi_g(y)$ are large. Hence the finite range of momenta probed by this method. Let us note that since we are dealing with only one Universe, averaging in (7.29) over an ensemble of Universes is replaced by averaging over independent balls of size R . The result is plagued by cosmic variance discussed in Section 5.4.

Yet another way to obtain information on the matter power spectrum at intermediate momenta is the counting of clusters of galaxies at different redshifts $z \lesssim 1$. The clusters are gravitationally bound systems, the largest virialized objects in the Universe. We will see in Section 7.2 that their number at given z is related to $P(k, z)$ in a relatively simple way. There are several methods of cluster counting. One makes use of the fact that the central parts of clusters host a lot of hot gas that emits X-rays; X-ray observations are then used to infer the number of baryons and total mass of a cluster. Another method makes use of the Sunyaev–Zeldovich effect: hot free electrons in clusters affect CMB.⁴

One more method of studying structures uses the gravitational lensing of bright sources (bright galaxies, quasars) and also of CMB. Here we are talking about *weak lensing*, that is the distortion of images of background objects by clusters of galaxies. The total effect is due to all lenses (gravitational potentials) along the line of sight. Unlike in conventional, strong lensing (see Figs. I.1.7, I.1.8), no multiple images of a source are generated, so the method is based on statistical analysis: a set of distorted images in a given direction carries the information on the gravitational potentials. Importantly, this method is sensitive to the total gravitational potentials produced by baryons and dark matter together. The length scales probed are determined by the distances between clusters of galaxies and sizes of regions from which matter fell into clusters. These scales are about 10 Mpc and a few times larger.

Finally, for determining the power spectrum at the shortest scales of a few Mpc and smaller, one makes use of the information on clouds of neutral hydrogen at

⁴This method counts baryons and does not tell us much about dark matter in clusters. However, the ratio of mass densities of baryons and dark matter in large objects, $M = 10^{14} - 10^{15} M_\odot$, is the same as on average in the Universe.

large redshifts, up to $z \sim 4$. The relevant scales are characteristic of the regions from which matter gathered into galaxies.⁵ Unlike galaxies at present, these clouds *began* to enter the non-linear regime of evolution, so their approximate analysis may possibly be based on linear theory. These clouds are detected by measuring the spectra of distant quasars. Although these spectra have strong emission lines, they are continuous in rather large frequency range. Neutral hydrogen between a quasar and observer absorbs photons whose frequencies coincide with the frequencies of transitions between hydrogen levels, notably, the Lyman- α line (1216 \AA). Hence, the position of the absorption line gives the redshift of a cloud, while the line width measures typical velocities of hydrogen atoms, and hence the mass of a cloud. The light from a quasar may cross several clouds at different redshifts. This gives a comb of dips in the spectrum, called *Lyman- α forest*. Improvement of the theory of these clouds will hopefully make it possible to accurately measure $P(k, z)$ at the highest momenta.

The results of the determination of the matter power spectrum obtained by the above methods [15] are plotted in Fig. 7.2. All of them are translated into the present spectrum $P(k, z = 0)$. The figure shows both the agreement between the results obtained by different methods and the validity of estimates given in this Section. Measurements of the spectrum $P(k)$ play an important role in the determination of the cosmological parameters and properties of primordial perturbations. This is clear, e.g., from Fig. 7.1 where the dependence of the spectrum on Ω_B/Ω_{CDM} is shown. Another example is given in Fig. 8.5 which illustrates the sensitivity of the spectrum to the neutrino fraction in the total energy density; the measurements of the spectrum give the strongest bound on the sum of neutrino masses, see (8.104).

7.1.2 *Baryon acoustic oscillations*

As we have pointed out, the second term in (7.18), and hence $\delta(k)$, contains the part that oscillates as function of momentum. This part is due to the oscillating terms in (7.5). The physics behind this property is the acoustic oscillations in the baryon-photon plasma before recombination. Waves of different momenta and hence frequencies have different phases at recombination. These phases in the density and velocity of baryon perturbations get frozen at the recombination epoch, and they yield the oscillatory dependence of the power spectrum on momenta [16–19]. This effect is called *baryon acoustic oscillations, BAO*.

We note that the phenomenon of oscillations in the spectrum of non-relativistic baryons was uncovered by Sakharov [20], who considered the evolution of perturbations in the cold Universe initially filled with relativistic baryons. Such a matter

⁵In future, this method may possibly be used to study the scales relevant for dwarf galaxies of masses of $10^7 M_\odot$ and above.

also

undergoes acoustic oscillations in time, and when baryons become non-relativistic, these oscillations freeze in as the oscillations in momentum. The oscillations in the matter spectrum originating from baryons are often called *the Sakharov oscillations*.

Let us discuss BAO phenomenon in some detail. It is clear from (7.5) that their period equals $2\pi/r_s$. It is important therefore to obtain an accurate estimate of the coordinate size of the sound horizon at recombination. Its definition is given by (7.6). We proceed along the lines of Section 2.1.2 and write it as follows,

$$r_s(\eta_r) = \frac{I_1[\Omega_{rad}/\Omega_M, R_B(\eta_r)]}{\sqrt{3}} \eta_r, \quad (7.33)$$

where

$$I_1 = \frac{\int_{z_r}^{\infty} \frac{dz}{\sqrt{[(1+z)^3 + \frac{\Omega_{rad}}{\Omega_M}(1+z)^4][1 + \frac{1+z_r}{1+z}R_B(\eta_r)]}}}{\int_{z_r}^{\infty} \frac{dz}{\sqrt{(1+z)^3 + \frac{\Omega_{rad}}{\Omega_M}(1+z)^4}}}. \quad (7.34)$$

Making use of our estimate for $R_B(\eta_r) \simeq 0.47$, see (6.41), and setting $\Omega_M/\Omega_{rad} = 1 + z_{eq} = 3.2 \cdot 10^3$, we find numerically

$$I_1 \approx 0.91. \quad (7.35)$$

The estimate for $a_0\eta_r$ given in (2.24) yields

$$a_0 r_s = \frac{I_1}{\sqrt{3}} a_0 \eta_r \simeq 150 \text{ Mpc}. \quad (7.36)$$

This is the value that determines the period of BAO.

Making use of (7.5), (7.18) and (7.19), we write for the oscillating contribution

$$\delta_{osc}(z) = \frac{3\Omega_B}{5\Omega_M} \frac{1+z_r}{1+z} A(k, \eta_r) (3 \cos kr_s - k\eta_r u_s \sin kr_s) \cdot \Phi_{(i)}.$$

The first term in parenthesis is due to the oscillations in the energy density, while the second term comes from the oscillations in the velocity. The velocity contribution dominates at sufficiently large momenta,

$$\frac{k}{a_0} > \frac{3}{u_s} \frac{1}{a_0 \eta_r} \simeq 0.02 \text{ Mpc}^{-1}. \quad (7.37)$$

We see from Fig. 7.1 that momenta (7.37) are the most interesting from the viewpoint of the oscillations in the power spectrum, so we can use the simplified formula

$$\delta_{osc}(z) = -\frac{3\Omega_B}{5\Omega_M} \frac{1+z_r}{1+z} A(k, \eta_r) k\eta_r u_s \sin kr_s \cdot g(z) \cdot \Phi_{(i)}, \quad (7.38)$$

where we restored $g(z)$. It is clear that the oscillating term is small compared to the monotonic term (7.21) at reasonably small k (in fact, BAO are damped at large k , see below); nevertheless, its effect is observable.

To obtain the oscillatory part of the power spectrum we note that both monotonic and oscillating terms in $\delta(k)$ are proportional to the primordial gravitational potential $\Phi_{(i)}(k)$,

$$\delta_{mon}(k) = T_{mon}(k)\Phi_{(i)}(k), \quad \delta_{osc}(k) = T_{osc}(k) \sin kr_s \cdot \Phi_{(i)}(k),$$

where $T_{mon}(k)$ and $T_{osc}(k)$ are *transfer functions*.⁶ The monotonic and oscillating terms in $P(k)$ add up coherently, so to the leading order in Ω_B/Ω_M we have

$$\begin{aligned} P(k, \eta) &= |T_{mon}(k, \eta) + T_{osc}(k, \eta) \sin kr_s|^2 P_\Phi(k) \\ &= (T_{mon}^2 + 2T_{mon}T_{osc} \sin kr_s) P_\Phi(k). \end{aligned} \quad (7.39)$$

The second term here oscillates as function of momentum, while the first one is monotonic. Since T_{mon} and T_{osc} have the same sign (see (7.20) and (7.8); the function $A(k, \eta_r)$ is positive), maxima and minima of $P(k)$ are at

$$\frac{k_{max}}{a_0} = \frac{1}{r_s a_0} \left(\frac{\pi}{2} + 2\pi n \right) \approx 0.01, 0.05, 0.09 \text{ Mpc}^{-1}, \quad n = 0, 1, \dots, \quad (7.40)$$

and

$$\frac{k_{min}}{a_0} = \frac{1}{r_s a_0} \left(\frac{3\pi}{2} + 2\pi n \right), \quad n = 0, 1, \dots, \quad (7.41)$$

respectively (the values at $n = 0$ cannot be trusted, since the inequality (7.37) is not valid for these momenta).

These oscillations lead to a feature in the correlation function (7.24). According to (7.25), it is a sum of contributions of monotonic and oscillating terms in (7.39),

$$\xi(y) = \xi_{mon}(y) + \xi_{BAO}(y),$$

where

$$\begin{aligned} \xi_{BAO}(y) &= \int \frac{k^2 dk}{2\pi^2} \frac{\sin ky}{ky} 2T_{mon}T_{osc} \sin kr_s P_\Phi \\ &= \int \frac{k^2 dk}{2\pi^2} \frac{\cos[(y - r_s)k] - \cos[(y + r_s)k]}{ky} T_{mon}T_{osc} P_\Phi. \end{aligned} \quad (7.42)$$

While the contribution of the monotonic term to the integral (7.25) is saturated at $k \lesssim y$, the contribution to the integral (7.42) at $y \approx r_s$ comes also from higher momenta, since the first term in the integrand oscillates slowly. At $y \approx r_s$ we have $T_{mon} \propto \log k \eta_{eq}$, $T_{osc} \propto k$ and $P_\Phi \propto k^{-3}$, so the high momentum contribution to $\xi_{BAO}(y)$ is fairly large. As a result, $\xi_{BAO}(y)$, and hence the entire correlation function $\xi(y)$ has a peak at $y \approx r_s$, see Fig. 7.3. Applying the above formulas for estimating the height of this peak, i.e., $\xi_{BAO}(y = r_s)$, would yield a divergent integral $\int \log k dk/k$; in fact, this integral is cut off due to the Silk effect, see below.

⁶The factor that oscillates as function of momentum is usually also included into $T_{osc}(k)$. This is inconvenient for our purposes.

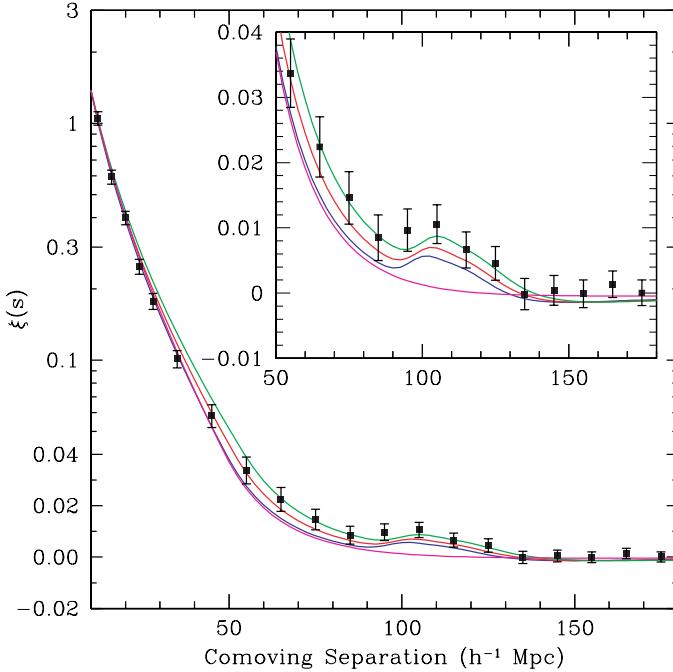


Fig. 7.3 The correlation function $\xi(s)$ determined by the analysis of the SDSS data on the distribution of distant galaxies [21]. Solid lines show the predictions of various cosmological models. See details in Fig. 17.3.

The peak in the correlation function $\xi(y)$ at $y \approx r_s$ has simple physical interpretation [22]. Let some small region of space near the point \mathbf{x} be initially overdense. For the adiabatic perturbations we consider, there is overdensity in both baryon-photon component and dark matter. The evolution of these perturbations after the sound horizon entry is different. The dark matter overdensity remains at the same place at all times, while the overdensity in the baryon-photon plasma generates a density wave expanding outwards from the point \mathbf{x} at the sound speed. This wave travels the coordinate distance r_s by recombination and then the baryon density contrast freezes in. As a result, there is overdensity both at the point \mathbf{x} (dark matter) and at distance r_s from this point (baryons). This yields the additional correlation between $\delta(\mathbf{x})$ and $\delta(\mathbf{x} + \mathbf{y})$ at $y \approx r_s$.

According to (7.32), the peak in the matter correlation function translates into the peak in the galaxy correlation function. It is this property that was detected by the analysis of data from SDSS Galaxy survey, see Fig. 7.3. Later on, the analysis of data of 2dFGRS and SDSS surveys at redshifts $z \simeq 0.2$ and $z \simeq 0.35$ has revealed the oscillating part in the power spectrum $P(k)$, see Fig. 7.4. The positions of maxima and minima in Fig. 7.4 are in good agreement with (7.40) and (7.41).

At recombination, the amplitude of the oscillating term in $P(k)$ grows approximately as k , see (7.38) and the estimate (7.7), while the monotonic part depends on

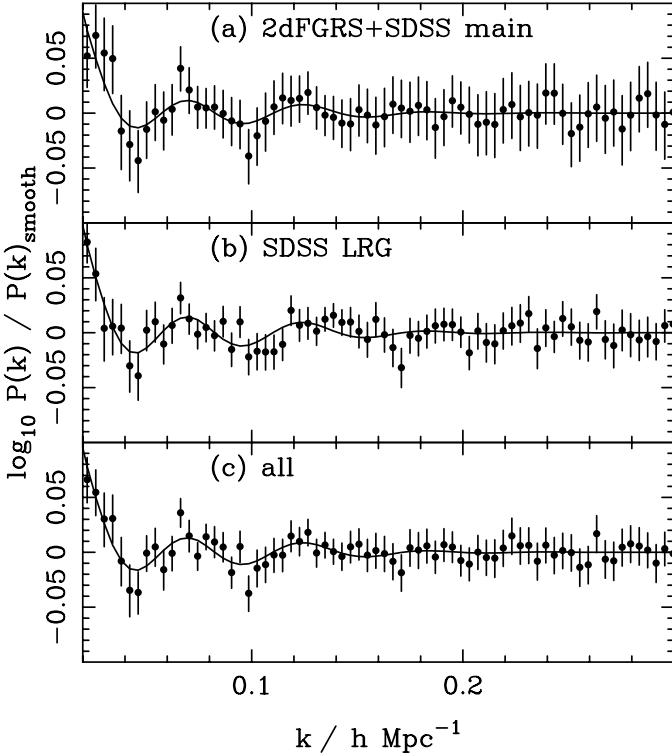


Fig. 7.4 The oscillating part in the power spectrum of matter perturbations obtained by the analysis of SDSS and 2dFGRS catalogs of galaxies [23]. Here $P_{\text{smooth}} \equiv P_{\text{mon}}$. Since the oscillating part is small, the quantity shown is approximately a half of the ratio of the oscillating part to the monotonic one. The scale in the horizontal axis is the same as in Fig. 7.1, so the maximum at $k/a_0 \simeq 0.05 \text{ Mpc}^{-1}$ (see (7.40)) is approximately at 0.07 in this figure.

momentum weakly, see the estimate (7.9). Therefore, the oscillating part grows with momentum relative to the monotonic one. However, this is true for low momenta only. The reason is that at high momenta,

$$k \gtrsim k_S \simeq a_0 \cdot 0.1 \text{ Mpc}^{-1},$$

the acoustic oscillations in the baryon-photon component get washed out due to the Silk damping. This is one of the effects occurring because the baryon-photon component is *not perfect fluid*; the photon mean free path is finite and in fact fairly long. We discuss the Silk damping in Section 8.5.2, and here we simply state that it can be approximately accounted for by multiplying the term δ_{osc} by the suppression factor $\exp(-k^2/k_S^2)$. The damping of the oscillations in the power spectrum at $k > k_S$ is clearly seen in Fig. 7.4.

Let us estimate the BAO amplitude. We use the result (7.38) and the estimates (7.7), (7.21), (7.22). Let us set the logarithmic factor equal to 1 for the

estimate.⁷ It then follows from (7.39) that the ratio of the amplitude (value at maximum) of the oscillating part to the monotonic part of the power spectrum is estimated as

$$\begin{aligned} \frac{P_{osc}(k_{max})}{P_{mon}(k_{max})} &= 2 \frac{T_{osc}(k_{max})}{T_{mon}(k_{max})} \\ &\sim \frac{2}{15} \frac{\Omega_B}{\Omega_{CDM}} \frac{1+z_r}{1+z_{eq}} \frac{u_s k_{max} \eta_r}{(1+R_B)^{1/4}} \cdot \exp\left(-\frac{k_{max}^2}{k_S^2}\right). \end{aligned} \quad (7.43)$$

In this way we obtain the numerical estimate for the maximum with $n = 1$,

$$\frac{P_{osc}(k_{max})}{P_{mon}(k_{max})} \sim 0.04. \quad (7.44)$$

This is in qualitative agreement with the result shown in Fig. 7.4.

Measuring BAO at different redshifts enables one to obtain information on the cosmological parameters. Indeed, the positions of maxima and minima in the power spectrum $P(k)$ are determined by the size of the acoustic horizon r_s at recombination. The dependence of this size on the cosmological parameters is rather weak, so it serves as (almost) standard ruler. The observable effects at various redshifts depend on the angular diameter distance $D_a(z)$ (its definition is given in Section I.4.7) and the Hubble parameter $H(z)$. In the procedure of averaging over angles in the momentum space, the length scale is given by distance scales in three directions. Two of these are related to angular measurements, and the corresponding distance scale is the angular diameter distance $D_a(\eta) = (\eta_0 - \eta) \cdot a(\eta)$. The third one is related to the measurement of the radial distance, so the scale is $H^{-1}(z)$. Hence, the length scale relevant for BAO at given redshift is $[D_a^2(z)/H(z)]^{1/3}$.

Problem 7.2. Show the validity of the last statement above.

Papers on BAO use the related distance scale,

$$D_V(z) = \left[\frac{z(1+z)^2 D_a^2(z)}{H(z)} \right]^{1/3}.$$

The results of measurements are expressed in terms of the ratio of $a_0 r_s$ to the latter scale. Hence, one measures

$$\frac{a_0 r_s}{D_V(z)}. \quad (7.45)$$

This quantity depends on the cosmological parameters in a fairly complex way. An example is shown in Fig. 7.5. It is clear that the interpretation of BAO measurements alone, like most others, suffers from the degeneracy in parameters.

⁷In fact, the quantity $0.2k_{max}\eta_{eq}$ in the argument of logarithm is close to 1 for interesting momenta. This means that our approximation is poor, and (7.44) must be viewed as the order-of-magnitude estimate. We emphasize, however, that the calculated *positions* of maxima and minima do not suffer from this problem, since the phase of oscillations does not rely upon the approximation used in obtaining (7.9), see also Section 6.3.

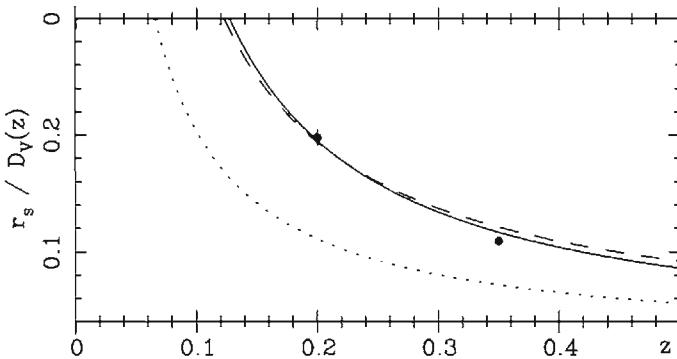


Fig. 7.5 Degeneracy in the cosmological parameters for BAO [23]. The points are observational results. The curves are predictions of cosmological models: spatially flat with $\Omega_M = 0.25$, $\Omega_\Lambda = 0.75$ (solid line), $\Omega_M = 1.0$, $\Omega_\Lambda = 0$ (short dashed line) and with negative spatial curvature $\Omega_M = 0.3$, $\Omega_\Lambda = 0$, $\Omega_{curv} = 0.7$ (long dashed line).

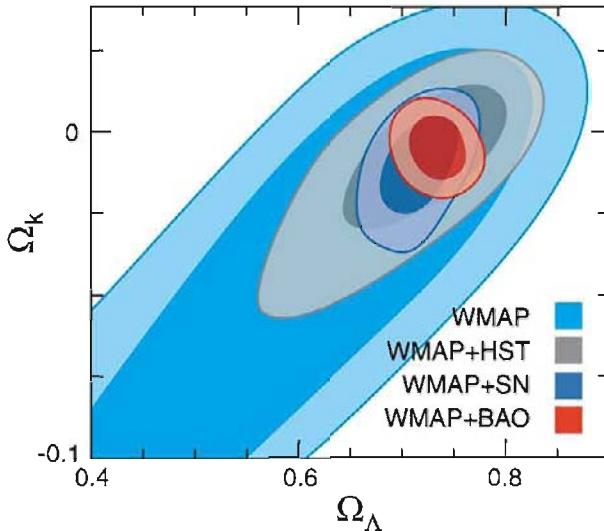


Fig. 7.6 The degeneracy in parameters (spatial curvature and dark energy density in this case) is removed by combining the data from different observations [2]. The notation WMAP refers to CMB observations, HST to the measurements of the present Hubble parameter and SN to the observations of SNe Ia. The central (dark) and light colored regions are allowed at 68% and 95% confidence level, respectively. See Fig. 17.4 for color version.

Let us note that according to (7.43), the amplitude of oscillations is sensitive to the parameter $\Omega_B/[\Omega_{CDM}(1 + z_{eq})] \propto \Omega_B/(\Omega_{CDM}\Omega_M)$ (see Fig. 7.1), and at relatively high momenta, also to the Silk momentum k_S . Combining the BAO data with the results of other cosmological observations enables one to remove the degeneracy in parameters. An example [2] is shown in Fig. 7.6.

7.2 Beginning of Non-linear Regime

7.2.1 Preliminaries

It follows from the above results that in the cold dark matter model, the dark matter density contrast after horizon entry at radiation domination grows, albeit weakly (logarithmically) with momentum. Hence, modes of the smallest spatial scale are the first to enter the non-linear regime, as they enter the horizon earlier; the first objects formed in the Universe are the smallest ones. From the viewpoint of formation of the first stars, the lowest mass scale of interest is $10^5 M_\odot$. The reason is as follows. The formation of the first stars begins with collapse of dark matter and baryons leading to the formation of protohalo. Hydrogen and helium get heated in the center of this protohalo and then cool down. The scale $M \sim 10^5 M_\odot$ is determined by the fact that smaller mass is insufficient to resist the radiation pressure; baryons are blown away from the halo and a star is not formed.⁸ This does not happen for protohalos of mass $M \gtrsim 10^5 M_\odot$, where the first stars of masses of order $100 M_\odot$ get formed.

Let us roughly estimate redshift of the first star formation. We use (7.21), (7.22) (the qualification “mon” is omitted in what follows). The perturbations enter the non-linear regime when

$$\delta(k, z) \sim 1. \quad (7.46)$$

Then the overdense region collapses pretty fast, so this is roughly the condition of the protohalo formation. We are interested in momenta $k/a_0 \sim \pi/R \sim (3 \text{ kpc})^{-1}$, where we use the estimate (1.29) for the present size of perturbation that corresponds to the mass $10^5 M_\odot$. We use (5.41) for the estimate of the primordial gravitational potential, and also the numerical values of z_{eq} and $a_0\eta_{eq}$ given in Section 2.1.2. In this way we obtain from (7.21), (7.22) that the relation (7.46) is valid at

$$z_5 \simeq 8, \quad (7.47)$$

where subscript 5 refers to the mass scale $M \sim 10^5 M_\odot$. In fact, the formation of the first stars begins somewhat earlier. One of the reasons is that $\delta(\mathbf{x})$ is random field whose values in some places are considerably larger than its typical fluctuation. We consider this and related issues in Section 7.2.2. In short, the intense formation of the first stars occurs at $z \sim 10$ and even somewhat earlier.

Notably, this picture has observable consequences. The formation of the first stars gives rise to the strong photon emission. This leads to secondary ionization (reionization) of hydrogen throughout the Universe. The latter affects CMB temperature anisotropy and especially CMB polarization, see Chapters 9, 10. The CMB

⁸Star formation at the subsequent epoch is affected by the presence of elements heavier than helium (“metals”) which are produced as the first stars burn and explode. This is the reason for the existence of lighter stars in the present Universe.

data yield the estimate of the reionization redshift, $z_{rei} \simeq 10$. This is in good agreement with the above estimate.

The structures of larger masses get formed later. As an example, the present spatial size of perturbations corresponding to galaxies ($M \sim 10^{11} M_\odot$) is of order $R \sim 1 \text{ Mpc}$, see Section 1.3. For this size, we obtain $z_{11} \simeq 2$ instead of (7.47). So, galaxies began to form relatively recently, at $z = 2 - 4$. Formation of even larger structures, clusters of galaxies, have begun even later, see Section 7.2.2.

Thus, structure formation is hierarchical in the cold dark matter model: larger structures form as a result of clumping of smaller ones.

7.2.2 Mass distribution of structures

Structure formation is a non-linear process. Its most adequate theoretical analysis makes use of numerical techniques. Nevertheless, some of its aspects, such as the mass distribution at various redshifts, can be studied at reasonable accuracy within the analytic approach known as the Press–Schechter formalism [24]. This is the approach we employ in this Section.

In the first place, let us state the result obtained in Section A.2 for homogeneous overdense sphere filled with non-relativistic matter: the collapse of this sphere ends at the time when the density contrast *calculated in the linear theory* is

$$\delta_c = 1.686. \quad (7.48)$$

Let us suppose that the same property holds in reality: if in the linearized theory the *average* density contrast in a ball of the present size R exceeds δ_c , this ball has just collapsed. The numerical analysis shows that this is a reasonably good approximation for smooth initial matter distributions.

It is then natural to introduce smoothed density contrast

$$\delta_R(\mathbf{x}, t) = \int d^3y \delta(\mathbf{x} + \mathbf{y}, t) W_R(\mathbf{y}),$$

where $W_R(\mathbf{y})$ is a window function which is substantially different from zero at $|\mathbf{y}| \lesssim R/a_0$ and normalized as

$$\int d^3y W_R(\mathbf{y}) = 1.$$

Note that we use comoving coordinates \mathbf{x} , \mathbf{y} here, while R is the *present* size of a cloud. The concrete form of the window function is not very important if the value of δ_c is chosen appropriately (see problem A.7 in this regard). The convenient choice is

$$W_R(\mathbf{y}) = \frac{3}{4\pi} \frac{a_0^3}{R^3} \theta(R - a_0|\mathbf{y}|).$$

This choice is dictated by the analogy to the homogeneous ball and corresponds to the value of δ_c given in (7.48). With this choice, the relation between the size R and

the mass of the compact object is standard,

$$M(R) = \frac{4\pi}{3} R^3 \rho_{M,0}, \quad (7.49)$$

where $\rho_{M,0}$ is the present mass density of baryons and dark matter.

Hence, a region of size R centered at the point \mathbf{x} has collapsed by the time t provided the following condition holds,

$$\delta_R(\mathbf{x}, t) \geq \delta_c. \quad (7.50)$$

The collapse leads to the formation of a gravitationally bound system. The further development of this structure proceeds locally, and we do not consider it. We repeat that only the first stage of the collapse occurs very similarly to the collapse of dust ball studied in Section A.2. Later on, at final stage, the absence of spherical symmetry, the interactions of baryons and the radiation processes become important.

Let us make a few comments concerning the formula (7.50). First, as we pointed out above (see also Section A.2), the quantity in the left hand side of (7.50) does not have direct physical significance. The function $\delta_R(\mathbf{x}, t)$ *would be* matter density contrast *if the linear theory were correct all the way until time t .* In reality, the density contrast in the center of a collapsing region greatly exceeds δ_c at the time when the linear theory gives $\delta_R(t) = \delta_c$; for homogeneous spherically symmetric matter distribution, the density in the center becomes infinite at that time, see Section A.2. The formalism we describe is convenient because it employs the results of formal *linear* theory to infer the properties of non-linear systems. In this Section we always understand the quantities $\delta(\mathbf{x}, t)$, $\delta_R(\mathbf{x}, t)$, etc., in this formal sense: they are calculated in the linear theory, and the results of the complete, non-linear theory are formulated in terms of approximate relations like (7.50). This approach naturally makes use of the power spectrum of the linear theory, which we have discussed in Section 7.1.1.

Second, once the condition (7.50) is satisfied, the size of the collapsing region can *exceed* R . Indeed, the condition (7.50) ensures that the region of size R has collapsed, but it does not specify what has happened to surrounding matter. In other words, the condition (7.50) states that the mass M of a compact object formed by the time t obeys

$$M \geq M(R), \quad (7.51)$$

where $M(R)$ is given by (7.49).

Third, the value (7.48) is obtained in Section A.2 in the cosmological model without dark energy. The presence of dark energy makes δ_c time-dependent. This dependence can be found by making use of Eq. (A.38). It is fairly weak [25], see Fig. 7.7, so we neglect this dependence in what follows.

Finally, structure formation at different spatial scales proceeds *independently* within this approach. Indeed, the condition (7.50) applied to a region of large size

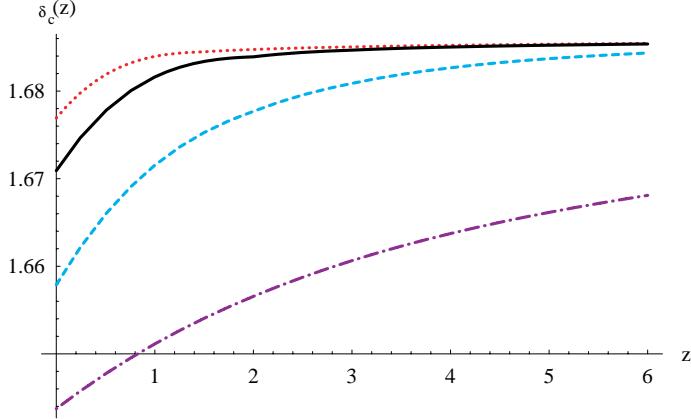


Fig. 7.7 Critical density contrast $\delta_c(z)$ as function of redshift for various models of dark energy [25]: cosmological constant (solid line), equation of state $p = w\rho$, $w(z) = w_0 + w_1z/(1+z)^2$, $w_0 = -1.3$, $w_1 = 4$ (dashed line), same with $w_0 = -1.3$, $w_1 = -2$ (dotted line), and Hubble law $H^2(z) = H_0^2 [\Omega_M(1+z)^3 + a_1(1+z) + a_2(1+z)^2 + (\Lambda - a_1 - a_2)]$, $a_1 = -4.16$, $a_2 = 1.67$ (dash-dotted line). All these models are consistent with SNe Ia data.

does not care about structures of smaller size inside this region. In particular, it is unimportant whether or not there are collapsed clumps of smaller sizes.

Clearly, the field $\delta_R(\mathbf{x}, t)$ defined above has the same statistical properties⁹ as the original perturbation field $\delta(\mathbf{x}, t)$. The field $\delta(\mathbf{x}, t)$ is assumed to be Gaussian. Therefore, $\delta_R(\mathbf{x}, t)$ is the Gaussian random variable at given \mathbf{x} . The Fourier-transforms of $\delta(\mathbf{x}, t)$ and $\delta_R(\mathbf{x}, t)$ are related as follows,

$$\begin{aligned} \delta_R(\mathbf{k}, t) &= \delta(\mathbf{k}, t) \cdot \int d^3y W_R(\mathbf{y}) e^{i\mathbf{ky}} \\ &= \delta(\mathbf{k}, t) W_R(\mathbf{k}) = \delta(\mathbf{k}, t) \frac{3j_1(kR/a_0)}{kR/a_0}. \end{aligned} \quad (7.52)$$

Here j_1 is the spherical Bessel function, see Appendix F. It follows from (7.52) that the main contribution to $\delta_R(\mathbf{x}, t)$ comes from perturbations of momenta $k \lesssim a_0/R$.

The variance σ_R of the random variable $\delta_R(\mathbf{x}, t)$ is given by

$$\begin{aligned} \sigma_R^2(t) \equiv \langle \delta_R^2(\mathbf{x}, t) \rangle &= \int d^3k d^3k' e^{i(\mathbf{k}+\mathbf{k}')\mathbf{x}} |W_R(\mathbf{k})|^2 \langle \delta(\mathbf{k}, t) \delta(\mathbf{k}', t) \rangle \\ &= \int_0^\infty \frac{dk}{k} \mathcal{P}(k, t) \frac{9j_1^2(kR/a_0)}{(kR/a_0)^2}. \end{aligned} \quad (7.53)$$

We note that both the smoothed density contrast δ_R and its variance σ_R can be considered as a function of mass M of the object to be formed; the relation between the mass and the size is given by (7.49). We use both σ_R and σ_M in what follows.

⁹It is important here that $\delta(\mathbf{x}, t)$ is a formal object calculated in the linear theory.

There is the obvious relation between them, $\sigma_R = \sigma_M$ for $M = M(R)$. The variable σ_M is particularly convenient, because the mass in the volume of comoving size k^{-1} is independent of redshift, so all dependence of σ_M on z is due to the redshift-dependence of the power spectrum $P(k, z)$.

Problem 7.3. Find the analog of Eq. (7.53) for the Gaussian window function.

In the physically interesting case of nearly scale-invariant primordial spectrum, we have at low momenta that $P(k) \propto k$, while the behavior at high momenta is $P(k) \propto k^{-3} \log^2 k$, see (7.3). Hence, the integral (7.53) is convergent. The result of numerical integration in the linear theory is shown in Fig. 7.8. We see from this figure that the protohalos of first stars, whose mass is $M \gtrsim 10^5 M_\odot$, are indeed formed at $z \sim 10$. The formation of clusters of galaxies has started fairly recently. We also see that the number of gravitationally bound systems of masses exceeding $M \sim 10^{14} M_\odot$ is small in the present Universe. These objects are being formed right now. As we pointed out in Section I.1.1, these objects (large clusters) are fairly loose: the density contrast in them is roughly of order 1. In fact, even larger structures, which have started to form only recently, are filaments. These are strongly elongated overdense regions, which usually connect clusters of galaxies. These filaments probably contain half of baryons and dark matter particles. Formation of filaments is not at all a spherically symmetric process, and we do not consider it in this book.

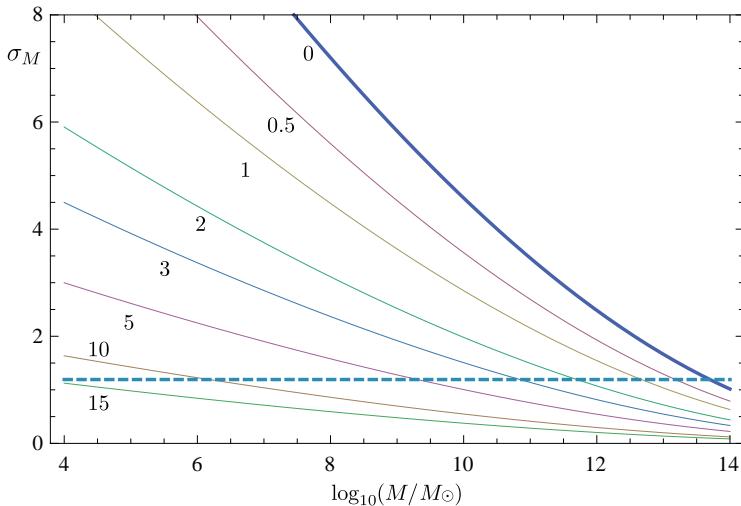


Fig. 7.8 The variance σ_M of the smoothed density contrast calculated in the linear theory for the Λ CDM model with cosmological parameters given in Section 2.1.2. Redshifts are shown by numbers near the curves. Thick line refers to the present epoch. The lower the curve, the earlier epoch it refers to, in accordance with (7.57). The horizontal dashed line shows the value $\sigma_M = \delta_c/\sqrt{2} \approx 1.192$ at which the exponent in the suppression factor of the differential spectrum (7.60) is equal to 1. The region above this line corresponds to intense structure formation.

Let us give simple analytical estimate for the integral (7.53) that determines the dependence of σ_R on the mass of a halo. This integral is saturated mostly at momenta $k \lesssim a_0/R$. Technically, this is due to the fall off in the spherical Bessel function j_1 at $k > a_0/R$ (see details in Appendix F.1). The physical reason is that high momenta correspond to small wavelengths, so the contributions of modes with $k \gg a_0/R$ cancel out when averaged over the large volume $V \sim (R/a_0)^3$. This is consistent with the general expectation that short scale perturbations are irrelevant for the formation of large halos. It follows from the estimates given in Section 1.3 that the heaviest structures correspond to the size R of a few tens of Mpc, so that for all perturbations of interest we have $R < a_0\eta_{eq}$, see (2.24). The main contribution to the integral (7.53) comes from the region $\eta_{eq}^{-1} \lesssim k \lesssim a_0/R$. The integral is logarithmic (the relevant integration region is large in the variable $\log(kR/a_0)$). Therefore, the last factor in (7.53) can be set equal to 1. We use the high-momentum asymptotics (7.21) for the power spectrum, i.e., $\mathcal{P}(k, z) = F^2(z) \mathcal{P}_\Phi(k) \log^2(0.2k\eta_{eq})$. Then the integral is saturated at $k \sim a_0/R$, and we get

$$\begin{aligned}\sigma_R^2(z) &\simeq F^2(z) \mathcal{P}_\Phi \int_{\eta_{eq}^{-1}}^{a_0/R} \frac{dk}{k} \log^2(0.2k\eta_{eq}) \\ &\simeq F^2(z) \mathcal{P}_\Phi \cdot \frac{1}{3} \log^3(0.2a_0\eta_{eq}/R).\end{aligned}\quad (7.54)$$

This is the desired expression; it is valid in the leading logarithmic approximation, i.e., when $\log(0.2a_0\eta_{eq}/R)$ is a large parameter. For nearly flat primordial spectrum, the variance σ_R decreases as R grows, $\sigma_R^2(z) \propto \log^3(R^{-1})$. Corrections to the result (7.54), which are formally suppressed by the parameter $\log^{-1}(0.2a_0\eta_{eq}/R)$, are actually rather large, see Problem 7.5. For large overdensity $\sigma_R \gtrsim 1$, numerical calculation gives the behavior $\sigma_R^2(z) \propto \log^{2.5}(R^{-1})$, and hence

$$\sigma_M^2(z) \propto \log^{2.5}(M^{-1}). \quad (7.55)$$

This dependence is in reasonable agreement with the behavior shown in Fig. 7.8.

Problem 7.4. Show that the result (7.54) is indeed valid in the leading logarithmic approximation. Hint: Use the quantity $\log(0.2k\eta_{eq})$ as the integration variable in (7.53).

Problem 7.5. Find the correction of order $\log^2(R^{-1})$ to the result (7.54). Hint: Show first that this correction comes solely from the region $k \sim a_0/R$.

We have seen that for flat primordial spectrum, the power spectrum $P(k, z)$ at $k/a \gg 0.01 \text{ Mpc}^{-1}$ depends on momentum exclusively through the logarithmic factor in (7.21). The dependence on redshift is contained in the function $F(z) \propto \frac{g(z)}{1+z}$, where $g(z)$ depends rather weakly on the ratio Ω_M/Ω_Λ . Hence, the power spectrum is determined mostly by one parameter, the overall amplitude. Traditionally, this parameter is chosen as $\sigma_8 \equiv \sigma_{R_8}(z = 0)$, where $R_8 = 8 h^{-1} \text{ Mpc}$. This scale

corresponds to the mass $M \simeq 2.5 \cdot 10^{14} M_{\odot}$. Of course, the parameter σ_8 is proportional to the primordial amplitude Δ_{Φ} (with the proportionality coefficient depending on the cosmological parameters). Therefore, the value of σ_8 is determined from both deep galaxy surveys and CMB observations. The existing data give the value [2]

$$\sigma_8 = 0.812 \pm 0.026.$$

This value is consistent with the value of the amplitude Δ_{Φ} given in Section 5.4.

Problem 7.6. Show that the last statement is indeed valid. Hint: When calculating the integral (7.53), use the numerical fit for the spectrum $P(k)$ given in the end of this Section.

Let us make use of the above value of σ_8 and numerical fit (7.55) to find σ_M for scales which are deep in the non-linear regime today. For the present Universe and mass scale of large galaxies, $M = M_* \simeq 10^{12} M_{\odot}$, Fig. 7.8 gives the estimate $\sigma_{M_*}(0) \simeq 2.5$. Then we find from (7.55) the following estimate for smaller masses and the present Universe,

$$\begin{aligned} \sigma_M(z=0) &\simeq \sigma_{M_*} \frac{\log^{1.25} \left[0.2a_0 \eta_{eq} \left(\frac{4\pi\rho_{M,0}}{3M} \right)^{1/3} \right]}{\log^{1.25} \left[0.2a_0 \eta_{eq} \left(\frac{4\pi\rho_{M,0}}{3M_*} \right)^{1/3} \right]} \\ &\simeq 0.25 \sigma_8 \log^{1.25} \left(\frac{2.3 \cdot 10^{15} M_{\odot}}{M} \right), \end{aligned} \quad (7.56)$$

where we use the traditional parameter $\sigma_8 \simeq 0.3 \sigma_{M_*}$ in the last formula. This result is in reasonable agreement with Fig. 7.8. The variance at redshift z is

$$\sigma_M(z) = \frac{g(z)}{g(0)} \frac{1}{1+z} \sigma_M(z=0). \quad (7.57)$$

Since the contrast $\delta_R(\mathbf{x}, t)$ is a random variable, overdense regions enter the non-linear regime earlier or later at different places in space. The probability that a given region of radius R has smoothed energy contrast exceeding δ_c is given by the probability integral

$$P(\delta_R > \delta_c) = \frac{1}{\sqrt{2\pi}} \int_{\delta_c}^{\infty} \frac{d\delta}{\sigma_R} e^{-\frac{\delta^2}{2\sigma_R^2}} = \frac{1 - erf\left(\frac{\delta_c}{\sqrt{2}\sigma_R}\right)}{2}. \quad (7.58)$$

This function determines the fraction of matter that resides in collapsed structures of mass $M \geq M(R)$, see (7.51). Let us denote by $dN(M)$ the average number density of structures in the mass interval $(M, M + dM)$. Then the mass fraction of structures of masses exceeding $M(R)$ is given by

$$\frac{1}{\rho_M} \int_{M(R)}^{\infty} M' dN(M') = 2P(\delta_R > \delta_c). \quad (7.59)$$

The factor 2 in the right hand side of Eq. (7.59) is introduced on the following grounds. It is clear that in the limit $M \rightarrow 0$, $\sigma_R \rightarrow \infty$, all particles reside in structures. On the other hand, using the formula (7.58) one would obtain $P = 1/2$ in this limit. The latter result has a simple interpretation: every point belongs to either overdense or underdense region, and formula (7.58) counts overdense regions only. On the other hand, in the limit considered, all matter has concentrated in structures, while underdense regions have become completely empty. The factor 2 restores this property. We had to introduce this factor superficially; this is because the linear theory is not valid at the late stage of structure formation: this theory simply discards matter in underdense regions. The discussion on more formal level can be found in Ref. [27].

It is clear from the derivation of (7.59) that this formula does not account for smaller mass objects inside larger ones. This is also seen from the following formal argument. Let the size R_c be such that the probability $2P(\delta_{R_c} > \delta_c)$ is nearly 1. Then lowering the limit of integration over masses in (7.59) should not add anything to the value of the integral. This would mean that there are no objects of smaller masses whatsoever. According to the discussion in Section 7.2.1, this is incorrect: small mass objects are formed earlier, so they may well exist inside larger structures. Thus, $dN(M)$ in (7.59) is the number density of *isolated* gravitationally bound structures of masses near M .

As the first application of the formula (7.59), let us estimate the masses of objects that confine considerable fraction of matter in the present Universe. We note that $2P(\delta_R > \delta_c) \simeq 0.32$ for $\sigma_R = \delta_c$ and $2P(\delta_R > \delta_c) \simeq 0.05$ for $\sigma_R = 0.5\delta_c$. We extract $\sigma_M(0)$ from Fig. 7.8, make use of (7.48) and find the masses at which $\sigma_M(0) = \delta_c$ and $\sigma_M(0) = 0.5\delta_c$,

$$M_{32\%} \simeq 1 \cdot 10^{13} M_\odot, \quad M_{5\%} \simeq 2 \cdot 10^{14} M_\odot,$$

where the notation is self-evident. The first of these values is characteristic of a small group of galaxies (similar to our Local Group), while the second one is typical for a large cluster. Hence, galaxies and their groups contain large fraction of matter (30% according to our estimate, and slightly less in reality), while large clusters of galaxies host much smaller fraction, a few percent.¹⁰ Despite the approximate character of our estimates, they agree with observations reasonably well (better than within a factor of 2), see, e.g., Fig. 7.10.

Problem 7.7. *Making use of Fig. 7.10, estimate the mass fraction of matter in the present Universe that resides in clusters of galaxies with $M \geq 2 \cdot 10^{14} M_\odot$.*

¹⁰It is worth noting that the largest *spherical* structures in the present Universe — groups and clusters of galaxies — comprise only about 10% of all galaxies and not more than a half of all baryons. The rest of matter is in more complicated non-spherical structures (the simplest among them are: filaments, which have elongated shape, and sheets). So, our approach based on spherical approximation is actually rather crude.

Problem 7.8. Estimate redshift at which 5% of matter collapses into halos of mass $M \geq 10^5 M_\odot$, thus refining (7.47).

Let us continue with the mass spectrum of structures, still making use of (7.59). We take the derivative with respect to M and obtain for the differential abundance

$$\begin{aligned} \frac{dN(M)}{dM} &= -\frac{\rho_M}{M} \frac{d\sigma_R}{dM} \frac{1}{\sigma_R^2} \cdot \frac{2}{\sqrt{2\pi}} \int_{\delta_c}^{\infty} d\delta \cdot \left(\frac{\delta^2}{\sigma_R^2} - 1 \right) \cdot e^{-\frac{\delta^2}{2\sigma_R^2}} \\ &= -\frac{2\delta_c}{\sqrt{2\pi}\sigma_M^2} \frac{\rho_M}{M^2} \frac{d\sigma_M}{d\log M} \cdot e^{-\frac{\delta_c^2}{2\sigma_M^2}}. \end{aligned} \quad (7.60)$$

This formula works both for the present epoch and for non-zero redshift. In the latter case, not only σ_M depends on z , but also $\rho_M(z) \propto (1+z)^3$. The function $(dN/dM)(M, z)$ is then the number of objects per unit mass per unit *physical* volume, while $(1+z)^{-3}dN/dM$ is the differential mass spectrum of structures in comoving volume. The latter function is shown in Fig. 7.9. The curves approach each other when σ_M reaches δ_c . In that case, the exponential factor in (7.60) is close to 1, and $dN/d\log M \propto M^{-1}$ modulo logarithmic factor. Hence, decimal intervals of masses that differ from each other by an order of magnitude contain the numbers of objects which differ by an order of magnitude too: there must be 10 times more galaxies of masses $10^9 M_\odot$ as compared to galaxies of masses $10^{10} M_\odot$. In other words, structures in decimal intervals of masses (say $(10^8 - 10^9) M_\odot$ and $(10^{10} - 10^{11}) M_\odot$) host approximately the same fraction of matter. These conclusions do not, of course, account for possible destruction of galaxies in their interactions. The redshift dependence is $(1+z)^{-3}dN/dM \propto (1+z)$; this is not visible in Fig. 7.9. At larger masses, when $\sigma_M \ll \delta_c$, the mass spectrum is exponentially suppressed, and the curves bend down. This effect takes place for smaller mass at larger redshift, according to the dependence $\sigma_M \propto (1+z)^{-1}$.

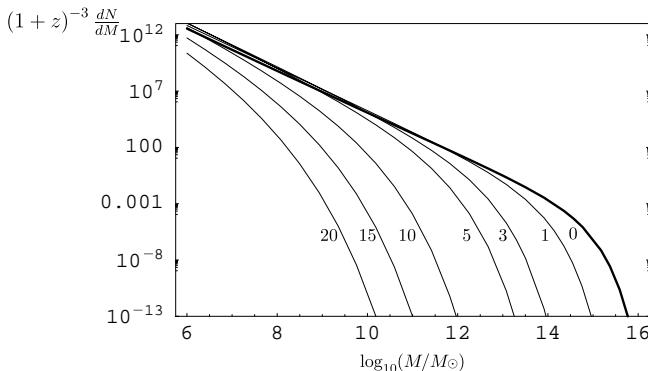


Fig. 7.9 Differential number of structures per comoving volume as a function of mass at different redshifts (shown by numbers near curves) in Λ CDM model. The normalization is arbitrary. The cosmological parameters are given in Section 2.1.2. Thick line refers to the present epoch.

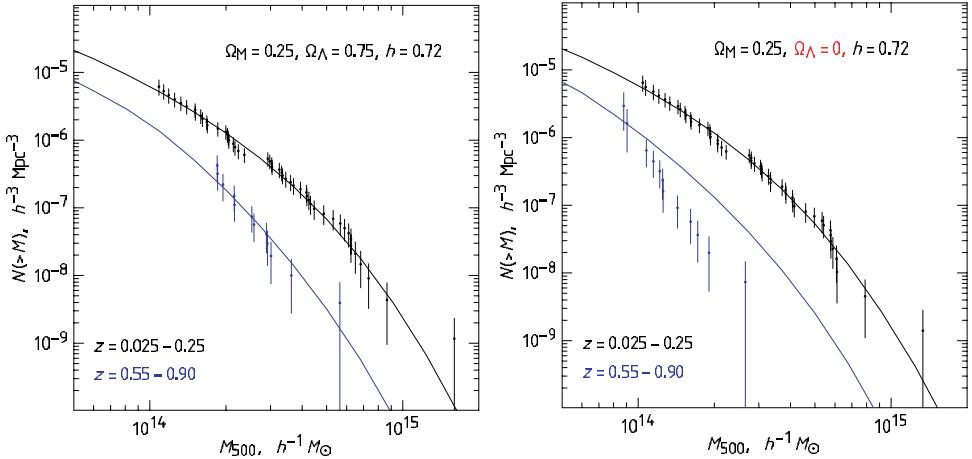


Fig. 7.10 Results for clusters of galaxies obtained by the analysis of observations by CHANDRA X-ray telescope and their comparison with predictions of cosmological models with different parameters [30]. Lower curves refer to more distant objects with larger redshifts.

The formula (7.60) and its generalizations (see the end of this Section) predict the abundance of structures in the Universe. This prediction depends on the amplitude of the power spectrum (i.e., on σ_8) and on the cosmological expansion rate at relatively recent past (i.e., on Ω_M , Ω_Λ and Ω_{curv}). Also, the abundance of small structures depends on the type of dark matter (cold versus warm, see Section 8.4). One of the observables is the number of objects with mass exceeding a given value per redshift interval,

$$N(\text{mass} > M) = \int_{z_1}^{z_2} \frac{dz}{(1+z)H(z)} \cdot D_a^2(z)d\Omega \int_M^\infty dM' \frac{dN(M', z)}{dM'}. \quad (7.61)$$

Here $D_a(z)$ is the angular diameter distance, so that $D_a^2 d\Omega$ is the physical area of a part of sphere seen at solid angle $d\Omega$; $(1+z)^{-1}H^{-1}(z)dz$ is the physical radial distance¹¹ between spheres of redshifts z and $z+dz$. The dependence on the cosmological parameters enters here through both $H^{-1}(z)D_a^2(z)$ and $\sigma_M(z)$. An example of the predicted number of clusters in comparison with observational data is shown in Fig. 7.10. Figure 7.11 shows the results for the cosmological parameters obtained in this way and their comparison with the results of CMB observations. We see that

¹¹The latter property follows from the fact the redshifts z and $z+dz$ of two photons emitted at physical distances l and $l+dl$ and arriving to the observer simultaneously, obey $dl = |dt| = |da/\dot{a}| = (1+z)^{-1}H^{-1}(z)dz$. Notice that the formula (7.61) can be written as

$$N(\text{mass} > M) = \int_{z_1}^{z_2} \frac{dz}{H(z)} \cdot r^2(z)d\Omega \int_M^\infty dM' \frac{1}{(1+z)^3} \frac{dN(M', z)}{dM'},$$

where $r(z)$ is the coordinate angular diameter distance, see Section I.4.7. The expression $r^2(z)H^{-1}(z)dzd\Omega$ is the comoving volume element, and $(1+z)^{-3}(dN/dM)dM$ is the number of objects per unit comoving volume. The latter is given by (7.60) with ρ_M equal to the present mass density.

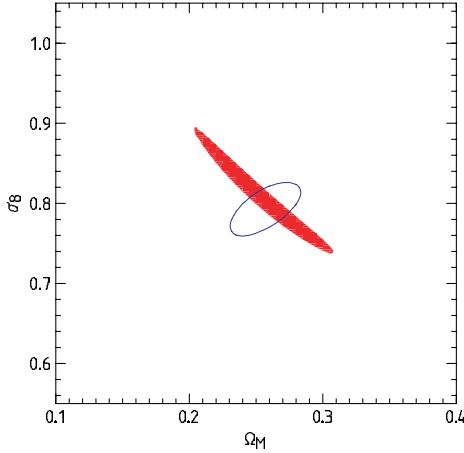


Fig. 7.11 Allowed region in the space of parameters (σ_8, Ω_M) in Λ CDM model, according to the analysis of clusters of galaxies [30] based on CHANDRA data (dark region) and the comparison with the result form the analysis of 5-year WMAP data (ellipse); the outlined regions are allowed at 68% C.L.

the measurements of structure abundance give important information on energy balance in the Universe, as well as on the spectrum of primordial perturbations. Data on structures are used in combined analysis for the determination of the cosmological parameters.

To end this Section, we give several formulas which are used for calculating structure formation with good accuracy, without full numerical modeling of this process. The analytical formula for $P(k, z)$ which gives good fit to the real spectrum and contains the dependence on the cosmological parameters is [26]

$$P(k, z) = \frac{9\mathcal{P}_\Phi k^{n_s}}{25} \frac{g^2(z)}{a(z) H^4(z) \Omega_M^2} T^2(K),$$

$$T(K) \equiv \frac{\log(1 + 2.34K)}{2.34K} (1 + 3.89K + (16.1K)^2 + (5.46K)^3 + (6.71K)^4)^{-1/4},$$

$$K \equiv \frac{k \exp(2\Omega_B)}{\Omega_M h^2} \text{ Mpc.}$$

More accurate condition for the collapse, as compared to the main text, takes into account deviations from spherical symmetry and has the form [28]

$$\delta_R > \delta_c \cdot \sqrt{A} \cdot \left(1 + \frac{\beta}{(A\nu)^\gamma}\right),$$

where $\nu \equiv \delta_c^2/\sigma_R^2$, and the parameters entering here are obtained by fitting numerical data [29]: $(A, \beta, \gamma) = (0.707, 0.47, 0.615)$. The refined expression for the differential spectrum, that replaces (7.60) is [28, 29]

$$\frac{dN}{dM} = -B\sqrt{A} \cdot \left(1 + \frac{\beta}{(A\nu)^\gamma}\right) \sqrt{\frac{2}{\pi}} \frac{\rho_M}{M} e^{-A\nu^2/2} \frac{d\sigma_R}{dM} \frac{\delta_c}{\sigma_R^2},$$

where $B = 0.322$. This value of B may be viewed as the fit to complete numerical simulation of structure formation.

This page is intentionally left blank

Chapter 8

*Beyond Ideal Fluid Approximation

In previous Chapters we have studied scalar perturbations under the assumption that the cosmic medium can be described in the ideal fluid approximation. This is a valid approximation for major components of the medium at various cosmological epochs. However, some components do not behave as ideal fluid. The important examples are neutrinos (they decouple completely at temperature of a few MeV), photons at the late stage of recombination and at later epoch, and possibly warm dark matter. Effects beyond the ideal fluid approximation often suppress perturbations, and hence the gravitational potentials. This has important consequences which can be tested observationally. We consider several effects of this sort in this Chapter.

8.1 Distribution Functions and Boltzmann Equation in Curved Space-Time

In what follows we encounter situations when gases of particles are dilute, while the spatial scales of gravitational fields are large. In these cases, the evolution of particle distribution functions is determined by the Boltzmann equation in mean gravitational field.

We begin with the derivation of the energy-momentum tensor for medium consisting of one type of particles. We consider massive particles for definiteness, and write the action for a system of N free particles of mass m in space-time with metric $g_{\mu\nu}(x)$,

$$S = -m \sum_{i=1}^N \int ds_{(i)} = -m \sum_i \int ds_{(i)} \sqrt{g_{\mu\nu}(x_{(i)}) \frac{dx_{(i)}^\mu}{ds_{(i)}} \frac{dx_{(i)}^\nu}{ds_{(i)}}}, \quad (8.1)$$

where $s_{(i)}$ are intervals along the world line of each of N particles, $x_{(i)} = x_{(i)}(s_{(i)})$ are coordinates of the world line. The contravariant 4-momentum of the i -th particle

is defined as follows,

$$P_{(i)}^\mu = m \frac{dx_{(i)}^\mu}{ds_{(i)}}. \quad (8.2)$$

It is convenient to also use the covariant 4-momentum $P_{(i)\mu} = g_{\mu\nu} P_{(i)}^\nu$. The evolution of the contravariant momentum is determined by the geodesic equation, see Section I.A.7,

$$\frac{d^2 x_{(i)}^\mu}{ds_{(i)}^2} + \Gamma_{\rho\lambda}^\mu \frac{dx_{(i)}^\rho}{ds_{(i)}} \frac{dx_{(i)}^\lambda}{ds_{(i)}} = 0, \quad (8.3)$$

which is the extremum equation for the action (8.1). The derivative of the covariant momentum is

$$\frac{dP_{(i)\mu}}{ds_{(i)}} = \frac{d(g_{\mu\nu} P_{(i)}^\nu)}{ds_{(i)}} = P_{(i)}^\nu \frac{dg_{\mu\nu}(x_{(i)})}{ds_{(i)}} + g_{\mu\nu} \frac{d^2 x_{(i)}^\nu}{ds_{(i)}^2}.$$

We notice that

$$\frac{dg_{\mu\nu}(x_{(i)})}{ds_{(i)}} = \frac{dx_{(i)}^\lambda}{ds_{(i)}} \partial_\lambda g_{\mu\nu} = \frac{P_{(i)}^\lambda}{m} \partial_\lambda g_{\mu\nu},$$

and find from (8.3)

$$m \frac{dP_{(i)\mu}}{ds_{(i)}} = P_{(i)}^\nu P_{(i)}^\lambda \partial_\lambda g_{\mu\nu} - g_{\mu\nu} \Gamma_{\lambda\rho}^\nu P_{(i)}^\lambda P_{(i)}^\rho.$$

Finally, we use the fact that the covariant derivative of metric vanishes, and obtain the equation for the covariant 4-momentum of the i -th particle,

$$m \frac{dP_{(i)\mu}}{ds_{(i)}} = \Gamma_{\mu\lambda}^\rho P_{(i)}^\lambda P_{(i)\rho}. \quad (8.4)$$

One can rewrite the expression for the action (8.1) by introducing formally the integration over (x^μ, P_ν) ,

$$S = -m \int d^4x d^4P \sum_i \int ds_{(i)} \delta[x - x_{(i)}(s_{(i)})] \delta[P - P_{(i)}(s_{(i)})], \quad (8.5)$$

where P is understood as the covariant momentum P_μ . One can see that the last integral in (8.5) is scalar under general coordinate transformations $x^\mu \rightarrow \tilde{x}^\mu(x)$; it is important that it involves covariant 4-momentum $P_{(i)\mu}$. Let us define the particle distribution function in 8-dimensional phase space,

$$F(x^\mu, P_\nu) = \left\langle \sum_i \int ds_{(i)} \delta[x - x_{(i)}(s_{(i)})] \delta[P - P_{(i)}(s_{(i)})] \right\rangle, \quad (8.6)$$

where the right hand side is the ensemble average. The distribution function is a scalar, and its definition (8.6) is valid for massless particles as well.

The energy-momentum tensor is obtained, as usual, by varying the action (8.1) with respect to the metric,

$$\begin{aligned} T^{\mu\nu} &= -\frac{2}{\sqrt{-g}} \frac{\delta S}{\delta g_{\mu\nu}} = \sum_i \int \frac{1}{\sqrt{-g}} \frac{ds_{(i)} \cdot m \frac{dx_{(i)}^\mu}{ds_{(i)}} \frac{dx_{(i)}^\nu}{ds_{(i)}}}{\sqrt{g_{\lambda\rho}(x_{(i)})} \frac{dx_{(i)}^\lambda}{ds_{(i)}} \frac{dx_{(i)}^\rho}{ds_{(i)}}} \delta(x - x_{(i)}(s)) \\ &= \sum_i \int \frac{ds_{(i)}}{\sqrt{-g}} \frac{P_{(i)}^\mu P_{(i)}^\nu}{m} \delta(x - x_{(i)}(s)). \end{aligned}$$

We introduce formally the integration over 4-momentum P_ν in a similar way as in (8.5), average over ensemble, use the definition of the distribution function (8.6) and find

$$T^{\mu\nu} = \int \frac{d^4 P}{\sqrt{-g}} \frac{P^\mu P^\nu}{m} F(x^\lambda, P_\rho), \quad (8.7)$$

where like in (8.5), the integration measure involves the covariant momentum. Both the distribution function F and phase space volume element $d^4 P / \sqrt{-g}$ are scalars, hence the integral in the right hand side of (8.7) is indeed a tensor.

The expression (8.7) for the energy-momentum in terms of 8-dimensional integral is inconvenient. Particles in the gas have definite masses, so the 4-momentum obeys

$$g^{\mu\nu} P_\mu P_\nu = m^2. \quad (8.8)$$

Hence, the energy and 3-momentum of each particle are not independent variables. Because of that, it is convenient to introduce the time-dependent distribution function in 6-dimensional phase space $f(x^0, x^i, P_j) \equiv f(x^\lambda, P_\rho)$ by

$$F(x^\lambda, P_\rho) = f(x^\lambda, P_j) \theta(P_0) \delta(\sqrt{g^{\mu\nu} P_\mu P_\nu} - m), \quad (8.9)$$

where $\theta(P_0)$ is step function. We note that the distribution function f is also a scalar, since both F and δ -function are scalars, whereas the step function is a scalar under physically relevant coordinate transformations that leave particle energies positive. By inserting Eq. (8.9) into Eq. (8.7) and integrating over P_0 we obtain finally

$$T^{\mu\nu}(x) = \int \frac{d^3 P_j}{\sqrt{-g}} \frac{P^\mu P^\nu}{P^0} f(x, \mathbf{P}). \quad (8.10)$$

Hereafter we use the notation $d^3 P_j$ to emphasize that the integration measure involves 3-dimensional components of the covariant momentum. To avoid confusion we point out that we always use the convention that the phase space volume element is positive. We also emphasize that hereafter we use the term distribution function for the dimensionless function $f(x, \mathbf{P})$ defined by (8.9). Since the expression (8.10) does not contain mass m explicitly, it is valid for massless particles as well.

Problem 8.1. Show explicitly that the right hand side of Eq. (8.10) is a tensor under coordinate transformations leaving intact the sign of energy. Thus, $d^3 P_j / (P^0 \sqrt{-g})$ is the invariant volume element in the momentum space.

Problem 8.2. Obtain the formula (8.10) for the gas of non-interacting massless particles.

The physical meaning of the function $f(x^\mu, P_j)$ can be understood in the following way. Consider locally Lorentz frame in the vicinity of the point x^μ and hypersurface $x^0 = \text{const}$ in this vicinity. According to the definition (8.6), one writes

$$\begin{aligned} F(x^0; x^i, P_\nu) = & \left\langle \sum_i \int ds_{(i)} \delta(x^0 - x_{(i)}^0(s_{(i)})) \right. \\ & \left. \times \delta(\mathbf{x} - \mathbf{x}_{(i)}(s_{(i)})) \delta(P - P_{(i)}(s_{(i)})) \right\rangle. \end{aligned}$$

One integrates over $ds_{(i)}$ using the definition (8.2) and obtains

$$F(x^0; x^i, P_\nu) = \left\langle \sum_i \frac{m}{P_{(i)}^0} \delta(\mathbf{x} - \mathbf{x}_{(i)}) \delta(\mathbf{P} - \mathbf{P}_{(i)}) \delta(P^0 - P_{(i)}^0) \right\rangle,$$

where $\mathbf{x}_{(i)}$ and $P_{(i)\nu}$ are spatial coordinate and covariant momentum of the i -th particle at a given moment of time x^0 . Making use of (8.8) with positive $P_{(i)}^0$ one casts the latter formula in the following form,

$$F(x^0; x^i, P_\nu) = \left\langle \sum_i \delta(\mathbf{x} - \mathbf{x}_{(i)}) \delta(\mathbf{P} - \mathbf{P}_{(i)}) \right\rangle \theta(P_0) \delta(\sqrt{g^{\mu\nu} P_\mu P_\nu} - m).$$

One compares this expression to (8.9) and observes that in the locally Lorentz frame, the function $f(x^0; \mathbf{x}, \mathbf{P})$ at time x^0 is given by

$$f(x^0; \mathbf{x}, \mathbf{P}) = \left\langle \sum_i \delta(\mathbf{x} - \mathbf{x}_{(i)}) \delta(\mathbf{P} - \mathbf{P}_{(i)}) \right\rangle.$$

Thus, this function is the number density of particles in the 6-dimensional phase space (\mathbf{x}, \mathbf{P}) . This is why it is legitimate to call $f(x^\mu, P_j)$ the distribution function.

Let us now derive the Boltzmann equation that governs the evolution of the distribution function. We begin with the case when particle interactions are totally absent, so we are dealing with the gas of collisionless particles. Let us make use of the obvious identity,

$$0 = \sum_i \int ds_{(i)} \cdot \frac{d}{ds_{(i)}} [m \delta(x - x_{(i)}(s_{(i)})) \cdot \delta(P - P_{(i)}(s_{(i)}))].$$

Now,

$$\frac{\partial}{\partial x_{(i)}^\mu} \delta(x - x_{(i)}) = -\frac{\partial}{\partial x^\mu} \delta(x - x_{(i)}), \quad \frac{\partial}{\partial P_{(i)\mu}} \delta(P - P_{(i)}) = -\frac{\partial}{\partial P_\mu} \delta(P - P_{(i)}).$$

Hence,

$$0 = \sum_i \int ds_{(i)} \left[m \frac{dx_{(i)}^\mu}{ds_{(i)}} \cdot \delta(P - P_{(i)}(s_{(i)})) \cdot \frac{\partial}{\partial x^\mu} \delta(x - x_{(i)}(s_{(i)})) + m \frac{dP_{(i)\mu}}{ds_{(i)}} \cdot \delta(x - x_{(i)}(s_{(i)})) \frac{\partial}{\partial P_\mu} \delta(P - P_{(i)}(s_{(i)})) \right]. \quad (8.11)$$

We use the definition (8.2) and geodesic equation (8.4) to write the equality (8.11) as follows,

$$0 = \sum_i \int ds_{(i)} \frac{\partial}{\partial x^\mu} [P_{(i)}^\mu \cdot \delta(x - x_{(i)}(s_{(i)})) \delta(P - P_{(i)}(s_{(i)}))] + \sum_i \int ds_{(i)} \frac{\partial}{\partial P_\mu} [\Gamma_{\mu\rho}^\nu (x_{(i)}) P_{(i)\nu} P_{(i)}^\rho \cdot \delta(x - x_{(i)}(s_{(i)})) \delta(P - P_{(i)}(s_{(i)}))]. \quad (8.12)$$

The ensemble average of the latter equation gives the Boltzmann equation for $F(x^\mu, P_\nu)$,

$$\frac{\partial}{\partial x^\mu} (FP^\mu) + \frac{\partial}{\partial P_\mu} (F\Gamma_{\mu\rho}^\nu P^\rho P_\nu) = 0. \quad (8.13)$$

This equation can be simplified. To this end, we evaluate the derivatives explicitly and obtain

$$P^\mu \frac{\partial F}{\partial x^\mu} + \Gamma_{\mu\rho}^\nu P^\rho P_\nu \frac{\partial F}{\partial P_\mu} + \left(\Gamma_{\nu\mu}^\nu g^{\mu\rho} P_\rho + \Gamma_{\rho\mu}^\nu g^{\mu\rho} P_\nu + \frac{\partial g^{\nu\mu}}{\partial x^\mu} P_\nu \right) F = 0. \quad (8.14)$$

The expression in parenthesis in the right hand side of (8.14) vanishes, since it is proportional to the covariant derivative of metric. Hence, Eq. (8.13) becomes

$$P^\mu \frac{\partial F}{\partial x^\mu} + \Gamma_{\mu\rho}^\nu P^\rho P_\nu \frac{\partial F}{\partial P_\mu} = 0. \quad (8.15)$$

We now make use of the definition (8.9) and integrate over P_0 to obtain

$$\begin{aligned} & \int dP_0 \theta(P_0) \delta(\sqrt{g^{\mu\nu} P_\mu P_\nu} - m) \left(P^\mu \frac{\partial}{\partial x^\mu} + \Gamma_{j\rho}^\nu P^\rho P_\nu \frac{\partial}{\partial P_j} \right) f(x^\lambda, P_i) \\ & + \int dP_0 \frac{\partial \theta(P_0)}{\partial P_0} f \Gamma_{0\rho}^\nu P_\nu P^\rho \delta(\sqrt{g^{\mu\nu} P_\mu P_\nu} - m) \\ & + \int dP_0 \theta(P_0) f \left(P^\mu \frac{\partial}{\partial x^\mu} + \Gamma_{\mu\rho}^\nu P^\rho P_\nu \frac{\partial}{\partial P_\mu} \right) \delta(\sqrt{g^{\mu\nu} P_\mu P_\nu} - m) = 0. \end{aligned} \quad (8.16)$$

Since the derivative of the step function is equal to δ -function, the second term in (8.16) vanishes for particles of non-zero energy. The third term in (8.16) is

$$\int dP_0 \theta(P_0) \frac{f}{\sqrt{g^{\mu\nu} P_\mu P_\nu}} \cdot \left(\frac{1}{2} P^\mu P_\lambda P_\rho \partial_\mu g^{\lambda\rho} + \Gamma_{\mu\rho}^\nu P^\rho P_\nu P^\mu \right) \delta' \left(\sqrt{g^{\mu\nu} P_\mu P_\nu} - m \right).$$

The expression in parenthesis multiplying δ' vanishes, again because the metric tensor is covariantly constant. Thus, only the first term in (8.16) is non-trivial. Upon integrating over energy we obtain the Boltzmann equation for the distribution function $f(x^\lambda, P_i)$,

$$P^\mu \frac{\partial f}{\partial x^\mu} + \Gamma_{j\rho}^\nu P^\rho P_\nu \frac{\partial f}{\partial P_j} = 0, \quad (8.17)$$

where P_0 is expressed through P_i via the mass shell condition (8.8).

The right hand side of Eq. (8.17) vanishes only in the case of collisionless gas we have studied so far. Equation (8.17) in this particular case is called the Liouville equation. In locally Lorentz frame, Eq. (8.17) takes the form, which is standard for the Liouville equation,

$$\frac{\partial f}{\partial t} + \mathbf{u} \frac{\partial f}{\partial \mathbf{x}} = 0, \quad (8.18)$$

where $u^i = P^i/P^0$ is the physical velocity. In curved space-time, the left hand side of Eq. (8.17) is

$$P^\mu \left[\frac{\partial f}{\partial x^\mu} + \Gamma_{j\mu}^\nu P_\nu \frac{\partial f}{\partial P_j} \right]. \quad (8.19)$$

The quantity in square brackets is actually a covariant vector under general coordinate transformations (we will give the proof in a moment), so the Liouville equation (8.17) is covariant.

The expression (8.19) involves the sum over spatial index j , so its covariance is not immediately clear. To see that it is in fact covariant, let us recall the transformation law of covariant momenta under coordinate transformation $x^\mu \rightarrow \tilde{x}^\mu(x^\lambda)$,

$$P_\nu \rightarrow \tilde{P}_\nu = P_\lambda \frac{\partial x^\lambda}{\partial \tilde{x}^\nu}, \quad (8.20)$$

The function f is a scalar, i.e.,

$$f(x, P) = f(\tilde{x}, \tilde{P}).$$

Therefore

$$\begin{aligned} \left(\frac{\partial}{\partial x^\mu} + \Gamma_{j\mu}^\nu P_\nu \frac{\partial}{\partial P_j} \right) f(x, P) &= \left(\frac{\partial}{\partial x^\mu} + \Gamma_{j\mu}^\nu P_\nu \frac{\partial}{\partial \tilde{P}_j} \right) f(\tilde{x}, \tilde{P}) \\ &= \frac{\partial f}{\partial \tilde{x}^\lambda} \frac{\partial \tilde{x}^\lambda}{\partial x^\mu} + \frac{\partial f}{\partial \tilde{P}_j} \frac{\partial \tilde{P}_j}{\partial x^\mu} + \Gamma_{j\mu}^\nu P_\nu \frac{\partial f}{\partial \tilde{P}_k} \frac{\partial \tilde{P}_k}{\partial P_j}. \end{aligned} \quad (8.21)$$

It follows from (8.20) that

$$\begin{aligned}\frac{\partial \tilde{P}_k}{\partial P_j} &= \frac{\partial x^j}{\partial \tilde{x}^k} + \frac{\partial P_0}{\partial P_j} \frac{\partial x^0}{\partial \tilde{x}^k}, \quad P_\lambda = \tilde{P}_\nu \frac{\partial \tilde{x}^\nu}{\partial x^\lambda}, \\ \frac{\partial \tilde{P}_k}{\partial x^\mu} &= \frac{\partial}{\partial x^\mu} \left(P_\lambda \frac{\partial x^\lambda}{\partial \tilde{x}^k} \right) = P_\lambda \frac{\partial^2 x^\lambda}{\partial \tilde{x}^\nu \partial \tilde{x}^k} \frac{\partial \tilde{x}^\nu}{\partial x^\mu} + \frac{\partial P_0}{\partial x^\mu} \frac{\partial x^0}{\partial \tilde{x}^k},\end{aligned}$$

while the mass shell condition (8.8) gives

$$\frac{\partial P_0}{\partial P_i} = -\frac{P^i}{P^0}, \quad P^\lambda \nabla_\mu P_\lambda = 0.$$

Hence,

$$\frac{\partial P_0}{\partial x^\mu} = \Gamma_{\mu 0}^\nu P_\nu - \frac{P^i \nabla_\mu P_i}{P^0}.$$

We see that the right hand side of (8.21) is

$$\begin{aligned}&\frac{\partial f}{\partial \tilde{x}^\lambda} \frac{\partial \tilde{x}^\lambda}{\partial x^\mu} + \left[P_\lambda \frac{\partial^2 x^\lambda}{\partial \tilde{x}^\nu \partial \tilde{x}^k} \frac{\partial \tilde{x}^\nu}{\partial x^\mu} + \left(\Gamma_{\mu 0}^\nu P_\nu - \frac{P^i \nabla_\mu P_i}{P^0} \right) \frac{\partial x^0}{\partial \tilde{x}^k} \right. \\ &\left. + \Gamma_{j\mu}^\nu P_\nu \left(\frac{\partial x^j}{\partial \tilde{x}^k} - \frac{P^j}{P^0} \frac{\partial x^0}{\partial \tilde{x}^k} \right) \right] \frac{\partial f}{\partial \tilde{P}_k}.\end{aligned}$$

The terms with P^0 in the denominator cancel out in this expression, and the rest is

$$\frac{\partial f}{\partial \tilde{x}^\lambda} \frac{\partial \tilde{x}^\lambda}{\partial x^\mu} + \left[\left(\frac{\partial^2 x^\nu}{\partial \tilde{x}^k \partial \tilde{x}^\lambda} \frac{\partial \tilde{x}^\lambda}{\partial x^\mu} + \Gamma_{\lambda\mu}^\nu \frac{\partial x^\lambda}{\partial \tilde{x}^k} \right) \frac{\partial \tilde{x}^\tau}{\partial x^\nu} \right] \tilde{P}_\tau \frac{\partial f}{\partial \tilde{P}_k}. \quad (8.22)$$

Finally, we recall the transformation law (I.A.16) of the Christoffel symbols under coordinate transformation and find that the quantity in square brackets in (8.22) is equal to

$$\tilde{\Gamma}_{k\lambda}^\tau \frac{\partial \tilde{x}^\lambda}{\partial x^\mu},$$

where $\tilde{\Gamma}_{k\lambda}^\tau$ are the Christoffel symbols in the new coordinate frame. Hence, we have

$$\left(\frac{\partial}{\partial x^\mu} + \Gamma_{j\mu}^\nu P_\nu \frac{\partial}{\partial P_j} \right) f(x, P) = \frac{\partial \tilde{x}^\lambda}{\partial x^\mu} \left(\frac{\partial}{\partial \tilde{x}^\lambda} + \tilde{\Gamma}_{j\lambda}^\nu P_\nu \frac{\partial}{\partial \tilde{P}_j} \right) f(\tilde{x}, \tilde{P}).$$

We conclude that the quantity in square brackets in (8.19) is indeed a covariant vector, so the Liouville equation (8.19) is indeed invariant under general coordinate transformations. The whole derivation has been given here to show that the fact that the function f is independent of P_0 , explicitly used when writing the Liouville equation (8.19) (there is no partial derivative over P_0), does not spoil the general coordinate invariance of this equation.

Problem 8.3. Show that any function of $P^2 \equiv P_\mu P^\mu$ is a formal solution to Eq. (8.15).

Equation (8.17) is definitely valid for collisionless matter *in the background metric*. It also remains valid in many cases when metric is determined by the medium itself. In these cases, Eq. (8.17), together with the Einstein equations for the metric, constitute the *Vlasov system of equations*. It is valid when the dominant force acting on a particle is due to long-ranged gravitational field (mean field) rather than due

to a few neighboring particles. This is indeed the case in cosmology, at least at the linear stage of the evolution of perturbations.

It is clear from (8.18) that for collisionless gas, Eq. (8.17) can be interpreted as particle number conservation in comoving phase space volume. This interpretation is instrumental in generalizing Eq. (8.17) to gases in which collisions between particles are important. We now turn to this generalization, which is the Boltzmann equation proper.

We consider two-particle collisions only, since these are the most relevant in the cosmological context (generalizing to multiparticle processes is straightforward). The collisions give rise to non-conservation of the number of particles in the comoving phase space volume. This phenomenon is conveniently studied in a locally Lorentz frame. The two-particle collisions we consider are

$$1 + 2 \rightarrow 1' + 2' \quad (8.23)$$

where particles 1, 2, 1', 2' are, in general, of different types. The object of interest is the number of collisions dN in spatial region $\mathbf{x} \pm d\mathbf{x}/2$ in the time interval $x^0 \pm dx^0/2$, occurring between particles 1 and 2 of momenta $\mathbf{P}_1 \pm d\mathbf{P}_1/2$ and $\mathbf{P}_2 \pm d\mathbf{P}_2/2$ and producing particles 1' and 2' of momenta $\mathbf{P}_{1'} \pm d\mathbf{P}_{1'}/2$ and $\mathbf{P}_{2'} \pm d\mathbf{P}_{2'}/2$. This number is proportional to the phase space number densities of particles 1 and 2, hence

$$\begin{aligned} dN = & \frac{|\mathcal{M}(\mathbf{P}_1, \mathbf{P}_2; \mathbf{P}_{1'}, \mathbf{P}_{2'})|^2}{P_1^0 P_2^0 P_{1'}^0 P_{2'}^0} d^3 P_{1'} d^3 P_{2'} \\ & \times d^4 x d^3 P_1 d^3 P_2 f_1(x, \mathbf{P}_1) f_2(x, \mathbf{P}_2). \end{aligned} \quad (8.24)$$

The quantity \mathcal{M} introduced in this way is a Lorentz scalar. According to the meaning of dN , this quantity is proportional to the amplitude of the transition (8.23). In general $|\mathcal{M}|^2$ is calculated by quantum field theory methods (see Refs. [31–35] for details). It is related to the cross section of the process (8.23) as follows,

$$|\mathcal{M}|^2 = (P_1 + P_2)^2 \delta(P_1 + P_2 - P_{1'} - P_{2'}) \frac{d\sigma}{d\Omega}, \quad (8.25)$$

where $d\sigma/d\Omega$ is the differential cross section per unit solid angle Ω in the center-of-mass frame. Indeed, by the definition of the differential cross section $d\sigma$, its product with the flux of incident particles 1 is the transition probability per unit time per one particle 2. In the locally Lorentz frame, the flux is equal to the product of the number density of incident particles and the relative velocity $|\mathbf{u}| = |\mathbf{P}_1/P_1^0 - \mathbf{P}_2/P_2^0|$. It follows from (8.24) that $|\mathcal{M}|^2$ is related to the cross section by

$$d\sigma = \frac{|\mathcal{M}|^2}{|\mathbf{u}| P_1^0 P_2^0} \frac{d^3 P_{1'} d^3 P_{2'}}{P_{1'}^0 P_{2'}^0}. \quad (8.26)$$

This gives (8.25). Indeed, inserting (8.25) into (8.26) and integrating over momenta, one obtains the identity.

Problem 8.4. Check the last statement on the previous page.

Kinematics of $2 \rightarrow 2$ processes involves only one variable¹ θ , the scattering angle in the reaction plane. In the center-of-mass frame we have

$$\frac{d\sigma}{d\Omega} = \frac{1}{2\pi} \frac{d\sigma}{d\cos\theta}, \quad (8.27)$$

and

$$\cos\theta = \frac{\mathbf{P}_1 \cdot \mathbf{P}_{1'}}{|\mathbf{P}_1| |\mathbf{P}_{1'}|}. \quad (8.28)$$

In this frame, $d\sigma/d\Omega$ is the function of $s = (P_1 + P_2)^2 = (P_1^0 + P_2^0)^2$ and $\cos\theta$,

$$\frac{d\sigma}{d\Omega} = \frac{d\sigma}{d\Omega}(s, \cos\theta). \quad (8.29)$$

To use the formula (8.25) in arbitrary reference frame, one makes use of the cross section calculated in the center-of-mass frame and expresses s and $\cos\theta$ in terms of Lorentz-invariants. For particles of one and the same mass (i.e., for processes like $e^+e^- \rightarrow e^+e^-$) the invariant expressions are

$$s = (P_1 + P_2)_\mu (P_1 + P_2)_\nu g^{\mu\nu} \equiv (P_1 + P_2)^2, \quad (8.30)$$

$$\cos\theta = \frac{(P_1 - P_2)_\mu (P_{1'} - P_{2'})_\nu g^{\mu\nu}}{(P_1 - P_2)^2}. \quad (8.31)$$

The formula (8.30) is valid for the scalar s in the general case of different masses. The general expression for the scattering angle is obtained by using the complete set of the Mandelstam variables s, t, u , where

$$t = (P_1 - P_{1'})^2, \quad u = (P_1 - P_{2'})^2.$$

These variables obey

$$s + t + u = m_1^2 + m_2^2 + m_{1'}^2 + m_{2'}^2.$$

The generalization of the formula (8.31) is

$$\cos\theta = \frac{u - t}{u + t}.$$

Hence, the quantity $d\sigma/d\Omega$ entering the right hand side of (8.25) is explicitly defined.

We are interested in the fate of particles of the type 1. So, we integrate the right hand side of (8.24) over momenta of other particles and obtain the number of particles of type 1 exiting the phase space volume $d^3x d^3P_1$,

$$\begin{aligned} \Delta N_1^- &= d^4x \frac{d^3P_1}{P_1^0} \int \frac{d^3P_2}{P_2^0} \frac{d^3P_{1'}}{P_{1'}^0} \frac{d^3P_{2'}}{P_{2'}^0} f_1(x, \mathbf{P}_1) \\ &\quad \times f_2(x, \mathbf{P}_2) \cdot |\mathcal{M}(\mathbf{P}_1, \mathbf{P}_2; \mathbf{P}_{1'}, \mathbf{P}_{2'})|^2. \end{aligned}$$

¹We disregard the subtleties related to particle polarizations.

Similarly, the number of particles entering this volume due to the inverse process $1' + 2' \rightarrow 1 + 2$ is

$$\begin{aligned}\Delta N_1^+ &= d^4x \frac{d^3P_1}{P_1^0} \int \frac{d^3P_2}{P_2^0} \frac{d^3P_{1'}}{P_{1'}^0} \frac{d^3P_{2'}}{P_{2'}^0} f_{1'}(x, \mathbf{P}_{1'}) \\ &\quad \times f_{2'}(x, \mathbf{P}_{2'}) \cdot |\mathcal{M}(\mathbf{P}_{1'}, \mathbf{P}_{2'}; \mathbf{P}_1, \mathbf{P}_2)|^2\end{aligned}$$

Hence, in the locally Lorentz frame, the number of particles per unit phase space volume changes per unit time by

$$\begin{aligned}\frac{(\Delta N_1^+ - \Delta N_1^-)}{d^4x d^3P_1} &= \frac{1}{P_1^0} \int \frac{d^3P_2}{P_2^0} \frac{d^3P_{1'}}{P_{1'}^0} \frac{d^3P_{2'}}{P_{2'}^0} \\ &\quad \times \left[f_{1'}(x, \mathbf{P}_{1'}) f_{2'}(x, \mathbf{P}_{2'}) |\mathcal{M}(\mathbf{P}_{1'}, \mathbf{P}_{2'}; \mathbf{P}_1, \mathbf{P}_2)|^2 \right. \\ &\quad \left. - f_1(x, \mathbf{P}_1) f_2(x, \mathbf{P}_2) |\mathcal{M}(\mathbf{P}_1, \mathbf{P}_2; \mathbf{P}_{1'}, \mathbf{P}_{2'})|^2 \right]. \quad (8.32)\end{aligned}$$

It is this quantity that enters the right hand side of Eq. (8.18) due to particle collisions. The covariant generalization of (8.32) is obtained by replacing the Lorentz-invariant quantities, such as the volume element in the momentum space, by generally covariant ones, e.g.,

$$\frac{d^3P_i}{P^0} \rightarrow \frac{d^3P_i}{\sqrt{-g}P^0},$$

and generalizing the relation (8.25),

$$|\mathcal{M}|^2 = (P_1 + P_2)^2 \sqrt{-g} \delta(P_1 + P_2 - P_{1'} - P_{2'}) \frac{d\sigma}{d\Omega}, \quad (8.33)$$

where $d\sigma/d\Omega$ is still calculated in the locally Lorentz frame. The extra factor $\sqrt{-g}$ in the right hand side of (8.33) makes the modified δ -function generally invariant. Finally, the invariant generalization of the right hand side of the Boltzmann equation is

$$\begin{aligned}\mathcal{C}(x, \mathbf{P}_1) &= \int \frac{d^3P_2}{\sqrt{-g}P_2^0} \frac{d^3P_{1'}}{\sqrt{-g}P_{1'}^0} \frac{d^3P_{2'}}{\sqrt{-g}P_{2'}^0} \\ &\quad \times \left[f_{1'}(x, \mathbf{P}_{1'}) f_{2'}(x, \mathbf{P}_{2'}) |\mathcal{M}(\mathbf{P}_{1'}, \mathbf{P}_{2'}; \mathbf{P}_1, \mathbf{P}_2)|^2 \right. \\ &\quad \left. - f_1(x, \mathbf{P}_1) f_2(x, \mathbf{P}_2) |\mathcal{M}(\mathbf{P}_1, \mathbf{P}_2; \mathbf{P}_{1'}, \mathbf{P}_{2'})|^2 \right], \quad (8.34)\end{aligned}$$

where $|\mathcal{M}|^2$ is calculated according to (8.33). The factor $1/P_1^0$ is absent in (8.34) as compared to (8.32) because the left hand side of Eq. (8.17) contains $P^0(\partial f/\partial x^0)$ rather than $\partial f/\partial x^0$, see also (8.37). Hence, the account of two-particle collisions yields the Boltzmann equation of the following form,

$$P^\mu \frac{\partial f}{\partial x^\mu} + \Gamma_{j\rho}^\nu P^\rho P_\nu \frac{\partial f}{\partial P_j} = \mathcal{C}(x^\lambda, P_i). \quad (8.35)$$

The right hand side of this equation is called *collision integral*. In general, its complete expression is the sum of the right hand side of (8.34) over particle types $2, 1', 2'$. One should also account for decays and inverse decays, as well as multiparticle processes.

It is worth noting that when deriving the Boltzmann equation, we implicitly assumed that the particles are classical. Quantum effects modify the collision integral. This modification is relevant when the distinction between the Bose–Einstein and Fermi–Dirac statistics is important. As an example, fermions experience Pauli blocking: their distribution function cannot exceed unity (more precisely, $1/(2\pi)^3$ with our normalization). The opposite effect of Bose-enhancement exists for bosons. The modified expressions for the collision integral are

$$\begin{aligned} \mathcal{C}(x, \mathbf{P}_1) = & \int \frac{d^3 P_2}{\sqrt{-g} P_2^0} \frac{d^3 P_{1'}}{\sqrt{-g} P_{1'}^0} \frac{d^3 P_{2'}}{\sqrt{-g} P_{2'}^0} \\ & \times \left[|\mathcal{M}_{1'2' \rightarrow 12}|^2 f_{1'} f_{2'} \left(1 \pm \frac{(2\pi)^3 f_{1'}}{g_{1'}} \right) \left(1 \pm \frac{(2\pi)^3 f_{2'}}{g_{2'}} \right) \right. \\ & \left. - |\mathcal{M}_{12 \rightarrow 1'2'}|^2 f_1 f_2 \left(1 \pm \frac{(2\pi)^3 f_1}{g_1} \right) \left(1 \pm \frac{(2\pi)^3 f_2}{g_2} \right) \right], \end{aligned} \quad (8.36)$$

in self-evident notations. Here g_i denotes the number of spin states; upper and lower signs are relevant for bosons and fermions, respectively. This modification is important in some cases, we do not encounter them in what follows.

Equation (8.35) can be written as

$$\frac{\partial f}{\partial x^0} + \frac{P^i}{P^0} \frac{\partial f}{\partial x^i} + \frac{\Gamma_{j\rho}^\nu P^\rho P_\nu}{P^0} \frac{\partial f}{\partial P_j} = \frac{\mathcal{C}(x^\lambda, P_i)}{P^0}. \quad (8.37)$$

To simplify formulas, we introduce the notation

$$\frac{dP_\mu}{dx^0} \equiv \frac{\Gamma_{\mu\rho}^\nu P^\rho P_\nu}{P^0} \quad (8.38)$$

We emphasize that the momenta P_i entering (8.37) are independent variables, arguments of the distribution function. They are time-independent and do not obey any equations, so the left hand side of (8.38) is to be understood as the *notation for the right hand side*, rather than the time derivative. This notation is motivated by the fact that particle momenta obey similar relation, which in that case has the meaning of equation of motion. Indeed, the left hand side of Eq. (8.4) can be written as

$$m \frac{dP_{(i)\mu}}{ds_{(i)}} = m \frac{dx_{(i)}^0}{ds_{(i)}} \frac{dP_{(i)\mu}}{dx_{(i)}^0} = P_{(i)}^0 \frac{dP_{(i)\mu}}{dx_{(i)}^0}.$$

Hence, Eq. (8.4) can be written in the form analogous to (8.38). Note that the definition (8.38) implies that the “derivative” of the contravariant momentum $P^i = g^{\mu i} P_\mu$ is

$$\frac{dP^i}{dx^0} = - \frac{\Gamma_{\mu\nu}^i P^\mu P^\nu}{P^0}. \quad (8.39)$$

Problem 8.5. Show that the definition (8.38) yields (8.39).

Let us introduce yet another notation,

$$\frac{dx^\mu}{dx^0} = \frac{P^\mu}{P^0}. \quad (8.40)$$

It should also be understood as the *notation for the right hand side*; it is also motivated by the fact that similar relation holds for particles. As a result, Eq. (8.35) is written as

$$\frac{df}{dx^0} \equiv \frac{\partial f}{\partial x^0} + \frac{\partial f}{\partial x^i} \frac{dx^i}{dx^0} + \frac{\partial f}{\partial P_i} \frac{dP_i}{dx^0} = C[f], \quad (8.41)$$

where

$$C[f] = \frac{\mathcal{C}(x^\mu, P_i)}{P^0}.$$

This form of Eq. (8.35) is convenient due to the fact that it admits changes of variables: instead of coordinates and momenta x^i and P_i , one can use arbitrary functions of them as independent variables; these functions may even depend on time,

$$y^i = y^i(x^0; x^j, P_k), \quad \kappa_i = \kappa_i(x^0; x^j, P_k). \quad (8.42)$$

The left hand side of the Boltzmann equation (8.41) has the same form after this change of variables.

Problem 8.6. Show that the distribution function considered as function of variables (8.42) obeys the equation

$$\frac{\partial f(x^0, y^i, \kappa_j)}{\partial x^0} + \frac{\partial f(x^0, y^i, \kappa_j)}{\partial y^i} \frac{dy^i}{dx^0} + \frac{\partial f(x^0, y^i, \kappa_j)}{\partial \kappa_i} \frac{d\kappa_i}{dx^0} = C[f],$$

where the “derivative” of κ_i is, by definition,

$$\frac{d\kappa_i}{dx^0} = \frac{\partial \kappa_i}{\partial x^0} + \frac{\partial \kappa_i}{\partial x^j} \frac{dx^j}{dx^0} + \frac{\partial \kappa_i}{\partial P_j} \frac{dP_j}{dx^0}, \quad (8.43)$$

and similarly for y^i . Here the “derivatives” of x^i and P_i are defined by (8.38), (8.40).

Equation (8.37) (or, equivalently, (8.41)) is valid in arbitrary reference frame. At the same time, it is often useful to work in locally Lorentz frame. The transformation to the latter is made by using vierbein e_μ^α , see Appendix C. Regarding its lower index, vierbein is a covariant vector with respect to general coordinate transformations. As to the upper index, vierbein is a Lorentz vector. Vierbein vectors obey

$$\eta^{\alpha\beta} = g^{\mu\nu} e_\mu^\alpha e_\nu^\beta, \quad g_{\mu\nu} = \eta_{\alpha\beta} e_\mu^\alpha e_\nu^\beta. \quad (8.44)$$

The transformation of the vector P^μ to the locally Lorentz frame gives

$$p^\alpha = e_\mu^\alpha P^\mu, \quad (8.45)$$

where the vector p^α has the meaning of the physical momentum in the locally Lorentz frame. It follows from (8.8) and (8.44) that the physical momentum obeys the standard Minkowski relation $\eta_{\alpha\beta} p^\alpha p^\beta = m^2$.

Physical momenta get redshifted in the expanding Universe, so it is convenient to use, instead of \mathbf{p} , conformally stretched momentum

$$\kappa = a\mathbf{p}. \quad (8.46)$$

This would be conformal momentum in the homogeneous Universe. In the case of homogeneous collisionless gas, the distribution function, considered as a function of conformal momentum, does not depend on time, $f = f(\kappa)$ (see Section I.2.5, note the change of notation here).

As the distribution function f is conveniently considered as the function of variables (x, κ) , we write Eq. (8.37) in the form

$$\frac{\partial f}{\partial x^0} + \frac{d\kappa}{dx^0} \cdot \frac{\partial f}{\partial \kappa} + \frac{d\mathbf{x}}{dx^0} \cdot \frac{\partial f}{\partial \mathbf{x}} = C[f], \quad (8.47)$$

where $dx^i/dx^0 \equiv P^i/P^0$, while $d\kappa/dx^0$ is calculated according to (8.43). The study of processes in the Universe is performed, as a rule, in these terms.

8.2 General Equations for Scalar Perturbations

In what follows, it is convenient to use the variable κ introduced in (8.46), (8.45). In the gauge $g_{0i} = 0$, one of the vierbein vectors can be chosen to coincide with the time direction, while three others are spatial, $e_k^0 = e_0^k = 0$. Then the following relation holds,

$$\kappa^2 \equiv \kappa^i \kappa^i = -a^2 g_{ij} P^i P^j. \quad (8.48)$$

It shows, in particular, that the relation between κ and P_i depends on the form of metric. We consider here scalar perturbations (tensor perturbations are discussed in the end of Section 8.4.3). We have in the conformal Newtonian gauge

$$g_{00} = a^2 (1 + 2\Phi), \quad g_{ij} = -a^2 (1 + 2\Psi) \cdot \delta_{ij}. \quad (8.49)$$

Hence, the three vierbein vectors are orthogonal to each other, $e_j^i \propto \delta_j^i$, and the 3-vector κ^i has the same direction as P^i . We do not distinguish between 3-vectors with upper and lower indices in what follows, unless there is a risk of confusion; as an example, we set $\kappa_i = \kappa^i$. We obtain from (8.48) and (8.8) that²

$$P^\mu = \left((1 - \Phi) \frac{\sqrt{\kappa^2 + m^2 a^2}}{a^2}, (1 - \Psi) \frac{\kappa}{a^2} \right), \quad (8.50a)$$

$$P_\mu = ((1 + \Phi) \sqrt{\kappa^2 + m^2 a^2}, -(1 + \Psi) \kappa). \quad (8.50b)$$

²The results of Section I.2.5 imply that the components of 3-momenta of massive non-interacting particles $P^i = m dx^i/ds$ decay in time as a^{-2} , physical momenta decrease as $p^i = m(adx^i)/ds \propto a^{-1}$, while conformal momenta are constant in time. This, of course, is consistent with (8.50a).

The evolution of the distribution function $f_\lambda = f_\lambda(\kappa; \eta, \mathbf{x})$ is governed by the Boltzmann equation (8.47). Let us derive the explicit linearized form of its left hand side.

When calculating the collision integral (8.34) in the situation described in the beginning of Section 8.1, we can approximate the metric as static, homogeneous and conformally flat. We recall that the collision integral $C[f]$ entering Eq. (8.47) differs from the invariant quantity (8.34) by the factor $1/P^0$. The latter is calculated according to (8.50a). It is convenient to introduce unit 3-vector n^i , directed along 3-momentum κ^i , so that

$$\kappa \equiv \mathbf{n} \cdot \kappa, \quad \kappa^2 = \kappa^2, \quad x^{i\prime} \equiv \frac{dx^i}{d\eta} = \frac{P^i}{P^0} = \frac{\kappa \cdot n_i}{\sqrt{\kappa^2 + m^2 a^2}} \cdot (1 + \Phi - \Psi) \quad (8.51)$$

(according to our convention, we have $n_i = n^i$). In analogy to the definition $\delta = \delta\rho/\rho$, we define the perturbation $\delta f_\lambda(\kappa; \eta, \mathbf{x})$ of the distribution function by

$$f_\lambda = f_\lambda^{(0)} \cdot (1 + \delta f_\lambda), \quad (8.52)$$

where $f_\lambda^{(0)}$ is the homogeneous and isotropic unperturbed distribution function of the component λ . As an example, for particles in approximate thermal equilibrium with cosmic plasma, the function $f_\lambda^{(0)}(\kappa)$ is the Bose–Einstein or Fermi–Dirac distribution, depending on statistics of particles. We insert (8.52) into the Boltzmann equation (8.47) and obtain, to the linear order,

$$\kappa' \cdot \frac{\partial \log f_\lambda^{(0)}}{\partial \kappa} + (\log f_\lambda^{(0)})' \cdot \delta f_\lambda + \delta f'_\lambda + \frac{\kappa \cdot \mathbf{n}}{\sqrt{\kappa^2 + m^2 a^2}} \cdot \frac{\partial \delta f_\lambda}{\partial \mathbf{x}} = \frac{C^{(1)}[f]}{f_\lambda^{(0)}},$$

where prime, as always, denotes the derivative with respect to conformal time (recall that \mathbf{x}' and κ' are to be understood as explained in the end of Section 8.1). Here $C^{(1)}[f]$ is the linear part of the collision integral, and we made use of the fact that for homogeneous medium, “time derivative” of conformal momentum κ vanishes, so that κ' is the first order quantity, see also (8.55). In many cases of interest, the unperturbed distribution function is independent of time, $f^{(0)} = f^{(0)}(\kappa)$, and collision integral vanishes in the absence of perturbations, $C[f^{(0)}] = 0$. In these cases, the Boltzmann equation for small perturbations has simpler form,

$$\kappa' \cdot \frac{\partial \log f_\lambda^{(0)}}{\partial \kappa} + \delta f'_\lambda + \frac{\kappa \cdot \mathbf{n}}{\sqrt{\kappa^2 + m^2 a^2}} \cdot \frac{\partial \delta f_\lambda}{\partial \mathbf{x}} = \frac{C[f]}{f_\lambda^{(0)}}. \quad (8.53)$$

It is this form that we use in what follows, unless we state the opposite.

To find κ' , we make use of the relation (8.50a) and formula (8.39), in which x^0 is conformal time η . The Christoffel symbols in the conformal Newtonian gauge are (see Appendix B)

$$\begin{aligned} \Gamma_{00}^0 &= \frac{a'}{a} + \Phi', & \Gamma_{0i}^0 = \Gamma_{00}^i = \partial_i \Phi, & \Gamma_{0j}^i = \left(\frac{a'}{a} + \Psi' \right) \delta_{ij}, \\ \Gamma_{ij}^0 &= \left(\frac{a'}{a} - 2 \frac{a'}{a} \Phi + 2 \frac{a'}{a} \Psi + \Psi' \right) \delta_{ij}, & \Gamma_{jk}^i = (\delta_{ik} \partial_j \Psi + \delta_{ij} \partial_k \Psi - \delta_{jk} \partial_i \Psi). \end{aligned}$$

As a result, the “derivative” of P^i is

$$\frac{dP^i}{d\eta} = -2(aH + \Psi') \cdot P^i - \partial_i \Phi \cdot P^0 - \frac{1}{P^0} (2\delta_{ij} P^k P^j \partial_k \Psi - \delta_{jk} P^j P^k \cdot \partial_i \Psi). \quad (8.54)$$

We now express P^0 and P^i in terms of κ by making use of (8.50a) and insert them into (8.54). We find to the linear order

$$\kappa' = -\kappa\Psi' - \sqrt{\kappa^2 + m^2 a^2} \cdot \mathbf{n} \frac{\partial \Phi}{\partial \mathbf{x}}. \quad (8.55)$$

Finally, using (8.55) in (8.53), we obtain the linearized Boltzmann equation in the following form,

$$\begin{aligned} & - \left(\Psi' + \frac{\sqrt{\kappa^2 + m^2 a^2}}{\kappa} \cdot \mathbf{n} \frac{\partial \Phi}{\partial \mathbf{x}} \right) \cdot \frac{\partial \log f_\lambda^{(0)}}{\partial \log \kappa} + \delta f'_\lambda \\ & + \frac{\kappa}{\sqrt{\kappa^2 + m^2 a^2}} \cdot \mathbf{n} \frac{\partial \delta f_\lambda}{\partial \mathbf{x}} = \frac{C[f]}{f_\lambda^{(0)}}. \end{aligned} \quad (8.56)$$

To write the complete system of equations for scalar perturbations, we need the linearized expression for the energy-momentum tensor (8.10). To the linear order, we find from (8.49) and (8.50b) that

$$\frac{d^3 P_i}{\sqrt{-g}} = \frac{\kappa^2 d\kappa}{a^4} d\mathbf{n} \cdot (1 - \Phi),$$

where $d\mathbf{n}$ is an element of solid angle in the momentum space. Note that the factor $(1 - \Phi)$ cancels out in the expression for the invariant phase space volume in the integrand of (8.10), as it should. Let us extract the diagonal part in the spatial components of the energy-momentum tensor, denote it by p in analogy to ideal fluid, and introduce traceless anisotropic stress tensor Π_{ij} . Then

$$T_j^i = -\delta_j^i \cdot p - \Pi_j^i, \quad \Pi_i^i = 0.$$

The helicity decomposition of Π_{ij} is given in (B.13b).

We write to the linear order

$$\delta T_0^0 \equiv \delta\rho = \frac{1}{a^4} \int \kappa^2 d\kappa d\mathbf{n} \sqrt{\kappa^2 + m^2 a^2} \cdot f_\lambda^{(0)}(\kappa) \cdot \delta f_\lambda, \quad (8.57a)$$

$$\delta T_i^0 \equiv -(\rho + p) \cdot v_i = -\frac{1}{a^4} \int \kappa^3 d\kappa d\mathbf{n} n_i \cdot f_\lambda^{(0)}(\kappa) \cdot \delta f_\lambda, \quad (8.57b)$$

$$\delta p = \frac{1}{3a^4} \int \frac{\kappa^4 d\kappa d\mathbf{n}}{\sqrt{\kappa^2 + m^2 a^2}} \cdot f_\lambda^{(0)}(\kappa) \cdot \delta f_\lambda, \quad (8.57c)$$

$$\Pi_j^i = \frac{1}{a^4} \int \frac{\kappa^4 d\kappa d\mathbf{n} (n^i n_j - \frac{1}{3} \delta_j^i)}{\sqrt{\kappa^2 + m^2 a^2}} \cdot f_\lambda^{(0)}(\kappa) \cdot \delta f_\lambda. \quad (8.57d)$$

Non-vanishing anisotropic stress is a new effect as compared to ideal fluid. It contributes to both the linearized covariant conservation equations and Einstein equations, see Appendix B. In particular, using the expressions (2.70)–(2.72) for the linearized Einstein tensor, we obtain the following equations in the scalar sector (see also Appendix B):

$$-\Delta\Psi + 3\frac{a'}{a}\Psi' - 3\frac{a'^2}{a^2}\Phi = 4\pi Ga^2 \cdot \delta\rho_{tot}, \quad (8.58a)$$

$$-\Psi' + \frac{a'}{a}\Phi = -4\pi Ga^2 \cdot [(\rho + p)v]_{tot}, \quad (8.58b)$$

$$\Psi'' - \frac{1}{3}\Delta(\Phi + \Psi) + \frac{a'}{a}(2\Psi' - \Phi') - 2\frac{a''}{a}\Phi + \frac{a'^2}{a^2}\Phi = -4\pi Ga^2 \cdot \delta p_{tot}, \quad (8.58c)$$

$$\Delta(\Phi + \Psi) = -12\pi Ga^2 \cdot [(\rho + p)\pi]_{tot}, \quad (8.58d)$$

Here we introduced the anisotropic stress potential π via

$$\Pi_{ij} = (p + \rho) \left(-\frac{1}{2}\delta_{ij} + \frac{3}{2}\frac{\partial_i\partial_j}{\Delta} \right) \pi. \quad (8.59)$$

Note that the expression in parenthesis here is the only helicity-0 traceless combination. Equations (8.58a), (8.58b) generalize Eqs. (2.74) and (2.75), while Eqs. (8.58c) and (8.58d) replace single equation (2.76) in the situation where the anisotropic stress does not vanish. The anisotropic stress ruins the relation (2.73) between the gravitational potentials Φ and Ψ , thus making the study of perturbations more complicated. When writing Eq. (8.58d) we introduced the “total” anisotropic stress potential by analogy to the “total” velocity potential v ,

$$[(\rho + p)\pi]_{tot} = \sum_{\lambda} [(\rho_{\lambda} + p_{\lambda})\pi_{\lambda}],$$

where the sum runs over all components of the cosmic medium. The covariant conservation equations (5.4) and (5.5) are generalized as follows (see Appendix B),

$$\delta\rho'_{\lambda} + 3\frac{a'}{a}(\delta\rho_{\lambda} + \delta p_{\lambda}) - (\rho_{\lambda} + p_{\lambda})(k^2 v_{\lambda} - 3\Psi') = 0, \quad (8.60)$$

$$[(\rho_{\lambda} + p_{\lambda})v_{\lambda}]' + 4\frac{a'}{a}(\rho_{\lambda} + p_{\lambda})v_{\lambda} + \delta p_{\lambda} + (\rho_{\lambda} + p_{\lambda})(\pi_{\lambda} + \Phi) = 0. \quad (8.61)$$

To end this Section, let us make a general comment. The ideal fluid approximation is not valid, generally speaking, for all components of the cosmic medium which are out of thermal equilibrium. Nevertheless, it can sometimes be used, formally, even in that situation, provided that the integrals (8.57) coincide with the perturbed energy-momentum tensor of ideal fluid. This situation is characteristic, e.g., of cold dark matter: it is clear from (8.57) that all integrals but (8.57a) are suppressed by the mass of dark matter particle, so dark matter after its decoupling is effectively ideal fluid with the equation of state $p = 0$. In fact, this is a fairly

general situation: for perturbations of most of the components, there exist wavelength intervals in which the formal use of the ideal fluid approximation gives the correct result.

8.3 Warm Dark Matter

8.3.1 Suppressed growth of perturbations

As a simple example, let us study perturbations in collisionless gas whose particles participate in gravitational interactions only. These perturbations get damped by the *free streaming* effect which is very similar to the Landau damping in electromagnetic plasma.³ The physics of this phenomenon is simple: free motion of particles in space washes out inhomogeneity in spatial distribution of these particles. The flux of particles from an overdense region is higher than that from an underdense one, so the free motion smoothes out the density contrast. This is a kinematic effect whose spatial scale is determined by the distance that particles travel in characteristic time, the Hubble time in the cosmological context. Perturbations of wavelengths shorter than this length are suppressed. In the first place, this happens for the density perturbations in the collisionless component itself, but gravitational potentials get suppressed too.

Clearly, the free streaming effect is most pronounced for light particles that decouple from cosmic plasma being relativistic. Among the Standard Matter particles these are neutrinos, and also photons after recombination. Free streaming effects for neutrinos is studied in Section 8.4, and here we concentrate on a simpler case of warm dark matter.

Recall (see Section I.9.1) that dark matter may in principle be *cold*, *hot* and *warm*. Cold dark matter particles decouple when they are non-relativistic, while hot and warm dark matter particles decouple being relativistic. The distinction between hot and warm dark matter is in the properties of the particles at radiation-matter equality: hot dark matter particles are still relativistic at that epoch, while warm dark matter particles are not. As we find in the Section, free streaming effect is negligible for cold dark matter, it rules out the possibility that most of non-baryonic dark matter is hot and places strong bounds on warm dark matter properties.

For concreteness, let us consider matter dominated epoch, when the cosmological expansion is mostly governed by dark matter. Unlike in the previous Chapters, we do not neglect velocities of dark matter particles. Let $f(x^0, \mathbf{x}, \mathbf{u})$ be the distribution function of these particles; it is somewhat more convenient here to use the distribution in velocity rather than in momentum. Our first purpose is to find the length

³Landau damping is a collective effect in electromagnetic plasma occurring when perturbations in the plasma density are “out of phase” with perturbations of electromagnetic field. In the cosmological context, the matter perturbations are out of phase with perturbations of the gravitational field.

scales of perturbations that grow in time due to the gravitational instability and the scales for which the free streaming effect suppresses the growth. To this end, we use the Newtonian approach of Section 1.

Recall that in the ideal fluid approximation, the relevant length scale is the Jeans length λ_J , see Chapter 1: perturbations of larger wavelengths grow, while shorter ones oscillate. The Jeans length in the expanding Universe filled with single component fluid is of the order of the sound horizon size, $u_s/\sqrt{4\pi G\rho} \sim u_s(t)/H(t)$, see Section 1.2. In analogy with the Jeans momentum (1.11) and wavelength (1.12), it is natural to introduce the free streaming conformal momentum k_{fs} and physical wavelength λ_{fs} by replacing the sound velocity with the averaged in some way velocity of dark matter particles $u(t)$,

$$\begin{aligned} k_{fs}(t) &= \left(\frac{4\pi G\rho(t) a^2(t)}{u^2(t)} \right)^{1/2} = \sqrt{\frac{3}{2}} \frac{H(t) a(t)}{u(t)}, \\ \lambda_{fs}(t) &= 2\pi \frac{a(t)}{k_{fs}(t)} = 2\pi \sqrt{\frac{2}{3}} \frac{u(t)}{H(t)}, \end{aligned} \quad (8.62)$$

where ρ is the energy density of the medium. The second equalities here reflect our assumption that the component we consider dominates the cosmological expansion. We are going to see that k_{fs} and λ_{fs} indeed are the scales at which free streaming starts to be important. Note that $\lambda_{fs} \sim uH^{-1}$ is of the order of the length the particles travel in the Hubble time, as we have already pointed out. We also note that for relativistic particles, λ_{fs} is of the order of the horizon size, so all subhorizon modes are suppressed. Hence, the non-trivial situation occurs for non-relativistic particles, and we consider this case in what follows. We emphasize that $\lambda_{fs}(t)$ in Eq. (8.62) is the current physical wavelength at time t ; the corresponding present size is $\lambda_0 = 2\pi a_0/k_{fs}(t) = (1+z)\lambda_{fs}(t)$, where $z = z(t)$.

To obtain the result (8.62), let us proceed along the lines of Section 1.1 and consider the Minkowski background. Since we study single-component medium, the gravitational potential Φ in the Newtonian approximation obeys

$$\Delta\Phi = 4\pi G\rho, \quad (8.63)$$

while the local energy density of the non-relativistic particles of mass m is given by the integral of the distribution function over velocity,

$$\rho(x) = m \int d^3 u f(x, \mathbf{u}). \quad (8.64)$$

The distribution function obeys the Vlasov equation (8.17) with obvious modification due to the fact that we consider distribution in velocities rather than in momenta. For the non-relativistic particles in the Minkowski background we have

$$x^0 = t, \quad \frac{P^i}{P^0} = u^i, \quad \frac{\Gamma_{j\rho}^\nu P^\rho P_\nu}{P^0} = -m \frac{\partial\Phi}{\partial x^j}.$$

The latter relation can be obtained either by expanding the metric near the Minkowski background or by noticing that the left hand side is the force entering the equation of motion of a particle (see (8.4)). Hence, the Vlasov equation reads

$$\frac{\partial f}{\partial t} + \mathbf{u} \frac{\partial f}{\partial \mathbf{x}} - \frac{\partial \Phi}{\partial \mathbf{x}} \frac{\partial f}{\partial \mathbf{u}} = 0, \quad (8.65)$$

where the gravitational potential in the mean field approximation is determined by Eq. (8.63).

Following the lines of Section 1.1, let us find the evolution of perturbations about static, homogeneous background distribution with

$$\rho = m \int d^3 u f^{(0)}(\mathbf{u}) = \text{const}; \quad (8.66)$$

the background gravitational potential is assumed to vanish. Like in Chapter 1, such a background is not a solution to Eqs. (8.63)–(8.65), but this is unimportant for calculating the free streaming wavelength λ_{fs} ; we confirm this later on.

Let us define the perturbation of the distribution function⁴

$$\delta f(x, \mathbf{u}) = f(x, \mathbf{u}) - f^{(0)}(\mathbf{u}), \quad \delta \Phi(x) = \Phi.$$

Then the linearized Vlasov equations are

$$\begin{aligned} \Delta \Phi &= 4\pi Gm \int d^3 u \delta f, \\ \frac{\partial \delta f}{\partial t} + \mathbf{u} \frac{\partial \delta f}{\partial \mathbf{x}} &= \frac{\partial \Phi}{\partial \mathbf{x}} \cdot \frac{\partial f^{(0)}}{\partial \mathbf{u}}. \end{aligned}$$

Note that the second of these equations is the non-relativistic limit of Eq. (8.56) with $a = 1$. Let us perform the Fourier transformation in variables t and \mathbf{x} ,

$$\Phi \rightarrow \Phi(\omega, \mathbf{q}) \cdot e^{-i\omega t + i\mathbf{q}\mathbf{x}}, \quad \delta f \rightarrow \delta f(\omega, \mathbf{q}, \mathbf{u}) \cdot e^{-i\omega t + i\mathbf{q}\mathbf{x}}.$$

The meaning of \mathbf{q} and ω is that \mathbf{q} is the momentum of perturbation (rather than momentum of a dark matter particle) while ω is its frequency. In terms of the Fourier-harmonics, the Vlasov equations read

$$-q^2 \Phi = 4\pi Gm \int d^3 u \delta f, \quad (8.67a)$$

$$(-\omega + \mathbf{q}\mathbf{u}) \delta f = \mathbf{q} \frac{\partial f^{(0)}}{\partial \mathbf{u}} \Phi. \quad (8.67b)$$

We now solve Eq. (8.67b) for δf and insert the solution into Eq. (8.67a). This gives the dispersion equation

$$q^2 = 4\pi Gm \int d^3 u \frac{1}{\omega - \mathbf{q}\mathbf{u}} \left(\mathbf{q} \frac{\partial f^{(0)}}{\partial \mathbf{u}} \right). \quad (8.68)$$

⁴The perturbation δf is related to the perturbation δf introduced in (8.52) via $\delta f = \delta f \cdot f^{(0)}$.

Since the integral over velocity of any even function with the weight $\partial f^{(0)}/\partial \mathbf{u}$ vanishes, Eq. (8.68) takes the final form,

$$1 = \frac{4\pi Gm}{q^2} \int d^3 u \frac{(\mathbf{q}\mathbf{u})}{\omega^2 - (\mathbf{q}\mathbf{u})^2} \left(\mathbf{q} \frac{\partial f^{(0)}}{\partial \mathbf{u}} \right). \quad (8.69)$$

Let us first consider large wavelength modes. Equation (8.69) gives for $q \rightarrow 0$

$$\omega^2 = \frac{4\pi Gm}{q^2} \int d^3 u (\mathbf{q}\mathbf{u}) \left(\mathbf{q} \frac{\partial f^{(0)}}{\partial \mathbf{u}} \right) = -4\pi G\rho,$$

where we integrated by parts to obtain the second equality. We see that the frequency is imaginary, so large wavelength perturbations grow exponentially. This is precisely the Jeans instability in collisionless gas. In fact, the increment of this instability is the same as for cold dark matter, see Section 1.1.

The equation for the critical momentum $q_{fs} = |\mathbf{q}_{fs}|$, above which the growth is absent, is found from Eq. (8.69) by setting $\omega = 0$,

$$q_{fs}^2 = -4\pi Gm \int d^3 u \frac{1}{\mathbf{q}_{fs}\mathbf{u}} \left(\mathbf{q}_{fs} \frac{\partial f^{(0)}}{\partial \mathbf{u}} \right).$$

By the isotropy of the background, the unperturbed distribution function depends only on the absolute value of the velocity $u \equiv |\mathbf{u}|$, and the latter equation can be written as follows,

$$q_{fs}^2 = -4\pi Gm \int d^3 u \frac{1}{u} \frac{\partial f^{(0)}}{\partial u} = -4\pi Gm \int_0^\infty 4\pi u du \frac{\partial f^{(0)}}{\partial u}.$$

Upon integrating by parts we obtain

$$q_{fs}^2 = 4\pi Gm \int_0^\infty 4\pi du f^{(0)} = 4\pi Gm \int d^3 u \frac{f^{(0)}}{u^2},$$

and finally [36]

$$q_{fs} = \sqrt{4\pi G\rho \cdot \left\langle \frac{1}{u^2} \right\rangle}, \quad (8.70)$$

where

$$\left\langle \frac{1}{u^2} \right\rangle = \frac{\int d^3 u f^{(0)}/u^2}{\int d^3 u f^{(0)}}.$$

This defines the average velocity entering (8.62). For physically interesting distributions, one has $\langle 1/u^2 \rangle \simeq 1/\langle u^2 \rangle$; we often do not specify the precise meaning of the average velocity in what follows. We emphasize that the above derivation has been performed in terms of physical momenta related to conformal momenta \mathbf{k} in the usual way, $\mathbf{q} = \mathbf{ka}$. Hence the factor a in the first of the relations (8.62).

The momentum (8.70) is critical also in the expanding Universe dominated by warm dark matter. The growth of perturbations occurs only for modes with $q < q_{fs}$, but this growth is power law, $\delta \propto a(t) \propto t^{2/3}$, rather than exponential. We give the corresponding analysis later on.

Let us discuss two cases within our single-component scenario. The first one is cold dark matter, CDM. By definition, CDM particles are non-relativistic at decoupling. Hence, their distribution function after decoupling has the Maxwell–Boltzmann form with effective temperature T_{eff} . In this case we have

$$\text{CDM: } q_{fs} = \left(4\pi G \rho \frac{m}{T_{eff}} \right)^{1/2}, \quad (8.71)$$

and the effective temperature decreases in time as (see Section I.2.5)

$$\text{CDM: } T_{eff} \propto a^{-2}.$$

Since $\rho \propto a^{-3}$, the critical conformal momentum grows in time,

$$k_{fs} = aq_{fs} \propto a^{1/2} \propto t^{1/3}. \quad (8.72)$$

Hence, as time increases, shorter and shorter waves become unstable.

As we know, rapid growth of perturbations begins at radiation-matter equality. So, the critical momentum and wavelength at this epoch are of particular interest. We find for CDM from (8.71) that

$$\lambda_{fs}(t_{eq}) \simeq 2\pi H_{eq}^{-1} \left(\frac{T_{eff}(t_{eq})}{m} \right)^{1/2}.$$

We roughly estimate the effective temperature as follows,

$$T_{eff}(t_{eq}) \simeq T_d \left(\frac{a_d}{a_{eq}} \right)^2 \simeq \frac{T_{eq}^2}{T_d},$$

where the subscript d refers to CDM decoupling (time at which CDM particles get out of kinetic equilibrium with the cosmic plasma). The present wavelength is

$$\lambda_0 = (1 + z_{eq})\lambda_{fs}(t_{eq}) \simeq 2\pi(1 + z_{eq})H_{eq}^{-1} \left(\frac{T_{eq}^2}{mT_d} \right)^{1/2} \sim \left(\frac{1 \text{ GeV}^2}{mT_d} \right)^{1/2} \text{ pc}, \quad (8.73)$$

where we use the values $z_{eq} = 3.2 \cdot 10^3$, $T_{eq} = (1 + z_{eq})T_0 \simeq 0.76 \text{ eV}$ and $2\pi(1 + z_{eq})H_{eq}^{-1} \simeq 640 \text{ Mpc}$, see Section I.4.4. We see that free streaming is irrelevant for most cosmologically interesting scales. As an example, the values typical for weakly interacting massive particles (WIMPs, see Section I.9.3) are $m \sim 100 \text{ GeV}$, $T_d \sim 10 \text{ MeV}$ (see below), so the scale (8.73) is of order 1 pc, which is very small by cosmological standards. We note though, that there is some discussion in literature on the possible existence of small dark matter clumps and ways to detect them. The mass spectrum of these clumps must be cut off from below due to free streaming.

Let us mention another effect inherent in many CDM models. It also leads to the suppression of small scale perturbations. The suppression is weaker, while the length scales are often larger as compared to free streaming.

Let us consider for definiteness stable neutralino of supersymmetric extensions of the Standard Model, see Section I.9.6.1. After freeze-out of the neutralino abundance, these particles remain in kinetic equilibrium with cosmic plasma for a long time: the collision rate of neutralino with other particles is high enough to change the neutralino momentum in the Hubble time. The rate at which neutralino momentum changes in time is (cf. Section I.6.3)

$$\Gamma_p = n \cdot \sigma_N \cdot \frac{T}{M_N}, \quad (8.74)$$

where M_N is the neutralino mass, $n \propto g_* T^3$ is the number density of other particles in the plasma, σ_N is the elastic scattering cross section. The cross section decreases as $\sigma_N \propto T^2$ at low temperatures. Let us define the mass parameter M by $\sigma_N = T^2/M^4$, then the temperature of exit from kinetic equilibrium, when $\Gamma_p \sim H$, is estimated by

$$T_d \sim \left(\frac{M^4 M_N}{\sqrt{g_*} M_{Pl}} \right)^{1/4} \simeq 5 \text{ MeV} \cdot \left(\frac{M_N}{100 \text{ GeV}} \right)^{1/4} \cdot \frac{M}{100 \text{ GeV}},$$

where we assume that $T_d \ll \Lambda_{QCD}$ in the estimate for the number of degrees of freedom. We see that, roughly speaking, $T_d \sim 10 \text{ MeV}$, i.e., decoupling occurs at radiation domination, and most likely before neutrino decoupling.

Before decoupling, neutralino and usual particles make single fluid with the relativistic equation of state. Density perturbations do not grow in such a fluid, while perturbations of decoupled dark matter would grow logarithmically. This is the difference between neutralino and “genuine” CDM that decouples very early (it is this genuine CDM that we have discussed in Section 6.2 without explicitly mentioning that). Thus, neutralino perturbations of momenta $k > k_d \simeq \eta_d^{-1}$ get somewhat suppressed. Making use of (2.13b) and recalling entropy conservation, $g_* a^3 T^3 = \text{const}$, we obtain the present wavelength,

$$\lambda_{d,0} = 2\pi \frac{a_0}{k_d} \simeq 2\pi \left(\frac{g_{*,0}}{g_*(T_d)} \right)^{1/6} \frac{1}{H_0 \sqrt{\Omega_{rad}}} \frac{T_0}{T_d} \simeq 60 \text{ pc} \cdot \frac{10 \text{ MeV}}{T_d}. \quad (8.75)$$

This may well be larger than the free streaming scale estimated by (8.73).

The suppression is, however, fairly mild. The logarithmic growth of neutralino perturbations with $k > k_d$ starts at the time η_d , whereas the growth of perturbations of genuine dark matter begins at the horizon entry. So, the logarithmic growth factor for neutralino is smaller by radiation-matter equality, but not very much smaller. In any case, the neutralino perturbations of these scales become non-linear rather early at matter domination. This makes the detection of the effect very difficult, if not impossible.

Problem 8.7. Find typical WIMP velocities for $m \sim 100 \text{ GeV}$, $T_d \sim 10 \text{ MeV}$ at $z = 100$. What distance do WIMPs travel in the Hubble time?

Problem 8.8. Find the mass of dark matter clump whose size was of the order of the free streaming scale at $t = t_{eq}$. Give numerical estimate for $m \sim 100 \text{ GeV}$, $T_d \sim 10 \text{ MeV}$. Find the same for the scale (8.75).

Problem 8.9. Find the free streaming scale λ_{fs} for baryons right after recombination in the (unrealistic) Universe without dark matter. Find the mass of baryons in a ball of size λ_{fs} . Take the baryon-to-photon ratio equal to its value in our Universe. Modulo factor of order 1, the result is the same as the Jeans length for the

baryonic Universe obtained in Section 1. Note that the physics behind the scales λ_{fs} and λ_J is entirely different. The former case refers to collisionless gas, while the latter is determined by pressure in the baryonic medium. Is the equality of the two scales a pure coincidence?

Let us now turn to the second case of warm dark matter, WDM. For definiteness, let WDM particles be fermions and have thermal distribution at decoupling. Since WDM particles are relativistic at decoupling, their distribution in momenta has the relativistic Fermi–Dirac form with the effective temperature decreasing as a^{-1} , see Section I.2.5. At late times, when WDM particles are non-relativistic, their distribution in velocities is

$$f^{(0)}(t) \propto \frac{1}{e^{um/T_{\text{eff}}(t)} + 1}. \quad (8.76)$$

The average of the inverse velocity squared that enters (8.70) is

$$\left\langle \frac{1}{u^2} \right\rangle = \frac{2 \log 2}{3\zeta(3)} \left(\frac{m}{T_{\text{eff}}} \right)^2 = 0.38 \left(\frac{m}{T_{\text{eff}}} \right)^2 \quad (8.77)$$

Problem 8.10. Derive the formula (8.77). Find its analog for bosons.

Let us use the entropy conservation in the plasma of the usual particles, $g_*(T) T^3(t) a^3(t) = \text{const}$, and write (cf. Section I.7.2)

$$T_{\text{eff}}(t) = \left(\frac{g_*(T)}{g_*(T_d)} \right)^{1/3} T(t), \quad (8.78)$$

where T_d is the decoupling temperature of WDM particles. Recall (see Section I.4.4) that at temperature below a few hundred keV (below electron mass), the effective number of degrees of freedom is $g_*(T < m_e) = 43/11 = 3.9$. Making use of the above expressions, we find from (8.70) the critical WDM wavelength

$$\lambda_{fs}(t) = 0.45 \cdot 2\pi H^{-1}(t) \frac{T(t)}{m} \left(\frac{100}{g_*(T_d)} \right)^{1/3}.$$

Like in the CDM case, the critical conformal momentum grows as $t^{1/3}$ at matter domination, cf. (8.72).

We again consider radiation-matter equality epoch. The present value of the corresponding wavelength is

$$\begin{aligned} \lambda_0 &= (1 + z_{eq}) \lambda_{fs}(t_{eq}) = 0.45 \cdot 2\pi (1 + z_{eq}) H^{-1}(t_{eq}) \frac{T_{eq}}{m} \left(\frac{100}{g_*(T_d)} \right)^{1/3} \\ &\simeq 220 \text{ kpc} \cdot \left(\frac{100}{g_*(T_d)} \right)^{1/3} \left(\frac{1 \text{ keV}}{m} \right). \end{aligned} \quad (8.79)$$

This estimate gives the maximum length of perturbations which are suppressed at matter domination.

We have seen in Section 7 that the growth of perturbations at radiation domination is important too. So, the analysis has to be extended by studying WDM perturbations at that epoch. We have to distinguish perturbations that enter the horizon at temperatures $T_{\text{eff}} > m$ and $T_{\text{eff}} < m$. Their momenta obey $k > k_{nr}$ and $k < k_{nr}$, respectively, where $k_{nr} \sim \eta_{nr}^{-1}$, and η_{nr} is conformal time at which WDM particles become non-relativistic, $T_{\text{eff}}(\eta_{nr}) \simeq m$. We estimate the latter by noticing that the effective number of degrees of freedom at $T \sim m$ for interesting values of m is the same as at present, and obtain from (2.13b) that

$$a_0 k_{nr}^{-1} \simeq a_0 \eta_{nr} \simeq \left(\frac{g_{*,0}}{g_*(T_d)} \right)^{1/3} \frac{1}{H_0 \sqrt{\Omega_{\text{rad}}}} \frac{T_0}{m} \simeq 100 \text{ kpc} \cdot \frac{1 \text{ keV}}{m}.$$

Perturbations with $k > k_{nr}$ enter the horizon when WDM particles are still relativistic. We leave the analysis of this situation to Section 8.4.1, and here we state that these perturbations decrease as $\delta_{DM} \propto (k\eta)^{-1}$ until $\eta \simeq \eta_{nr}$. Hence, these perturbations get suppressed even before WDM particles become non-relativistic.

The suppression at $\eta > \eta_{nr}$, which is due to free streaming of non-relativistic particles, is important for subhorizon modes. In this regime, gravitational potentials are small at radiation domination, so Eq. (8.56) with vanishing collision integral reduces to the simple equation for the WDM distribution function,

$$\delta f' + \frac{i\kappa\mathbf{k}}{ma(\eta)} \delta f = 0,$$

where we have performed the Fourier transformation in coordinate \mathbf{x} . We recall that \mathbf{k} is the conformal momentum characterizing the *perturbation mode*, whereas κ is the *conformal momentum of dark matter particle*. The solution to the latter equation is

$$\delta f(\eta) = \exp \left(-\frac{i\kappa\mathbf{k}}{m} \int_{\eta_{\times}}^{\eta} \frac{d\tilde{\eta}}{a(\tilde{\eta})} \right) \cdot \delta f(\eta_{\times}), \quad (8.80)$$

where η_{\times} denotes the horizon entry time for modes with $k < k_{nr}$, while for $k > k_{nr}$ one should set $\eta_{\times} \sim \eta_{nr}$. It follows from (8.80) that the perturbation δf rapidly oscillates at large k as function of κ . In that case, the expressions (8.57) for the components of the energy-momentum tensor involve the integrals of rapidly oscillating functions, so they are suppressed. Hence, WDM perturbations are small by radiation-matter equality, provided that

$$\frac{\kappa k}{m} \int_{\eta_{\times}}^{\eta_{eq}} \frac{d\eta}{a(\eta)} \gg 1, \quad (8.81)$$

where κ is the typical conformal momentum of WDM particle. For thermal spectrum (8.76), the typical momentum is estimated from the relation $\kappa/a \sim T_{\text{eff}}$ (recall that aT_{eff} is independent of time), so the condition (8.81) takes the form,

$$\frac{T_{\text{eff}}ak}{m} \int_{\eta_{\times}}^{\eta_{eq}} \frac{d\eta}{a(\eta)} = \frac{T_{\text{eff}}k\eta}{m} \log \frac{\eta_{eq}}{\eta_{\times}} = k\eta_{nr} \log \frac{k}{k^{(eq)}} \gg 1,$$

where $k^{(eq)}$ is the momentum of the mode that enters the horizon at $\eta = \eta_{eq}$, see Section 2.4. We see that free streaming at radiation domination determines the maximum wavelength of suppressed perturbations: the minimum momentum is obtained from the equation

$$k_{min}^{-1} = k_{nr}^{-1} \log \frac{k_{min}}{k^{(eq)}}.$$

For $m = 1 \text{ keV}$, the value of $a_0 k_{min}^{-1}$ is of order 500 kpc, and the present wavelength is $\lambda_{max} = 2\pi a_0 k_{min}^{-1} \sim 3 \text{ Mpc}$.

Problem 8.11. Estimate the maximum wavelength for which the free streaming suppression at radiation domination is important for CDM. Show that for CDM with $m \sim 100 \text{ GeV}$, $T_d \sim 10 \text{ MeV}$, this effect is relevant for shorter waves as compared to (8.75).

It follows from the above analysis that WDM particles cannot have mass much less than 1 keV. As an example, if the mass of WDM particle were a few $\cdot 100 \text{ eV}$, perturbations of the present length scale of about 1 Mpc would be suppressed even at matter domination, let alone radiation domination. Perturbations of this scale correspond to galaxies, and their suppression would contradict observations. Note that this is even more so for hot dark matter; this means that hot dark matter particles cannot be the dominant component of dark matter in our Universe.

On the other hand, WDM whose particles have masses exceeding ten keV does not experience free streaming at present scales of 100 kpc and above. These are the scales which correspond to dwarf galaxies and heavier structures. Hence, insofar as galaxy formation is concerned, dark matter of this sort behaves in the same way as CDM.

WDM consisting of particles with mass $m \sim 1 - 10 \text{ keV}$ is of particular interest. The growth of perturbations at matter domination occurs in almost the same way as for CDM. Yet the perturbations are suppressed at relatively short, but still cosmologically interesting scales. The suppression of the power spectrum $P(k)$ in WDM model is shown in Fig. 8.1. For definiteness, the initial distribution function is chosen in the form (8.76), normalized to the dark matter mass density $\rho_{DM} = \Omega_{DM} \rho_c$, i.e.,

$$f^{(0)}(u) = \frac{\rho_{DM}}{6\pi\zeta(3)mT_{0,eff}^3} \frac{1}{e^{mu/T_{0,eff}} + 1}; \quad (8.82)$$

the effective number of degrees of freedom is assumed to be $g_*(T_d) = g_{*,MSSM} = 228.75$ (the number of degrees of freedom in the Minimal Supersymmetric Standard Model). We see that for $m = 1 - 10 \text{ keV}$, perturbations corresponding to structures of masses $M \lesssim 10^9 - 10^7 M_\odot$ are indeed suppressed considerably, while the spectrum is almost the same as in CDM model for larger masses. Accordingly, WDM model predicts smaller abundance of structures of relatively small mass, see Fig. 8.2.

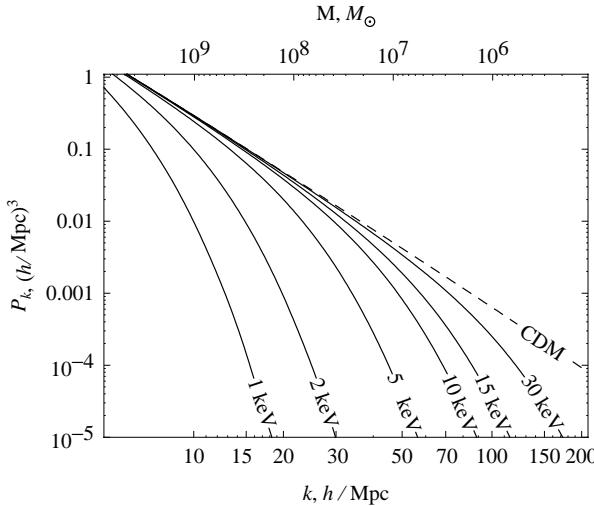


Fig. 8.1 The present power spectrum of the linear theory in CDM model (dashed line) and WDM models with the distribution function (8.82) for realistic cosmological parameters [37]; numbers at curves show the WDM particle mass.

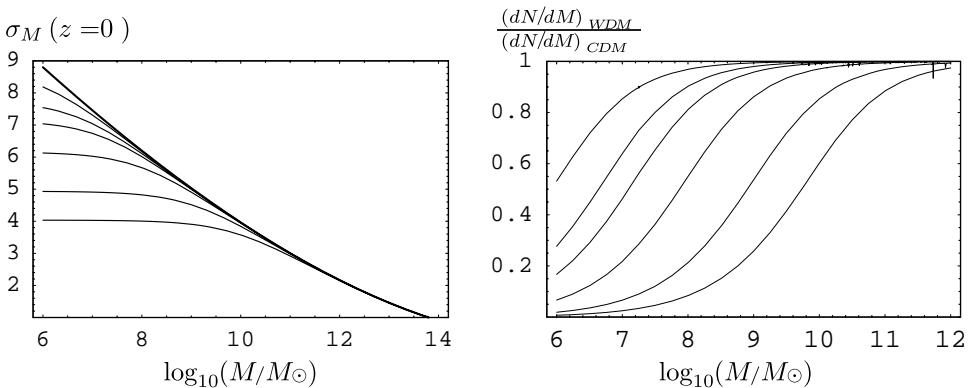


Fig. 8.2 Left: the present variance of the linear density contrast δ_M (see the definition in Section 7.2) in CDM (thick line) and WDM models. Right: the WDM/CDM ratio of the differential mass spectra at the present epoch. The curves in both plots correspond to models with power spectra shown in Fig. 8.1; the smaller the mass, the lower the line.

There are several observational results that may point towards the advantage of WDM over CDM. The first one has to do with dwarf galaxies. These are objects of masses $(10^7 - 10^9) M_\odot$. The observed number of dwarf satellites of our Galaxy is about an order of magnitude smaller than that predicted by the CDM model [38, 39]. A similar situation with dwarf galaxies takes place in the region of size 8 Mpc around our Galaxy [40, 41]. In WDM model, perturbations of sizes corresponding to dwarf galaxies are suppressed, so this model may, possibly, better describe the data. The second point is that numerical simulations with CDM show strong increase (cusp) of the mass density near the centers of galaxies, $\rho \propto r^{-1}$ to $r^{-1.5}$ as $r \rightarrow 0$. This is not confirmed by observations. There is a hope

that this difficulty may be overcome in WDM model, again because of the suppression of small scale perturbations. Finally, CDM model has a potential problem with the formation of spiral galaxies: fast collapse leads to the considerable loss of angular momentum, so the disks are too small in the end. The collapse is slower in WDM model, and larger disks may be formed. We note, however, that quantitative analysis of these issues in WDM model is a difficult and largely unsolved problem.

On the other hand, WDM model has its own problems. One is that the studies of hydrogen clouds at $z \sim 3$, based on the Lyman- α forest, show that the power spectrum is not strongly suppressed at relatively small scales [42]. If the Lyman- α analysis gives the $z \sim 3$ spectrum consistent with CDM at the smallest scales, and at the same time the deficit of dwarf galaxies in the present Universe is confirmed, one will have to conclude that the power spectrum evolved non-trivially between $z \sim 3$ and $z \sim 0$. The latter property is present in neither CDM nor WDM model.

Another hard problem of WDM model is that the suppression of the small scale perturbations inhibits the formation of the first stars. However, the formation of the first stars yields reionization of hydrogen, thus affecting CMB properties, see Sections 9.2.3 and 10.3.2. CMB data are fully consistent with CDM model, and disfavor WDM. A possible solution of this problem is the production of energetic photons in decays of WDM particles. Interestingly, this mechanism can work for light sterile neutrino of mass $m_s \sim 1 - 10$ keV as WDM particle. The decay of this neutrino, $\nu_s \rightarrow \nu_{e,\mu,\tau} + \gamma$ may occur due to mixing with the usual neutrino. The Universe may get reionized by photons produced in these decays.

WDM candidates are gravitino, sterile neutrino and even more exotic particles. The assumption of thermal distribution of these particles may or may not be valid. Indeed, for the mass in the range $m \sim 1 - 10$ keV, the number density of dark matter particles at freeze out is much smaller than in equilibrium, otherwise their present mass density would be too large, see Section I.9.6.3 (barring the possibility of extremely large number of degrees of freedom at that time, $g_*(T_f) \sim 10^3 - 10^4$). This observation points towards non-thermal mechanism of WDM generation, and such a mechanism may well result in non-thermal spectrum. Hence, in concrete WDM models the estimate like (8.79) are at best valid within an order of magnitude.

Problem 8.12. Estimate the number of effective degrees of freedom $g_*(T_f)$ at freeze-out of the abundance of dark matter particles whose mass is $m = 1$ keV, assuming that just before the freeze-out these particles were in complete thermal equilibrium with cosmic plasma.

To end this Section, let us show that the condition for the growth of perturbations *in the expanding Universe* is still

$$k \lesssim k_{fs}(t),$$

where $k_{fs}(t)$ is the current critical value of conformal momentum (8.62). As a by-product, we show that once this condition is satisfied, the perturbations grow in the same way as for perfect fluid with $p = 0$, i.e., $\delta \propto a(t)$. This will support the discussion earlier in this Section and also will give the accurate proof that perturbations in collisionless CDM are correctly described in the ideal fluid approximation.

Let us make use of the Newtonian approach analogous to that given in Section 1.2. As before, we assume that collisionless particles are non-relativistic. Let us come back to the system of equations (8.63)–(8.65), but now we consider another background, which corresponds to the homogeneous expanding Universe, cf. (1.14),

$$\bar{\Phi} = \frac{2}{3}\pi G \mathbf{x}^2 \rho(t), \quad (8.83a)$$

$$f = f^{(0)}(U), \quad (8.83b)$$

where

$$\mathbf{U} = (\mathbf{u} - \mathbf{x}H(t))a(t), \quad (8.84)$$

$H = \dot{a}/a$, and $\rho(t)$ and $H(t)$ obey the continuity equation (1.15a) and the Friedmann equation. Their useful consequence is Eq. (1.15b). Recall that the background space-time is flat in the Newtonian approach, so that \mathbf{u} entering (8.84) is the physical velocity of a particle with respect to the static reference frame in this flat space, and $\mathbf{x}H$ is the velocity of the gas at the point \mathbf{x} . The variable \mathbf{U} has the meaning of the conformal velocity of a particle in the reference frame of the gas, while $\mathbf{U}/a(t)$ is the physical velocity in the latter frame. In accordance with the fact that conformal velocity of non-relativistic particle is time-independent in General Relativity, the unperturbed distribution function $f^{(0)}(U)$ is time-independent. Self-consistency of the entire approach is confirmed by the fact that the background (8.83) is a solution to the system (8.63)–(8.65).

Problem 8.13. Prove the last statement above.

It is convenient to use the variable $\mathbf{y} \equiv \mathbf{x}/a$ instead of \mathbf{x} and write the linearized equations for perturbations $\delta\phi = \Phi$ and $\delta\mathfrak{f}$ as follows,

$$\Delta\Phi = \frac{4\pi Gm}{a} \int d^3U \delta\mathfrak{f}, \quad (8.85a)$$

$$\frac{\partial\delta\mathfrak{f}}{\partial t} + \frac{\mathbf{U}}{a^2} \frac{\partial\delta\mathfrak{f}}{\partial\mathbf{y}} = \frac{\partial f^{(0)}}{\partial\mathbf{U}} \frac{\partial\Phi}{\partial\mathbf{y}}, \quad (8.85b)$$

where Δ is the Laplacian in coordinates y^i .

Problem 8.14. Derive the system of equations (8.85). Hint: Make use of Eq. (1.15b).

Upon the Fourier transformation in \mathbf{y} , $\Phi \rightarrow \Phi(t)e^{i\mathbf{k}\mathbf{y}}$, $\delta\mathfrak{f} \rightarrow \delta\mathfrak{f}(t, \mathbf{U})e^{i\mathbf{k}\mathbf{y}}$ we obtain the equations

$$-\mathbf{k}^2\Phi = \frac{4\pi Gm}{a} \int d^3U \delta\mathfrak{f} \equiv 4\pi Ga^2 \delta\rho, \quad (8.86a)$$

$$\frac{\partial\delta\mathfrak{f}}{\partial t} + i\frac{\mathbf{k}\mathbf{U}}{a^2} \delta\mathfrak{f} = i\mathbf{k} \frac{\partial f^{(0)}}{\partial\mathbf{U}} \Phi \equiv \frac{i\mathbf{k}\mathbf{U}}{U} \frac{\partial f^{(0)}}{\partial U} \Phi. \quad (8.86b)$$

When writing Eq. (8.86a) we recalled that the mass density is related to the distribution function in conformal velocity by

$$(\rho + \delta\rho)(t) = m \int d^3u f(t, \mathbf{U}) = \frac{m}{a^3(t)} \int d^3U f(t, \mathbf{U}).$$

Vector \mathbf{k} has the meaning of conformal momentum of the perturbation wave in the gas.

Let us first solve the system (8.86) in the large wavelength limit, $k \equiv |\mathbf{k}| \rightarrow 0$. To this end, we expand the unknown function δf in \mathbf{kU} . We are interested in the solution to Eq. (8.86b) which non-trivially depends on $\Phi(t)$ (the general solution is discussed below). This solution is

$$\delta f = i \frac{\mathbf{kU}}{U} \frac{\partial f^{(0)}}{\partial U} \int_{t_*}^t dt' \Phi(t') + \frac{(\mathbf{kU})^2}{U} \frac{\partial f^{(0)}}{\partial U} \int_{t_*}^t dt' \frac{1}{a^2(t')} \int_{t_*}^{t'} dt'' \Phi(t''), \quad (8.87)$$

where t_* is the time at which the large wavelength approximation becomes valid. We have written here two terms in the expansion and used the fact that the unperturbed distribution function $f^{(0)}$ is time-independent. We now use the solution (8.87) to find the mass density perturbation. The contribution of the first term in (8.87) vanishes, and we get

$$\begin{aligned} \delta\rho &= \frac{m}{a^3} \int d^3U \delta f = \frac{m}{a^3} \int d^3U \frac{(\mathbf{kU})^2}{u} \frac{\partial f^{(0)}}{\partial u} \int_{t_*}^t dt' \frac{1}{a^2(t')} \int_{t_*}^{t'} dt'' \Phi(t'') \\ &= -\rho \mathbf{k}^2 \int_{t_*}^t dt' \frac{1}{a^2(t')} \int_{t_*}^{t'} dt'' \Phi(t''), \end{aligned} \quad (8.88)$$

where we first evaluated the integral over the direction of the velocity and then integrated by parts. We divide the latter expression by ρ , take two time derivatives and use Eq. (8.86a). This gives the equation for the density perturbation

$$\left(\frac{\partial^2}{\partial t^2} + 2H \frac{\partial}{\partial t} - 4\pi G \rho \right) \delta = 0. \quad (8.89)$$

This equation coincides with Eq. (1.23) for pressureless ideal fluid, and also with Eq. (6.17) obtained in the complete relativistic theory. One of its solutions grows as $\delta(t) \propto a(t)$.

The limit of applicability of the large wavelength approximation is obtained by requiring that the higher order terms in the expansion (8.87) be small compared to the lower order ones. We know (see Sections 1.2 and 4.3) that the Newtonian potential Φ is time-independent at matter domination. Since $a \propto t^{2/3}$, the integrals in (8.87) are saturated at the upper limit, and the second term is suppressed by the factor $kUt/a^2(t)$. Hence, the large wavelength approximation is valid for

$$k \ll \frac{H(t)a(t)}{u(t)} \sim k_{fs}, \quad (8.90)$$

where $u(t) \equiv U/a(t)$. We see that k_{fs} is indeed as given in (8.62). Perturbations with $k \ll k_{fs}$ can be effectively described in the approximation of pressureless ideal fluid, while perturbations of higher momenta experience free streaming.

Let us consider for completeness the solution to the homogeneous equation (8.86b) in the large wavelength approximation,

$$\delta f = \exp \left(3i \cdot \mathbf{kU} \cdot t^{-1/3} \right) c(\mathbf{U}, \mathbf{k}),$$

where c is an arbitrary function, and we have set $a(t) = t^{2/3}$. Repeating the above calculation, we find that Eq. (8.89) is modified at small k by the term in the right hand side equal to

$$\text{const} \cdot \frac{1}{a^4(t)} \int d^3u \exp \left(3i \cdot \mathbf{kU} \cdot t^{-1/3} \right) (\mathbf{kU})^2 c(\mathbf{U}, \mathbf{k}).$$

This term is irrelevant at late times for $k \ll k_{fs}$, since it decays as a^{-4} , whereas each term in the left hand side of Eq. (8.89) behaves as a^{-2} . Hence, the expression (8.88) is indeed the dominant term in δf in the large wavelength case.

In the short wavelength case, opposite to (8.90), we use the general solution to Eq. (8.86b),

$$\delta f(t, \mathbf{U}) = \delta f(\mathbf{U}, \mathbf{k}, t_i) e^{-i\mathbf{k}\mathbf{U}(s-s_i)} + \int_{s_i}^s ds' a^2(s') e^{-i\mathbf{k}\mathbf{U}(s-s')} i\mathbf{k} \frac{\partial f^{(0)}}{\partial \mathbf{U}} \Phi(s'), \quad (8.91)$$

where t_i is the initial moment of time, at which the distribution function is assumed to be known, and

$$s = \int \frac{dt}{a^2(t)} = -3 \frac{t}{a^2(t)} = -\frac{2}{a^2 H} \propto -\frac{1}{t^{1/3}}.$$

We consider high momenta,

$$kUs \sim \frac{kU}{a^2 H} \sim \frac{k}{k_{fs}} \gg 1. \quad (8.92)$$

Let us study sufficiently late times, for which $(s_i - s)kU \gg 1$. Then the integral involving rapidly oscillating function in the second term in (8.91) is saturated at the upper limit of integration, and we find

$$\delta f_{(2)} = \frac{a^2(t)}{U} \frac{\partial f^{(0)}}{\partial U} \Phi(t). \quad (8.93)$$

The corresponding contribution to the density perturbation is

$$m\Phi \frac{1}{a} \int 4\pi dU f^{(0)} = \left\langle \frac{1}{u^2} \right\rangle \Phi \rho.$$

Its contribution to Eq. (8.86a) is small compared to the left hand side of that equation, so the gravitational potential is determined by the first term in Eq. (8.91),

$$\Phi = -\frac{4\pi G}{a\mathbf{k}^2} \int d^3U \delta f(\mathbf{U}, \mathbf{k}, s_i) e^{-i\mathbf{k}\mathbf{U}(s-s_i)}. \quad (8.94)$$

Because of the rapidly oscillating factor in the integrand, it decays exponentially as $|s - s_i|$ increases, so the contribution to the Newtonian potential due to particles obeying (8.92) gets damped in the time interval $\Delta t \sim (kU/a^2)^{-1} \ll H^{-1}$ after the initial time t_i . Note that our analysis is not valid for slow particles that do not obey (8.92), but the fraction of these particles is small in the situation we consider.

Density perturbations are also given by the integral with the rapidly oscillating factor,

$$\delta = \frac{\int d^3U \delta f(t_i, \mathbf{U}) e^{-i\mathbf{k}\mathbf{U}(s-s_i)}}{\int d^3U f^{(0)}(U)},$$

so they also get washed out. All these properties are in full agreement with the study of perturbations about the static background.

8.3.2 Bound on WDM particle mass from phase space density

This Section is a digression from our main line of presentation. We discuss here an approach to obtaining bounds on the mass of dark matter particle, which is independent of the studies of structure development. It is based on the following observation. Let the initial distribution function of dark matter particles be homogeneous, to the zeroth approximation, $f_i = f_i(\kappa)$, where κ is the conformal momentum. This function determines the number of particles in the phase space volume $d^3\kappa d^3x$,

i.e., it is the initial phase space density. The distribution function evolves in time and becomes inhomogeneous; it is different in galaxies and voids. The Liouville theorem states that the phase space density of collisionless particles does not change in time; rather, it flows from one region of phase space to another. This is an exact statement valid at *microscopic* level. In reality, one is interested in the phase space density averaged over macroscopic regions of phase space. Generally speaking, this *coarse grained phase space density*, which we still denote by $f(\kappa, \mathbf{x})$ in this Section, gets diluted in overdense regions of phase space and grows in underdense regions, see Fig. 8.3. Hence, the coarse grained phase space density obeys

$$f(\kappa, \mathbf{x}, t) \leq \max_{\kappa} f_i(\kappa). \quad (8.95)$$

This property also holds in the expanding Universe⁵ in the presence of large scale (by microscopic standards) gravitational fields. So, the phase space density of dark matter particles in gravitationally bound systems should obey the inequality (8.95). We set $a_0 = 1$ in this Section, so $\kappa = \mathbf{P}$ is the present momentum of dark matter particle, $f(\mathbf{P}, \mathbf{x}, t_0)$ is the present phase space density and $f_i(\mathbf{P})$ is the phase space density in the unrealistic homogeneous Universe.

The observational data obtained for various galaxies enable one to estimate the quantity

$$Q = \frac{\rho}{\langle v_{||}^2 \rangle^{3/2}},$$

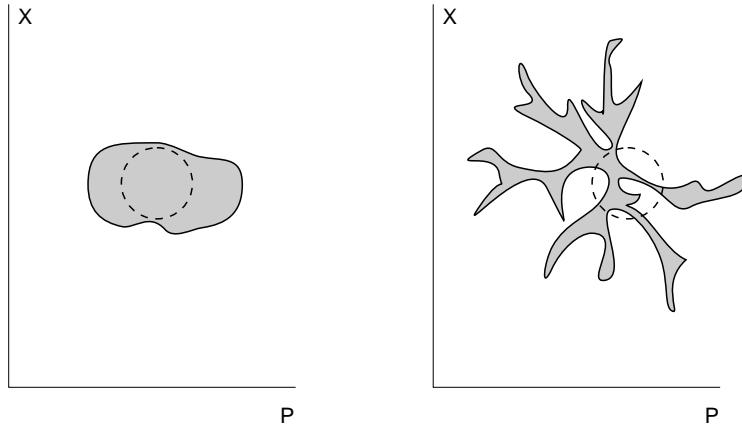


Fig. 8.3 In the beginning (left panel), particles occupy a certain region in phase space. The volume of this region does not change as the system evolves (right panel). However, the phase space density averaged over a macroscopic region (encircled) changes in time and decreases in the dense regions of phase space.

⁵It is important here that the phase space element $d^3 P d^3 l = (d^3 \kappa / a^3)(a^3 d^3 x)$ is time-independent.

where ρ is the mass density in the central region (which is determined from the gravitational potential) and $\langle v_{||}^2 \rangle$ is the average velocity squared along the line of sight in that region. We will see in a moment that the largest values of Q are most relevant. They are obtained for dwarf galaxies and belong to the range [43]

$$Q \equiv \mathcal{Q} \cdot 1 \frac{M_\odot/\text{pc}^3}{(\text{km/s})^3} = (5 \cdot 10^{-3} - 2 \cdot 10^{-2}) \frac{M_\odot/\text{pc}^3}{(\text{km/s})^3}. \quad (8.96)$$

Most of the mass in dwarf galaxies is due to dark matter, so that $\rho = \rho_{DM} = mn$, where n is the number density of dark matter particles. Dwarfs are spherically symmetric to a reasonable approximation, hence $\langle v_{||}^2 \rangle = (1/3)\langle v^2 \rangle$. The velocities of dark matter particles and stars are likely to be the same (and obey the virial theorem). Hence, we have

$$Q \simeq 3^{3/2} \frac{\rho_{DM}}{\langle v_{DM}^2 \rangle^{3/2}} = 3^{3/2} m^4 \frac{n}{\langle P^2 \rangle^{3/2}}.$$

The quantity $n/\langle P^2 \rangle^{3/2}$ estimates the number of particles per unit spatial volume per unit volume in momentum space, i.e., the phase space density of dark matter particles in a galaxy. Hence, we have an estimate

$$f(\mathbf{P}, \mathbf{x}) \simeq \frac{Q}{3^{3/2} m^4}.$$

The inequality (8.95) now gives a bound on the mass of dark matter particle, known as bound of the Tremaine–Gunn type [44],

$$m^4 \gtrsim \frac{Q}{3^{3/2} \max f_i}. \quad (8.97)$$

The concrete form of this bound depends on the initial distribution function $f_i(P)$, translated to the present epoch; it is often quite restrictive. We note that if f_i reaches the maximum value consistent with the Pauli principle for fermions, $f_i = g_X/(2\pi)^3$ (where g_X is the number of spin states), this bound is basically the same as that discussed in Section I.9.1. Most of the mechanisms of dark matter generation yield smaller maximum values of the initial distribution function, and the bound (8.97) is stronger than that given in Section I.9.1.

As an example, let us consider fermionic dark matter whose initial distribution function has the thermal form, normalized to the present mass density, Eq. (8.82). In this case the bound (8.97) gives

$$m^3 \gtrsim \frac{12\pi}{3^{3/2}} T_{0,\text{eff}}^3 \frac{Q}{\Omega_{DM} \rho_c},$$

Making use of (8.78) we find numerically

$$m \gtrsim 6 \text{ keV} \cdot \left(\frac{0.2}{\Omega_{DM}} \right)^{1/3} \left(\frac{\mathcal{Q}}{5 \cdot 10^{-3}} \right)^{1/3} \left(\frac{g_*(T_d)}{43/4} \right)^{1/3}, \quad (8.98)$$

where \mathcal{Q} is defined in (8.96), and we normalized the effective number of degrees of freedom g_* to its value at the epoch of neutrino decoupling.

The bound (8.98) works almost literally for dark matter made of sterile neutrinos ν_s , if they are produced in the early Universe via non-resonant transitions of the conventional neutrinos ν . These transitions are possible if there is small $\nu - \nu_s$ mixing, and they occur at the epoch just before neutrino decoupling. All factors in parentheses in (8.98) are close to 1, and the bound is [45, 46]

$$m_{\nu_s} \gtrsim 6 \text{ keV}.$$

Similar bounds are obtained for other WDM candidates, including gravitino [37].

We note that bounds at the level comparable to (8.98), and even stronger, are obtained by the analysis of Lyman- α forest, see, e.g., Ref. [47]. We emphasize in this regard that bounds discussed in this Section are very conservative, since they do not account for the fact that the maximum value of the distribution function *decreases* in the course of the evolution. In future, with better understanding of the structure formation in WDM model, these bounds will get stronger, and it is conceivable that the approach outlined here will give *estimates for* rather than bounds on the mass of dark matter particle, if dark matter is indeed warm.

To end this Section, we mention that the non-resonant mechanism of the sterile neutrino production involves two parameters, the sterile neutrino mass and $\nu_s - \nu$ mixing angle, see Section I.7.3. One combination of these parameters is determined from the dark matter density, and the remaining one can be chosen as the mass m . It is unambiguously related to the lifetime of the sterile neutrino, whose dominant decay channel is $\nu_s \rightarrow \nu\gamma$. These decays would generate diffuse γ -ray flux in the present Universe at the current level of sensitivity of X-ray telescopes. This effect is not observed, which implies the *upper* bound $m_{\nu_s} < 4 \text{ keV}$ [48]. Combined with the bound (8.98), it disfavors the simple mechanism of the non-resonant production of sterile neutrinos as the mechanism of dark matter generation.

8.4 Neutrino Free Streaming

Neutrinos give considerable contribution to the total energy density throughout the entire Hot Big Bang epoch. Therefore, they affect the evolution of the cosmological perturbations fairly strongly (for a review, see, e.g, Ref. [50]). Neutrino decoupling, as well as freeze-out of their abundance, occurs at temperatures of $2 - 3 \text{ MeV}$, see Section I.7.1, so free streaming phenomenon is fully relevant for neutrinos.

Let us recall, see Section I.7.2 and Eq. (8.78), that before e^+e^- -annihilation, neutrino temperature T_ν coincides with that of photons, while after the annihilation it is related to the photon temperature T as follows,

$$T_\nu(t) = \left(\frac{4}{11} \right)^{1/3} T(t). \quad (8.99)$$

The homogeneous part of the neutrino distribution function $f_\nu^{(0)}$ coincides with the Fermi-Dirac distribution to good accuracy. The number density of neutrinos decreases as a^{-3} , and its present value for every type of active neutrinos and its

antineutrino is

$$n_{\nu,0} = 112 \text{ cm}^{-3}. \quad (8.100)$$

Detailed studies show that the approximation of instantaneous decoupling that leads to (8.99) and (8.100) is actually quite good. The major corrections are due to the fact that electron neutrinos from the high energy tail of the spectrum still interact with electrons and positrons at e^+e^- -annihilation epoch. As a result, small part of energy released in e^+e^- -annihilations is transferred to neutrinos, and their spectrum gets slightly distorted. Conversely, photons get heated up slightly less. There are other, more subtle corrections. All these effects are accounted for in the following way: neutrino contribution to the total energy density is still written in the standard form using (8.99) for neutrino temperature, while corrections are encoded in the effective number of neutrino species N_ν entering the neutrino energy-momentum tensor. The value of the latter is [50] $N_\nu \simeq 3.046$. We see that the instantaneous decoupling approximation, which yields $N_\nu = 3$, works very well indeed. Let us note that another possible source of the spectrum distortion and modification of the effective number of neutrino species is the lepton asymmetry of the Universe. No evidence for sizeable lepton asymmetry of our Universe has been found so far, though.

At radiation domination, and after e^+e^- -annihilation, the following relation holds,

$$\rho_\nu = \rho \cdot R_\nu,$$

with

$$R_\nu \equiv \frac{\rho_\nu}{\rho} = \frac{\rho_\nu}{\rho_\nu + \rho_\gamma} = \frac{1}{1 + \frac{8}{7N_\nu} \left(\frac{T_\gamma}{T_\nu} \right)^4}, \quad (8.101)$$

where N_ν is the effective number of neutrino species, $N_\nu \simeq 3.05$ in the Standard Model (see above). Hence, the neutrino fraction in the total energy density is quite large,

$$R_\nu \simeq 0.41. \quad (8.102)$$

The neutrino fraction at matter domination is considerably smaller. Recall in this regard that neutrino oscillation experiments show that at least two of the three neutrino species are massive, and the lower limits on the masses are given by (see Appendix I.B)

$$m_1 \geq m_{sol} \simeq 0.009 \text{ eV}, \quad m_2 \geq m_{atm} \simeq 0.05 \text{ eV}. \quad (8.103)$$

The third neutrino may have very small and even zero mass; in the latter case the inequalities (8.103) are saturated. On the other hand, the neutrino masses may be much higher than 0.05 eV; then all three neutrinos must be degenerate in mass. The cosmological bound on the sum of neutrino masses is

$$\sum m_{\nu_i} < 0.2 - 1.0 \text{ eV}, \quad (8.104)$$

depending on priors on the cosmological parameters. We explain how this bound is obtained in Section 8.4.3.

Since the critical density in our Universe is $5.2 \cdot 10^{-6} \text{ GeV} \cdot \text{cm}^{-3}$, and $\Omega_M = 0.27$, we find from (8.100) that

$$\begin{aligned}\Omega_\nu &= \frac{n_{\nu 0} \sum m_{\nu i}}{\rho_c} = 0.001 - 0.02, \\ \frac{\Omega_\nu}{\Omega_M} &= 0.005 - 0.08,\end{aligned}\quad (8.105)$$

where smaller values are obtained for neutrino masses $(0, m_{sol}, m_{atm})$ and larger ones correspond to the sum of neutrino masses equal to 1 eV.

Neutrinos become non-relativistic at redshift $z = z_{nr}$, when (the numerical factor here is somewhat arbitrary; its choice is motivated by the fact that the average energy of relativistic fermion at temperature T equals $3.15 T$)

$$m_\nu \simeq 3.15 T_\nu(z) = 3.15 T_{\nu,0} \cdot (1 + z_{nr}) = 3.15 \left(\frac{4}{11} \right)^{1/3} T_0 \cdot (1 + z_{nr}). \quad (8.106)$$

This gives for $m = m_{sol}$ and $m = m_{atm}$

$$z_{1,nr} \simeq 16 \quad \text{and} \quad z_{2,nr} \simeq 94.$$

The maximum value of neutrino mass allowed by (8.104) is about 0.33 eV; neutrinos of this mass become non-relativistic at

$$z_{nr,max} \simeq 630. \quad (8.107)$$

All these redshifts correspond to matter domination. Hence, relic neutrinos are *hot dark matter* particles.

8.4.1 Early evolution: relativistic neutrinos

We begin with the evolution of perturbations at the early epoch when neutrinos are relativistic (effectively massless) and do not interact with the plasma. There is no interaction between neutrinos themselves either. At that epoch, the unperturbed energy-momentum tensor is expressed in the standard way in terms of the energy density ρ_ν and pressure $p_\nu = \rho_\nu/3$. Its perturbation is given by (8.57), i.e.,

$$\delta\rho_\nu(\eta, \mathbf{x}) = 3\delta p_\nu(\mathbf{x}) = \frac{1}{a^4} \int \kappa^3 d\kappa d\mathbf{n} f_\nu^{(0)}(\kappa) \cdot \delta f_\nu(\boldsymbol{\kappa}; \eta, \mathbf{x}) \quad (8.108a)$$

$$\delta T_i^0(\eta, \mathbf{x}) = -(\rho + p)v_i(\eta, \mathbf{x}) = -\frac{1}{a^4} \int \kappa^3 d\kappa d\mathbf{n} n_i f_\nu^{(0)}(\kappa) \cdot \delta f_\nu(\boldsymbol{\kappa}; \eta, \mathbf{x}) \quad (8.108b)$$

$$\Pi_j^i(\eta, \mathbf{x}) = \frac{1}{a^4} \int \kappa^3 d\kappa d\mathbf{n} \cdot \left(n_i n_j - \frac{1}{3} \delta_{ij} \right) \cdot f_\nu^{(0)}(\kappa) \cdot \delta f_\nu(\boldsymbol{\kappa}; \eta, \mathbf{x}), \quad (8.108c)$$

where $\boldsymbol{\kappa}$ is still conformal momentum of a particle and $\mathbf{n} = \boldsymbol{\kappa}/\kappa$. The evolution of the perturbations is governed by the Boltzmann equation (8.56) with vanishing

collision integral. In the relativistic case, this equation has the form

$$-\left(\Psi' + \mathbf{n} \frac{\partial \Phi}{\partial \mathbf{x}}\right) \cdot \frac{\partial \log f_\nu^{(0)}}{\partial \log \kappa} + \delta f'_\nu + \mathbf{n} \frac{\partial \delta f_\nu}{\partial \mathbf{x}} = 0. \quad (8.109)$$

It is convenient to write it in the Fourier representation in the variable \mathbf{x} ,

$$\delta f(\boldsymbol{\kappa}, \mathbf{x}, \eta) \rightarrow \delta f(\boldsymbol{\kappa}, \mathbf{k}, \eta) e^{i\mathbf{k}\mathbf{x}}, \quad \Phi(\mathbf{x}, \eta) \rightarrow \Phi(\mathbf{k}, \eta) e^{i\mathbf{k}\mathbf{x}}, \quad \text{etc.}$$

As usual, the vector \mathbf{k} is the conformal momentum of the perturbation wave. Equation (8.109), written in momentum representation, is

$$-(\Psi' + i\mathbf{k}\mathbf{n}\Phi) \cdot \frac{\partial \log f_\nu^{(0)}}{\partial \log \kappa} + \delta f'_\nu + i\mathbf{k}\mathbf{n}\delta f_\nu = 0. \quad (8.110)$$

Note that the gravitational potentials entering (8.109) are generated not only by the neutrino perturbations but also perturbations in other components of the cosmic medium: neutrinos interact with other particles via gravitational force.

The homogeneous part of the linear equation (8.109) does not depend on κ , while the source term depends on κ only through $f_\nu^{(0)}(\kappa)$ which is independent of time. Furthermore, the total neutrino distribution function initially has the thermal form in the case of the adiabatic perturbations. Hence, the dependence of the distribution function on κ remains locally thermal,

$$f_\nu \propto \left[\exp \left(\frac{\kappa}{a(T_\nu + \delta T_\nu)} \right) + 1 \right]^{-1}, \quad \delta T_\nu = \delta T_\nu(\mathbf{n}; \eta, \mathbf{x}).$$

The evolution of neutrino perturbations is encoded in the dependence of local temperature on \mathbf{x} and η . In our notations, the perturbation of the distribution function is related to the neutrino temperature perturbation as follows,

$$\delta f_\nu(\boldsymbol{\kappa}; \eta, \mathbf{x}) = -\frac{\partial \log f_\nu^{(0)}}{\partial \log \kappa} \cdot \frac{\delta T_\nu(\mathbf{n}; \eta, \mathbf{x})}{T_\nu}.$$

We do not use this “temperature” formalism in this book.

Let us first study superhorizon modes at radiation domination. In other words, we consider initial data for neutrino perturbations. As before, we discuss the adiabatic mode. We know from Section 5.2 that the superhorizon adiabatic mode has the following property,

$$\delta_\nu = \delta_\gamma = -2\Phi, \quad (8.111)$$

and that these quantities do not depend on time. To find the gravitational potential Ψ in the conformal Newtonian gauge, we use the Einstein equations in the scalar sector (8.58). To this end, we note that Eq. (8.58b) with time-independent potentials, $\Psi' = \Phi' = 0$, gives

$$v_\nu = v_\gamma = -\frac{1}{2}\eta\Phi. \quad (8.112)$$

We used here the following relations valid at radiation domination: $2\pi G(\rho + p) = H^2$, $aH = \eta^{-1}$. The relation $v_\nu = -\frac{1}{2}\eta\Phi$ can also be obtained from the Vlasov equation (8.110). Indeed, let us expand the perturbation δf_ν in $(\mathbf{k}\mathbf{n})$,

$$\delta f_\nu = \delta f_{\nu 0} + \mathbf{k}\mathbf{n}\delta f_{\nu 1} + (\mathbf{k}\mathbf{n})^2\delta f_{\nu 2} + \dots$$

The initial data for the adiabatic mode are

$$\delta f_{\nu 1}, \delta f_{\nu 2}, \dots \rightarrow 0 \quad \text{as } \eta \rightarrow 0, \quad (8.113)$$

while $\delta f_{\nu 0}$ is independent of κ as $\eta \rightarrow 0$. These initial data follow from the fact that the properties of the medium are the same everywhere in space for the superhorizon adiabatic mode, and they coincide with the properties of the unperturbed medium modulo local advance time ϵ , see (5.9), (5.10). In other words, the energy density perturbations vanish for *all* components at constant energy density hypersurfaces as $\eta \rightarrow 0$, while metric perturbations do not vanish.⁶ So, the neutrino distribution function in any gauge is proportional to $f_\nu^{(0)}$ with the proportionality coefficient independent of κ (but depending on \mathbf{x}). In this way we arrive at (8.113).

The solution to Eq. (8.110) with the initial data (8.113) is straightforwardly found at small η . To the lowest non-trivial order in η we have

$$\delta f_{\nu 1} = -i \left(\delta f_{\nu 0} - \Phi \frac{\partial \log f_\nu^{(0)}}{\partial \log \kappa} \right) \eta, \quad \delta f_{\nu 2} = -\frac{1}{2} \left(\delta f_{\nu 0} - \Phi \frac{\partial \log f_\nu^{(0)}}{\partial \log \kappa} \right) \eta^2.$$

The integrals in (8.108b) and (8.108c) are calculated by making use of the identities

$$\langle n_i n_j \rangle = \frac{1}{3} \delta_{ij}, \quad \langle n_i n_j n_l n_m \rangle = \frac{1}{15} (\delta_{ij} \delta_{lm} + \delta_{il} \delta_{jm} + \delta_{im} \delta_{jl}),$$

where brackets mean *angular average*. As a result, we obtain

$$-(\rho_\nu + p_\nu)v_{\nu i} = \frac{ik_i}{3}(\delta\rho_\nu + 4\Phi\rho_\nu)\eta \quad (8.114)$$

$$\Pi_{ij} = -\frac{1}{15} \left(k_i k_j - \frac{1}{3} \delta_{ij} \right) (\delta\rho_\nu + 4\Phi\rho_\nu)\eta^2 \quad (8.115)$$

The first of these relations is indeed reduced to (8.112) by using (8.111). Let us recall the definition (8.59) of the anisotropic stress potential π and obtain its initial value from (8.115),

$$\pi_\nu = -\frac{1}{15} k^2 \eta^2 \Phi. \quad (8.116)$$

We are now in a position to find the initial value of the gravitational potential Ψ from Eq. (8.58d). At radiation domination after neutrino decoupling we have

$$\rho_\nu + p_\nu = \frac{4}{3}\rho_\nu = \frac{4}{3}\rho \cdot R_\nu,$$

⁶Note that the definition of the adiabatic mode is generalized here in an obvious way.

where R_ν is defined in (8.101). We express the total unperturbed energy density in terms of the Hubble parameter using the Friedmann equation and obtain from (8.58d) the following equation (assuming that the anisotropic stress is solely due to the neutrino component)

$$\Delta(\Phi + \Psi) = -6a^2H^2R_\nu\pi_\nu = -\frac{6}{\eta^2}R_\nu\pi_\nu.$$

With the initial condition (8.116), this equation gives the non-trivial relation between the gravitational potentials for the superhorizon adiabatic mode,

$$\Psi = -\left(1 + \frac{2}{5}R_\nu\right)\Phi.$$

Numerically, we find from (8.102) that

$$\Psi = -1.16\Phi.$$

This effect has observational consequences and, moreover, there is experimental evidence that it does affect CMB properties [49].

Problem 8.15. *Find initial data for perturbations entering the horizon at matter domination but before recombination.*

We now turn to the analysis of the subhorizon neutrino perturbations. We begin with radiation domination. At that epoch, gravitational potentials decrease in time (see, e.g., Eq. (4.13); we will see in a moment that Ψ also decreases). Hence, the Vlasov equation (8.110) reduces to the Liouville equation. The solution to the latter is the wave propagating without distortion,

$$\delta f_\nu = \delta f_\nu(\boldsymbol{\kappa}; \mathbf{k}) \cdot e^{-i\eta\mathbf{k}\mathbf{n}} = \delta f_\nu(\boldsymbol{\kappa}; \mathbf{k}) \cdot e^{-ik\eta \cos \theta}, \quad (8.117)$$

where θ is the angle between vectors \mathbf{k} and $\boldsymbol{\kappa}$. Since $\rho_\nu, p_\nu \propto a^{-4}$, we obtain after performing the angular integration in (8.108) that the perturbations of neutrino energy density, velocity and anisotropic stress decrease in time at $k\eta \gg 1$ as

$$\delta_\nu \propto \frac{\sin k\eta}{k\eta}, \quad v_\nu \propto \frac{\cos k\eta}{k\eta}, \quad \pi_\nu \propto \frac{\sin k\eta}{k\eta}.$$

The neutrino perturbations are different in this respect from the perturbations of relativistic ideal fluid whose energy density perturbations stay constant in time. This is precisely the effect of free streaming.

We can now justify the irrelevance of the gravitational potentials in Eq. (8.110). It follows from the above solution that the neutrino contributions to the right hand side of the Einstein equations (8.58) decrease as $a^{-2}\eta^{-1} \propto \eta^{-3}$. Therefore, the potential Φ is determined by the baryon-photon component and its amplitude decays as η^{-2} , while the potential Ψ gradually approaches Φ . Their effect in Eq. (8.110) is indeed negligible for $k\eta \gg 1$.

Let us finally turn to the subhorizon neutrino perturbations at matter domination, still for relativistic neutrinos. Non-relativistic matter generates constant in time gravitational potentials. Therefore, the solution to Eq. (8.110) is

$$\delta f_\nu = \Phi \cdot \frac{\partial \log f_\nu^{(0)}}{\partial \log \kappa} + \text{const} \cdot e^{-i\mathbf{k}\mathbf{n}\eta}.$$

Accordingly, the density contrast has a part proportional to the gravitational potential,

$$\delta_\nu = -4\Phi + \text{const} \cdot \frac{\sin(k\eta)}{k\eta}. \quad (8.118)$$

The contrast δ_ν approaches the value (-4Φ) at late times. This value corresponds to stationary state in which neutrino pressure balances the gravitational force. Thus, neutrino perturbations are again suppressed, now compared to dark matter and baryon perturbations which grow as $a(\eta)$.

We note that so far we have neglected the effect of neutrino component on the evolution of other components at matter domination. We turn to this effect later on, and here we state that its back reaction *on the neutrino component itself* is small, so one can indeed consider gravitational potentials unaffected by neutrinos when studying the evolution of the neutrino perturbations.

To end this Section, we briefly describe an efficient technique for obtaining solutions to the Vlasov and Einstein equations numerically, which also gives additional analytic insight into the properties of neutrino perturbations. Let us introduce, in momentum space, the following variable,

$$F_\nu(\mathbf{n}; \eta, \mathbf{k}) = \frac{4\pi}{\rho_\nu a^4} \int \kappa^3 d\kappa f_\nu^{(0)}(\kappa) \cdot \delta f_\nu(\kappa; \eta, \mathbf{x}) \cdot e^{i\mathbf{k}\mathbf{x}} d^3\mathbf{x}, \quad (8.119)$$

where \mathbf{n} is still the unit vector along κ . We multiply Eq. (8.110) by $f_\nu^{(0)}$ and integrate the result over κ with weight κ^3 . In this way we obtain the equation

$$4(\Psi' + i\mathbf{k}\mathbf{n}\Phi) + F'_\nu + i\mathbf{k}\mathbf{n}F_\nu = 0. \quad (8.120)$$

The formal solution to this equation can be written as follows,

$$F_\nu(\mathbf{n}; \eta, \mathbf{k}) = e^{-i\mathbf{k}\mathbf{n}\eta} F_\nu(\mathbf{n}; 0, \mathbf{k}) - 4 \int_0^\eta e^{-i\mathbf{k}\mathbf{n}(\eta-\eta')} (i\mathbf{k}\mathbf{n}\Phi + \Psi') d\eta'. \quad (8.121)$$

Note that in the absence of the gravitational potentials, Eq. (8.120) is homogeneous, and its solution is given by the first term in (8.121), cf. (8.117). Upon integrating by parts the first term in the integrand of (8.121) we obtain

$$\begin{aligned} F_\nu(\mathbf{n}; \eta, \mathbf{k}) &= e^{-i\mathbf{k}\mathbf{n}\eta} [F_\nu(\mathbf{n}; 0, \mathbf{k}) + 4\Phi(0)] \\ &\quad - 4\Phi(\eta) + 4 \int_0^\eta e^{-i\mathbf{k}\mathbf{n}(\eta-\eta')} (\Phi' - \Psi') d\eta'. \end{aligned} \quad (8.122)$$

This solution shows that for time-independent gravitational potentials, the function F_ν effectively (when integrated over \mathbf{n} with smooth weight) tends to (-4Φ) at late times, cf. (8.118).

It is convenient to expand $F_\nu(\mathbf{n}; \eta, \mathbf{k})$ in the Legendre polynomials of the variable $\mathbf{k}\mathbf{n}/k = \cos\theta$,

$$F_\nu(\mathbf{n}; \eta, \mathbf{k}) = \sum_{l=0}^{\infty} (2l+1) (-i)^l \cdot F_{\nu,l}(\eta, \mathbf{k}) \cdot P_l\left(\frac{\mathbf{k}\mathbf{n}}{k}\right). \quad (8.123)$$

The choice of the numerical coefficients here is dictated by the normalization of the Legendre polynomials, see (F.33). According to (8.108), the neutrino density contrast, velocity potential and anisotropic stress potential are expressed as follows,

$$\begin{aligned} \delta_\nu &= \frac{1}{4\pi} \int d\mathbf{n} F_\nu = \frac{2\pi}{4\pi} \int_{-1}^1 dx P_0(x) F_\nu = F_{\nu,0}, \\ v_\nu &= -i \frac{3}{16\pi} \int d\mathbf{n} \frac{\mathbf{k}\mathbf{n}}{k^2} F_\nu = -i \frac{6\pi}{16\pi k} \int_{-1}^1 dx P_1(x) F_\nu = -\frac{3}{4k} F_{\nu,1}, \\ \pi_\nu &= \frac{3}{16\pi} \int d\mathbf{n} \left(\left(\frac{\mathbf{k}\mathbf{n}}{k} \right)^2 - \frac{1}{3} \right) F_\nu = \frac{6\pi}{16\pi} \int_{-1}^1 dx \frac{2}{3} P_2(x) F_\nu = -\frac{1}{2} F_{\nu,2}. \end{aligned}$$

We now insert the expansion (8.119), (8.123) into the Vlasov equation (8.120). We then see that the evolution of the function F_ν depends on \mathbf{n} only through the projection of \mathbf{n} on vector \mathbf{k} . This is a consequence of the homogeneity and isotropy of the background. Due to this property, the representation in terms of the Legendre polynomials (8.123) is convenient, as it reduces Eq. (8.120) to the system of linear first order equations for moments $F_{\nu,l}(k, \eta)$. To obtain this system, we use Eqs. (F.27) and (F.28) and write the Vlasov equation (8.120) as follows

$$4[\Psi' \cdot P_0(\cos\theta) + ik\Phi P_1(\cos\theta)] + F'_\nu + ikP_1(\cos\theta) \cdot F_\nu = 0.$$

We insert the expansion (8.123) into this equation and again expand the result in the Legendre polynomials with the use of the recurrence relation (F.26). In this way we obtain the system of equations

$$\delta'_\nu = \frac{4}{3}k^2 v_\nu - 4\Psi', \quad (8.124a)$$

$$v'_\nu = -\Phi - \frac{1}{4}\delta_\nu - \pi_\nu, \quad (8.124b)$$

$$\pi'_\nu = \frac{4}{15}k^2 v_\nu + \frac{3}{10}k F_{\nu,3}, \quad (8.124c)$$

$$F'_{\nu,l} = k \cdot \left(\frac{l}{2l+1} F_{\nu,l-1} - \frac{l+1}{2l+1} F_{\nu,l+1} \right), \quad l = 3, 4, \dots \quad (8.124d)$$

Note that the first two equations of this system, (8.124a) and (8.124b), are equivalent to the continuity equation (8.60) and Euler equation (8.61), with $p_\nu = \rho_\nu/3$ for relativistic neutrinos.

Problem 8.16. Derive the system of equations (8.124).

The properties of neutrino perturbations in the superhorizon regime are reproduced quite straightforwardly by the analysis of the system (8.124). Namely, (8.114) and (8.115) are obvious from (8.124), while the higher neutrino multipoles behave as $F_{\nu,l}(\eta) \propto \eta^l$ as $\eta \rightarrow 0$. Let us discuss the dynamics of multipoles $F_{\nu,l}$ after the horizon entry. To this end, we analyze the system (8.124) at qualitative level. Clearly, higher multipoles grow at early times due to the redistribution of power between multipoles, while the lowest multipoles decrease. In this “wave of power”, the higher multipoles first grow, and then decrease as the power is transferred to even higher multipoles. The initial growth of higher multipoles is clearly seen, since at small $k\eta$ the system (8.124) gives $F'_{\nu,l+1} \sim F_{\nu,l}$. On the other hand, this effect leads to strong suppression of lower multipoles, in particular, the density contrast and velocity, at late times.

The role of higher multipoles as “parasiting” on lower ones is clearly seen by using the density contrast δ_ν (monopole) as an example. It enters only the first two equations of the system (8.124). By differentiating Eq. (8.124a) over conformal time and using (8.124b) to eliminate the velocity, we obtain the second order equation for the monopole,

$$\delta''_\nu + \frac{k^2}{3} \delta_\nu = -\frac{4k^2}{3} \pi_\nu - \frac{4}{3} k^2 \Phi - 4\Psi''. \quad (8.125)$$

If not for the anisotropic stress potential, the equation for the monopole would coincide with the equation for the density contrast of relativistic ideal fluid. The monopole δ_ν would oscillate with constant amplitude. The anisotropic stress potential connects the monopole to other multipoles, so that power leaks into higher multipoles, while the amplitude of the monopole decreases in time.

Let us study in somewhat more detail what happens to multipoles after the horizon entry. The evolution of the multipoles $F_{\nu,l}$ can be found by making use of the formal solution (8.122). We insert there the Legendre polynomial expansion (8.123) and similar expansion of the plane wave (see (F.35)),

$$e^{-i\mathbf{k}\mathbf{n}\eta} = \sum_{l=0}^{\infty} (2l+1) (-i)^l \cdot j_l(k\eta) \cdot P_l\left(\frac{\mathbf{k}\mathbf{n}}{k}\right), \quad (8.126)$$

where j_l is the spherical Bessel function. Making use of the initial data just described, we obtain the formal solution for multipoles,⁷

$$\begin{aligned} F_{\nu,l}(\eta, \mathbf{k}) &= (-2\Phi_{(i)}\delta_{l,0} + 4\Phi_{(i)}) j_l(k\eta) \\ &\quad - 4\Phi(\eta)\delta_{l,0} + 4 \int_0^\eta d\eta' (\Phi' - \Psi') j_l[k(\eta - \eta')]. \end{aligned} \quad (8.127)$$

⁷The solution (8.127) is, generally speaking, not explicit, since the gravitational potentials Φ and Ψ are themselves determined in part by the neutrino perturbations. The general analysis of multipoles goes through, nevertheless.

Since the potentials $\Phi(\eta)$ and $\Psi(\eta)$ are negligible at late times, this formal solution shows that multipoles (with the exception of the monopole) behave as

$$F_{\nu,l} \propto j_l(k\eta). \quad (8.128)$$

This again shows that higher multipoles grow just after the horizon entry, and then decrease. The maximum of the l -th multipole is reached at $\eta \sim l/k$, and then this multipole decreases as $1/(k\eta)$. The result that the neutrino multipoles decrease is valid both at radiation and matter domination as long as neutrinos stay relativistic.

The exact solution to the system (8.124) gives more complicated dependence on time, especially at radiation domination, when neutrinos affect the gravitational potentials considerably. The qualitative behavior, however, is the same: the gravitational potentials decrease at radiation domination, so their effect on $F_{\nu,l}$ does not save the neutrino perturbations from the decay.

8.4.2 Non-relativistic neutrinos

We now study later cosmological epoch when neutrinos are non-relativistic. As we discussed earlier, this situation occurs at matter domination when $z < z_{nr}$. At that time, dark matter and baryon perturbations evolve together; we use the notation DB for this component, and reserve the notation M for dark matter, baryons and non-relativistic neutrino together.

Since neutrinos are not the dominant component at matter domination, the analysis of Section 8.3.1 does not apply here in its part concerning the gravitational potentials. Instead, the neutrino effects are small (suppressed by the ratio Ω_ν/Ω_M), and the gravitational potentials are determined by dark matter and baryons. Neglecting the neutrino effects for the time being, we have $\Psi = \Phi$, and Φ is independent of time. The formula (8.88) remains valid in the large wavelength approximation, if ρ and $\delta\rho$ are understood as neutrino mass density and its perturbation. We recall that $a \propto t^{2/3}$ and obtain from (8.88) that the density contrast is

$$\delta_\nu = -\mathbf{k}^2 \Phi \int_{t_*}^t dt' \frac{t' - t_*}{a^2(t')} = -\frac{3}{2} \mathbf{k}^2 \Phi \frac{t^2}{a^2(t)} \quad \text{at } t \gg t_*.$$

Making use of the background Einstein equations and their perturbations we find, still neglecting the neutrino effects, that the late time asymptotics is

$$\delta_\nu = \delta_{DB}, \quad (8.129)$$

i.e., the neutrino perturbations catch up with dark matter and baryons. This is precisely the same result as the one we obtained in Section 7.1.1 for the gas of baryons. This is not surprising, since neutrino velocity is irrelevant for large wavelength modes, and hence neutrinos are indistinguishable from pressureless fluid in this sense.

Since the rest energy is much greater than the kinetic energy in the non-relativistic limit, π_ν and δp_ν are strongly suppressed as compared to $\delta\rho_\nu$. This is

clearly seen from the general formulas (8.57). Since anisotropic stress is negligible, one has $\Phi + \Psi = 0$. This is valid irrespective of whether or not neutrino effects are neglected.

Repeating the discussion leading to (8.90), we find that the conformal momentum critical for free streaming at time t is given by

$$k_{fs} \sim \frac{H(t)a(t)}{u_\nu(t)}, \quad (8.130)$$

where $u_\nu(t)$ is the typical neutrino velocity. The estimate for the latter is obtained from (8.77) and reads,

$$u_\nu = 1.6 \frac{T_\nu}{m_\nu} = 1.6 \left(\frac{4}{11} \right)^{1/3} \frac{T_0}{m_\nu} (1 + z(t)).$$

Numerically,

$$u_\nu(z) \simeq 0.27 \cdot 10^{-3} (1 + z) \cdot \left(\frac{1 \text{ eV}}{m_\nu} \right), \quad (8.131)$$

and the critical conformal momentum and physical wavelength (8.62) are

$$\frac{k_{fs}(z)}{a_0} = 1.1 \cdot \frac{\sqrt{\Omega_\Lambda + (1+z)^3 \Omega_M}}{(1+z)^2} \left(\frac{m_\nu}{1 \text{ eV}} \right) \left(\frac{h}{0.705} \right) \text{Mpc}^{-1}, \quad (8.132)$$

$$\lambda_{fs}(z) = 5.9 \cdot \frac{1+z}{\sqrt{\Omega_\Lambda + (1+z)^3 \Omega_M}} \left(\frac{1 \text{ eV}}{m_\nu} \right) \left(\frac{0.705}{h} \right) \text{Mpc}. \quad (8.133)$$

Here we used the standard representation for the Hubble parameter, $H^2(z) = H_0^2 \cdot [\Omega_\Lambda + (1+z)^3 \Omega_M]$. It is clear from (8.132) that the free streaming suppression is characteristic of the most interesting scales. As an example, for neutrino mass 0.3 eV the present critical wavelength is about 20 Mpc. More realistic estimate is obtained by noting that the asymptotics (8.129) is established in a few Hubble times rather than instantaneously. We set $z = 1$ in Eq. (8.132) for an estimate, and obtain the present wavelength $\lambda_0 \simeq 45$ Mpc. Smaller scale perturbations in the neutrino component are strongly suppressed today.

Let us make a comment concerning the value of velocity given in (8.131). Comparing it with the typical velocities in structures (stars in galaxies, galaxies in clusters), we see that neutrinos can fall into compact objects quite late. As an example, the velocities of stars in large galaxies are of order $\sim 10^{-3}$, so neutrinos of mass of a fraction of eV can get trapped in a galaxy today at best.

Problem 8.17. Estimate pressure and anisotropic stress potential of the non-relativistic neutrino component for $k \gg k_{fs}$. Hint: Use the general formulas (8.57), calculate π_ν by making use of the higher order terms in the expansion (8.87).

Let us show explicitly that neutrino perturbations are suppressed due to free streaming for $k \gg k_{fs}$. To this end, we use the Liouville equation (8.85b) where the potential Φ is

external and does not depend on time. Its solution is given by (8.91) with zero initial value of δf_ν , since neutrino perturbations are small when neutrinos are relativistic. We recall the formula (8.93), which is valid in the small wavelength approximation, and obtain

$$\delta_\nu = \frac{\int d^3U \delta f_\nu}{\int d^3U f_\nu^{(0)}} = -a^2 \left\langle \frac{1}{U^2} \right\rangle \Phi = \left\langle \frac{1}{u_\nu^2} \right\rangle \Phi.$$

To compare this perturbation with the baryon and dark matter perturbation δ_{DB} , we express the gravitational potential through the latter. We have for subhorizon modes

$$-\Phi = \frac{1}{k^2} \cdot 4\pi G a^2 \rho_{DB} \delta_{DB} \sim \frac{a^2 H^2}{k^2} \delta_{DB}.$$

Hence, the neutrino density contrast is indeed small,

$$\frac{\delta_\nu}{\delta_{DB}} \sim \frac{H^2 a^2}{k^2 u_\nu^2} \sim \left(\frac{k_{fs}}{k} \right)^2.$$

This calculation shows also that δ_ν tends to δ_{DB} as k approaches k_{fs} . This reiterates the result (8.130).

Let us briefly discuss the modification of neutrino multipole approach to the case of non-relativistic neutrino. Since the integrands in the expressions (8.57) for energy-momentum tensor now depend on time (through the scale factor $a = a(\eta)$), these integrals cannot be evaluated explicitly. So, instead of the representation (8.119), (8.123) one now introduces the expansion in the Legendre polynomials with coefficients depending on the absolute value of the conformal momentum,

$$\delta f_\nu(\kappa, \mathbf{n}; \eta, \mathbf{k}) = \sum_{l=0}^{\infty} (2l+1) (-i)^l \cdot F_{\nu,l}(\kappa; \eta, \mathbf{k}) \cdot P_l\left(\frac{\mathbf{k}\mathbf{n}}{k}\right). \quad (8.134)$$

As a result, the expressions for basic quantities are

$$\delta \rho_\nu = \frac{4\pi}{a^4} \int \kappa^2 \sqrt{\kappa^2 + m_\nu^2 a^2} \cdot d\kappa f_\nu^{(0)}(\kappa) \cdot F_{\nu,0}, \quad (8.135a)$$

$$\delta p_\nu = \frac{4\pi}{3a^4} \int \frac{\kappa^4 d\kappa}{\sqrt{\kappa^2 + m_\nu^2 a^2}} d\kappa f_\nu^{(0)}(\kappa) \cdot F_{\nu,0}, \quad (8.135b)$$

$$(\rho_\nu + p_\nu) \cdot v_\nu = -\frac{4\pi}{a^4 k} \int \kappa^3 d\kappa f_\nu^{(0)}(\kappa) \cdot F_{\nu,1}, \quad (8.135c)$$

$$(\rho_\nu + p_\nu) \cdot \pi_\nu = -\frac{8\pi}{3a^4} \int \frac{\kappa^4 d\kappa}{\sqrt{\kappa^2 + m_\nu^2 a^2}} f_\nu^{(0)}(\kappa) \cdot F_{\nu,2}. \quad (8.135d)$$

The system of equations for $F_{\nu,l}$ is obtained by inserting the expansion (8.134) into the Vlasov equation (Eq. (8.56) with $C[f_\nu] = 0$) and performing the Fourier transformation in the coordinate \mathbf{x} . This system is

$$F'_{\nu,0} = -\frac{\kappa k}{\sqrt{\kappa^2 + m_\nu^2 a^2}} F_{\nu,1} + \frac{\partial \log f_\nu^{(0)}}{\partial \log \kappa} \Psi', \quad (8.136a)$$

$$F'_{\nu,1} = \frac{\kappa k}{3\sqrt{\kappa^2 + m_\nu^2 a^2}} (F_{\nu,0} - 2F_{\nu,2}) - \frac{k}{3\kappa} \sqrt{\kappa^2 + m_\nu^2 a^2} \cdot \frac{\partial \log f_\nu^{(0)}}{\partial \log \kappa} \Phi, \quad (8.136b)$$

$$F'_{\nu,l} = \frac{\kappa k}{(2l+1)\sqrt{\kappa^2 + m_\nu^2 a^2}} (lF_{\nu,l-1} - (l+1)F_{\nu,l+1}), \quad l = 2, 3, \dots \quad (8.136c)$$

Equations (8.136) together with the relations (8.135) and Einstein equations determine the evolution of the massive neutrino component.

Problem 8.18. Derive the equations (8.136) and relations (8.135).

Problem 8.19. Find initial conditions for the multipoles $F_{\nu,l}$.

We note that in practice, the system of equations (8.136) is truncated by setting at some large $l = l'$

$$F_{\nu,l'+1} \simeq (2l' + 1) \frac{\sqrt{\kappa^2 + m_\nu^2 a^2}}{\kappa k \eta} F_{\nu,l'} - F_{\nu,l'-1}.$$

This approximation is suggested by the recurrence relation (F.3) for the Bessel functions. The approximation works well for higher multipoles, since the coefficients multiplying the Legendre polynomials in the expansion (8.134) rapidly decay as soon as neutrino becomes non-relativistic. In fact, the system of equations for the multipoles is truncated in a similar way for massless neutrino too; the maximum multipole number l' is taken quite large in that case to ensure good accuracy.

8.4.3 Neutrino effect on matter perturbations. Cosmological bound on neutrino masses

Peculiarities of cosmology with massive neutrinos enable one to search for the cosmological evidence for the neutrino masses. Presently, only bounds on neutrino masses are obtained in this way, but these bounds are already stronger than direct bounds obtained in terrestrial experiments.

Let us compare two cosmological models, one with massless and another with massive neutrinos. We consider spatially flat Universe and take the present value of the Hubble parameter, the sum $\Omega_M = \Omega_{DB} + \Omega_\nu$ and the parameter Ω_Λ to be the same in the two models. Also, the two models are assumed to have one and the same number of neutrino species. The parameter Ω_ν is unambiguously related to the sum of neutrino masses, and does not depend on the hierarchy between these masses. So, an additional parameter in the second model is the number of the heaviest neutrinos species, $N_{\nu,m}$. If neutrinos have masses substantially exceeding $m_{atm} \simeq 0.05$ eV, then all three neutrino species are degenerate in mass and $N_{\nu,m} = 3$; in the case of inverse hierarchy $m_1 \approx m_2 \approx m_{atm}$ and $m_3 \ll m_{atm}$, so that $N_{\nu,m} = 2$; direct hierarchy implies $N_{\nu,m} = 1$, since one neutrino has mass m_{atm} , and others are much lighter (see Appendix I.C). Neutrinos of the largest mass are the most important to us; this mass and the parameter Ω_ν are related by

$$\Omega_\nu = \frac{N_{\nu,m} m_\nu n_{\nu,0}}{\rho_c}.$$

This gives

$$m_\nu = 23 \text{ eV} \cdot \left(\frac{h}{0.705} \right)^2 \cdot \left(\frac{2}{N_{\nu,m}} \right) \cdot \Omega_\nu. \quad (8.137)$$

Our first purpose is to find the dependence of perturbations on Ω_ν and $N_{\nu,m}$ for $\Omega_\nu \ll \Omega_M$. We note that for given Ω_ν , and hence $\Omega_{DB} = \Omega_M - \Omega_\nu$, the values of Ω_{CDM} and Ω_B separately are unimportant for us: baryons and dark matter evolve together soon after recombination, while the neutrino behavior is the same in the two models until $z = z_{nr}$. Choosing Ω_M the same in the two models is motivated by the fact that this quantity can be determined independently, say, by SNe Ia observations (see Section I.4.6). These observations measure the cosmological expansion rate at fairly low z , so the result is sensitive to Ω_M , rather than to Ω_{DB} and Ω_ν separately.

Let us consider perturbations, for which neutrino free streaming occurs up until today. Neutrino perturbations are negligible for these modes, but neutrinos affect dark matter and baryon perturbations. For definiteness, let us analyze the effect in terms of the present value of $\delta_M = \delta\rho_{DB}/\rho_M$.

One (but not the only, see below) effect due to the massive neutrinos is the suppression of the *growth rate* of the baryon and dark matter perturbations. It is completely analogous to the effect of baryons on dark matter perturbations before recombination. At the time when dark energy is negligible, the neutrino suppression is given by Eq. (6.24), where Ω_B and Ω_{CDM} are replaced by Ω_ν and Ω_M , respectively. With dark energy accounted for, the growth proceeds as follows,

$$\delta_{DB} \propto (ag(\eta))^{1-\frac{3}{5}\frac{\Omega_\nu}{\Omega_M}}, \quad (8.138)$$

where we use the function $g(\eta)$ introduced in Section 4.4; recall that its present value is $g(\eta_0) \approx 0.76$.

Problem 8.20. *Prove the formula (8.138).*

This effect discriminates between the two models at $z \lesssim z_{nr}$, see (8.106). Therefore, it suppresses the density contrast δ_{DB} in the model with massive neutrinos by the factor

$$\left(\frac{a(z_{nr})}{a_0 g(\eta_0)} \right)^{\frac{3}{5}\frac{\Omega_\nu}{\Omega_M}} = \left(\frac{1}{(1+z_{nr}) g(\eta_0)} \right)^{\frac{3}{5}\frac{\Omega_\nu}{\Omega_M}},$$

where we recalled that neutrinos become non-relativistic well before the dark energy becomes relevant, so that $g(\eta_{nr}) = 1$.

The second effect is that neutrinos affect the time of radiation-matter equality. For given Ω_M , its part Ω_{DB} is smaller in the model with massive neutrinos. Neutrinos are relativistic at the equality epoch, hence Ω_{rad} is the same in the two models. Therefore, the equality $\Omega_{DB} = \Omega_{rad}$ occurs in the massive neutrino model at

$$1 + z_{eq} \equiv \frac{a_0}{a_{eq}} = \frac{\Omega_{CDM} + \Omega_B}{\Omega_{rad}} = \frac{\Omega_M - \Omega_\nu}{\Omega_{rad}} = \left(1 - \frac{\Omega_\nu}{\Omega_M} \right) \cdot \frac{\Omega_M}{\Omega_{rad}}.$$

This means that the ratio of the scale factors at equality in the two models is

$$\frac{a_{eq}^{\Omega_\nu \neq 0}}{a_{eq}^{\Omega_\nu = 0}} = \left(1 - \frac{\Omega_\nu}{\Omega_M}\right)^{-1}.$$

Hereafter the notations $\Omega_\nu = 0$ and $\Omega_\nu \neq 0$, refer to the models with massless and massive neutrinos, respectively. Since δ_{DB} grows at matter domination proportionally to the scale factor (modulo the effect we have already accounted for), the delay in the transition to matter domination yields the suppression of δ_{DB} by the factor $(1 - \Omega_\nu/\Omega_M)$. These two effects result in the suppression factor for δ_{DB}

$$\frac{\delta_{DB}^{\Omega_\nu \neq 0}}{\delta_{DB}^{\Omega_\nu = 0}} = \left(1 - \frac{\Omega_\nu}{\Omega_M}\right) \left(\frac{1}{(1 + z_{nr})g(\eta_0)}\right)^{\frac{3}{5}\frac{\Omega_\nu}{\Omega_M}}. \quad (8.139)$$

Finally, massive neutrinos do not contribute to matter perturbations, but do contribute to the total mass density. So, we have for massive model

$$\delta_M^{\Omega_\nu \neq 0} = \delta_{DB}^{\Omega_\nu \neq 0} \frac{\Omega_{DB}}{\Omega_M} = \delta_{DB}^{\Omega_\nu \neq 0} \left(1 - \frac{\Omega_\nu}{\Omega_M}\right). \quad (8.140)$$

This effect is absent in the massless model, where $\delta_{DB}^{\Omega_\nu = 0} = \delta_M$.

Combining the formulas (8.139) and (8.140), we find for the present epoch

$$\frac{\delta_M^{\Omega_\nu \neq 0}}{\delta_M^{\Omega_\nu = 0}} = \left(1 - \frac{\Omega_\nu}{\Omega_M}\right)^2 \left(\frac{1}{(1 + z_{nr})g(\eta_0)}\right)^{\frac{3}{5}\frac{\Omega_\nu}{\Omega_M}}. \quad (8.141)$$

Redshift of the epoch when neutrinos become non-relativistic can be expressed from (8.106) through the neutrino mass,

$$1 + z_{nr} \simeq 1.9 \cdot 10^3 \frac{m_\nu}{1 \text{ eV}}.$$

The neutrino mass, in turn, is given by (8.137). Therefore, the estimate (8.141) gives

$$\begin{aligned} \frac{\delta_M^{\Omega_\nu \neq 0}(a_0)}{\delta_M^{\Omega_\nu = 0}(a_0)} &= \left(1 - \frac{\Omega_\nu}{\Omega_M}\right)^2 \\ &\times \left(4.3 \cdot 10^4 \cdot \left(\frac{h}{0.71}\right)^2 \cdot \left(\frac{2}{N_{\nu,m}}\right) \cdot \Omega_\nu \cdot g(a_0)\right)^{-\frac{3}{5}\frac{\Omega_\nu}{\Omega_M}}. \end{aligned} \quad (8.142)$$

Since Ω_ν/Ω_M is small, we expand (8.142) to the linear order in it. The dependence on other cosmological parameters in weak (logarithmic), and for their realistic values we obtain numerically

$$\frac{\delta_M^{\Omega_\nu \neq 0}(a_0)}{\delta_M^{\Omega_\nu = 0}(a_0)} \simeq 1 - 4 \frac{\Omega_\nu}{\Omega_M} \quad (8.143)$$

(the estimate (8.142) actually gives slightly different numerical coefficient, 4.5 instead of 4; we write here the value obtained by accurate analysis). We see that the correction due to massive neutrino is rather large. For the power spectrum, $P(k) \propto \delta_M^2$, it ranges from tens to a few percent depending on the mass hierarchy. Obviously, the correction is large for heavy degenerate neutrinos with $m_\nu \gg m_{atm}$, and small for the “lightest” neutrino sector with masses m_{atm} , m_{sol} and zero. The power spectra of matter perturbations are shown in Fig. 8.4 for various neutrino masses.

Problem 8.21. Perform the similar analysis for fixed value of Ω_{DB} instead of Ω_M .

The suppression of matter perturbations affects CMB properties and especially structure formation. The power spectrum at the relevant scales is determined from the analysis of galaxy catalogs, see Fig. 8.5, left panel. This analysis rules out the sum of neutrino masses exceeding ~ 1 eV. CMB data alone are not so sensitive, see Fig. 8.5. Combined analysis of several datasets gives the upper bound on the sum of neutrino masses at the level of better than 1 eV, see Fig. 8.5, right panel.

Let us mention that there are several independent methods of studying the structures at short scales. They give upper bounds on the neutrino masses in the ballpark given in (8.104). It will hopefully be possible in future to detect neutrino masses by cosmological methods and even figure out their hierarchy pattern. The latter will require measuring both Ω_ν and k_{nr} . Discriminating between the model with relatively heavy degenerate neutrinos and models with light neutrinos will be possible in near future. The prospects to detect the effect of neutrino with $m_\nu = m_{atm} \simeq 0.05$ eV and find out whether the hierarchy is direct or inverse are less

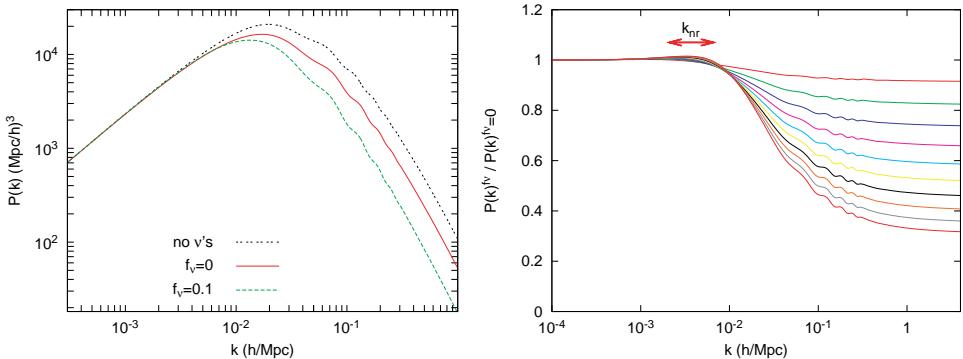


Fig. 8.4 Suppression of the present power spectrum of matter perturbations due to neutrinos [50]. The cosmological parameters are $h = 0.7$, $\Omega_\Lambda = 0.7$, $\Omega_M = 0.3$, $n_s = 1$ (flat primordial power spectrum). Left panel: power spectra $P(k) \propto \delta_M^2$ in models without neutrinos, (“no ν ’s”), with three massless neutrinos ($f_\nu \equiv \Omega_\nu/\Omega_M = 0$) and with three degenerate massive neutrinos and $\Omega_\nu = 0.1\Omega_M$ ($f_\nu = 0.1$). The primordial spectra are the same in the three models. Right panel: ratios of power spectra in models with massive neutrino ($P(k)^{f_\nu}$) to that in the model with three massless neutrinos ($P(k)^{f_\nu=0}$); the lines correspond to $f_\nu = 0.01$ (upper curve), $0.02, \dots, 0.1$ (lower curve).

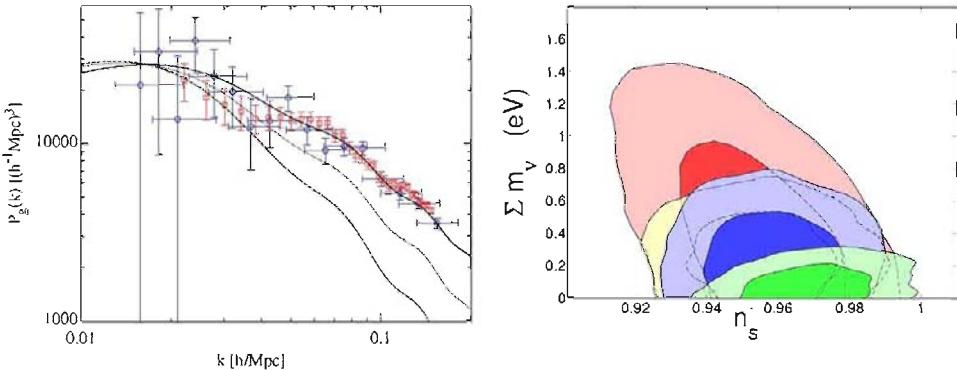


Fig. 8.5 *Left:* observational determination of the power spectrum of matter perturbations based on SDSS (circles) and 2dFGRS (squares) catalogs and its comparison with models with massive neutrinos [51]. Solid line is the prediction of the model with the sum of neutrino masses $\sum m_\nu = 0.28 \text{ eV}$, short and long dashed lines correspond to models with $\sum m_\nu = 1.5 \text{ eV}$ and $\sum m_\nu = 3 \text{ eV}$, respectively. Values of other cosmological parameters are similar to those used in this book. The spectra are normalized at momentum $k = 0.017 \text{ h/Mpc}$. It is clear that predictions of models with heavy neutrinos contradict observational data at short scales. *Right:* cosmological bounds (light and dark shades refer to 68% and 95% C.L., respectively) on the sum of neutrino masses and the scalar tilt from WMAP CMB data (largest contours) and their combination with other cosmological observations (see Ref. [52] for details): distribution of Large Red Galaxies (smaller region), observations of SNe Ia and baryon acoustic oscillations (even smaller region), distribution of Large Red Galaxies plus observations of SNe Ia and baryon acoustic oscillations plus measurement of the present value of the Hubble parameter (the smallest region).

clear, although there are projects for studying clusters of galaxies with the precision sufficient to explore the region $\sum m_\nu \sim 0.03 \text{ eV}$.

Let us note that the anisotropic stress tensor of the neutrino perturbations is a source of gravity waves, since it contributes to the right hand side of the tensor equation (B.15). Unfortunately, this source is very weak, and the gravity waves produced by neutrino perturbations cannot be discovered in near future.

Problem 8.22. Estimate the amplitude of gravity waves produced by neutrino perturbations.

Let us consider the effect of the neutrino component on the evolution of gravity waves in the Universe. We are not interested in the *generation* of gravity waves by the neutrino perturbations: we have noticed above that it is small. What we are going to study are relic gravity waves that could be produced, e.g., at inflation, see Section 13.3. Neutrinos affect them via their anisotropic stress tensor [54, 55]; note that similar effect is absent for baryon-photon component before recombination. The anisotropic stress tensor rapidly decreases after neutrino becomes non-relativistic, so we consider the case of relativistic neutrino and set neutrino masses equal to zero. We limit ourselves to the gravity waves that enter the horizon at radiation domination but after neutrino decoupling.

Since we study tensor perturbations here, we write the metric as follows,

$$g_{00} = a^2, \quad g_{ij} = -a^2 (\delta_{ij} - h_{ij}), \quad (8.144)$$

where $h_{ij} = h_{ij}^{T\prime T}$ is transverse traceless 3-tensor. We wish to calculate the neutrino contribution to the energy-momentum tensor (8.10). The analysis can be performed in analogy

to Section 8.2. However, we present an alternative, equivalent way of treating the cosmological perturbations [55].

Let us introduce unit 3-vector n_i in the direction opposite to the covariant vector P_i and denote the absolute value of the latter by κ , so that $P_i \equiv -n_i\kappa$. Then the contravariant neutrino momentum is, to the linear order in metric perturbation,

$$P^i = g^{ij} P_j = \frac{\kappa}{a^2} (n_i + n_j h_{ij}), \quad P^0 = \sqrt{-\frac{g^{ij}}{a^2} P_i P_j} = \frac{\kappa}{a^2} \left(1 + \frac{1}{2} n_i n_j h_{ij}\right).$$

The neutrino distribution function $f_\nu = f_\nu(\eta, x^i, P_j)$ obeys the Vlasov equation (Boltzmann equation (8.37) with vanishing collision integral). The combination entering there is

$$\frac{P^i}{P^0} = n_i + n_j h_{ij} - \frac{1}{2} n_i n_k n_l h_{kl}.$$

Inserting the explicit expressions for the Christoffel symbols, and having in mind linearized analysis, we write the Vlasov equation in the following form,

$$f'_\nu + \left(n_i + n_j h_{ij} - \frac{1}{2} n_i n_m n_l h_{ml}\right) \frac{\partial f_\nu}{\partial x^i} + \frac{1}{2} \kappa n_m n_i \partial_j h_{im} \frac{\partial f_\nu}{\partial P_j} = 0. \quad (8.145)$$

Non-vanishing anisotropic stress tensor emerges after neutrino decoupling. At the time of decoupling η_d , neutrino distribution function is the Fermi–Dirac distribution f_0 whose argument is the momentum $\sqrt{-g^{ij}(\eta_d, \mathbf{x}) P_i P_j}$,

$$f_\nu(\eta_d, x^i, P_j) \equiv f_0(x^i, P_j).$$

Hence, the distribution function at later times is naturally written as

$$f_\nu = f_0(x^i, P_j) + \delta f(\eta, x^i, P_j). \quad (8.146)$$

The initial condition at $\eta = \eta_d$ is, by definition,

$$\delta f(\eta_d, x^i, P_j) = 0. \quad (8.147)$$

In this formalism, metric perturbations affect both f_0 (through $g^{ij}(\eta_d, \mathbf{x})$) and induced perturbation δf . The latter is found by solving the Vlasov equation. Namely, we insert the expression (8.146) into Eq. (8.145) and perform the Fourier transformation in spatial coordinates. In this way we obtain the equation for δf induced by gravity waves,

$$\delta f' + i\mathbf{k}\mathbf{n}\delta f - \frac{i}{2} \kappa \frac{\partial f_0}{\partial \kappa} \mathbf{k}\mathbf{n} n_j n_m \cdot (h_{jm}(\eta) - h_{jm}(\eta_d)) = 0. \quad (8.148)$$

The term proportional to $h_{jk}(\eta_d)$ ensures the validity of the initial condition (8.147). It results from the spatial derivative of f_0 entering Eq. (8.145). The solution to Eq. (8.148) is

$$\delta f = \frac{i}{2} \frac{\partial f_0}{\partial \kappa} \mathbf{n}\mathbf{k} n_i n_j \int_{\eta_d}^\eta d\eta' \cdot (h_{ij}(\eta') - h_{ij}(\eta_d)) e^{i\mathbf{n}\mathbf{k} \cdot (\eta' - \eta)}.$$

We see that gravity waves indeed induce the *additional contribution* to the perturbation in the neutrino component. Similar effect exists for the photon component after recombination.

Now that we know the distribution function $f_\nu = f_0 + \delta f$, we insert it into the spatial components of the energy-momentum tensor (8.10) and extract the traceless part (see (B.13b)),

$$\Pi^i_j = -T^i_j + \frac{1}{3} \delta^i_j T^l_m \delta^m_l = \int \frac{d^3 P_m}{\sqrt{-g}} \frac{-P^i P_j + \frac{1}{3} \delta^i_j P_l P^l}{P^0} f_\nu(\eta, x^i, P_j). \quad (8.149)$$

We need the term linear in metric perturbation. It comes from both the perturbation δf and the Fermi-Dirac function f_0 , since the latter is a function of $\sqrt{-g^{ij}}(\eta_d, \mathbf{x}) P_i P_j$. Other quantities in (8.149) that involve metric perturbations are P^i and P^0 . With our parameterization of the neutrino momentum, the integration measure is independent of h_{ij} to the linear order,

$$\int \frac{d^3 P_m}{\sqrt{-g}} = \frac{1}{a^4} \int \kappa^2 d\kappa d\mathbf{n},$$

while the linearized expression for the integrand in (8.149) is

$$\left(n_i n_j - \frac{1}{3} \delta_{ij} \right) \left(\delta f + \kappa \frac{\partial f_0}{\partial \kappa} \frac{1}{2} n_l n_m h_{lm}(\eta_d) \right) \quad (8.150a)$$

$$+ \left(n_j n_k h_{ik}(\eta) - \frac{1}{2} \left(n_i n_j + \frac{1}{3} \delta_{ij} \right) n_l n_m h_{lm}(\eta) \right) f_0. \quad (8.150b)$$

Let us begin with the contribution due to the perturbation δf , the first term in (8.150a),

$$\begin{aligned} \Pi_{ij}^{(1)} &= \frac{i}{2} \frac{1}{2a^4} \int_0^\infty d\kappa \cdot \kappa^3 \cdot \kappa \frac{\partial f_0}{\partial \kappa} \int d\mathbf{n} \left(n_i n_j - \frac{1}{3} \delta_{ij} \right) n_l n_m (i\mathbf{n}\mathbf{k}) \\ &\times \int_{\eta_d}^\eta d\eta' (h_{lm}(\eta') - h_{lm}(\eta_d)) \cdot e^{i\mathbf{n}\mathbf{k}(\eta' - \eta)}. \end{aligned}$$

The integral over conformal time is conveniently rewritten via integrating by parts,

$$\begin{aligned} i\mathbf{n}\mathbf{k} \int_{\eta_d}^\eta d\eta' e^{i\mathbf{n}\mathbf{k}(\eta' - \eta)} (h_{lm}(\eta') - h_{lm}(\eta_d)) \\ = \int_{\eta_*}^\eta d\eta' \cdot (h_{lm}(\eta') - h_{lm}(\eta_d)) \frac{d}{d\eta'} e^{i\mathbf{n}\mathbf{k}(\eta' - \eta)} \\ = - \int_{\eta_d}^\eta d\eta' e^{i\mathbf{n}\mathbf{k}(\eta' - \eta)} h'_{lm}(\eta') + h_{lm}(\eta) - h_{lm}(\eta_d). \end{aligned} \quad (8.151)$$

We see that the last term here cancels out the contribution to (8.150a), proportional to $\partial f_0 / \partial \kappa$. We integrate the remaining term by parts in the variable κ ,

$$\int d\kappa \cdot \kappa^4 \cdot \frac{\partial f_0}{\partial \kappa} = -4 \int d\kappa \cdot \kappa^3 \cdot f_0 = -\frac{1}{\pi} \rho_\nu.$$

We now consider the second term in (8.151). It cancels out similar contribution coming from (8.150b) upon angular integration,

$$\begin{aligned} -\frac{\rho_\nu}{2\pi a^4} \int d\mathbf{n} h_{lm} n_l n_m \cdot \left(n_i n_j - \frac{1}{3} \delta_{ij} \right) \\ + \frac{\rho_\nu}{4\pi a^4} \int d\mathbf{n} \left(n_j n_l h_{il} - \frac{1}{2} \left(n_i n_j + \frac{1}{3} \delta_{ij} \right) h_{lm} n_l n_m \right) = 0. \end{aligned}$$

We perform the integration over azimuthal angle in the remaining integral term in (8.151) and introduce the integration variable $\xi = \cos \theta$ where θ is zenith angle. This gives the final expression for the transverse anisotropic stress,

$$\Pi_j^{iTT} = -\frac{\rho_\nu}{4a^4} \int_{-1}^1 d\xi (1 - \xi^2)^2 \int_{\eta_d}^\eta d\eta' e^{i\xi k(\eta - \eta')} h'_{ij}(\eta'). \quad (8.152)$$

It is this expression that enters the right hand side of Eq. (B.15).

Problem 8.23. Obtain the result (8.152) by making use of the standard approach of Section 8.2.

The integral over ξ is equal to

$$S(y) \equiv \frac{1}{16} \int_{-1}^1 d\xi (1 - \xi^2)^2 e^{i\xi y} = -\frac{1}{y^2} \left(\frac{\sin y}{y} + \frac{3 \cos y}{y^2} - \frac{3 \sin y}{y^3} \right).$$

So the equation for gravitational waves, with neutrino back reaction included, is

$$\frac{d^2 h_{ij}}{dy^2} + \frac{2}{a} \frac{da}{dy} \frac{dh_{ij}}{dy} + h_{ij} = 24 R_\nu(y) \left(\frac{1}{a} \frac{da}{dy} \right)^2 \int_{y_d}^y dy' \cdot S(y - y') \frac{dh_{ij}(y')}{dy'}, \quad (8.153)$$

where we introduced the variable $y = k\eta$ (and $y_d \equiv k\eta_d$), used the Friedmann equation at radiation domination and recalled the definition of the neutrino fraction in the energy density, $R_\nu = \rho_\nu(y)/\rho_{rad}(y)$. It follows from (8.153) that neutrino back reaction suppresses gravity wave amplitudes. The integral form of this effect is due to the fact that the neutrino anisotropic stress is generated prior to the current moment of time η . Since the tensor perturbation stays constant in time before the horizon entry, the initial condition for Eq. (8.153) is

$$\frac{dh_{ij}(y_d)}{dy} = 0. \quad (8.154)$$

We note that Eq. (8.153) does not mix polarizations of the tensor modes.

At radiation domination and after neutrino decoupling we have $a(\eta) \propto \eta$, $\eta_{eq} \gg \eta \gg \eta_d$. Hence, Eq. (8.153) is, explicitly,

$$\frac{d^2 h_{ij}}{dy^2} + \frac{2}{y} \frac{dh_{ij}}{dy} + h_{ij} = 24 \frac{R_\nu}{y^2} \int_{y_d}^y dy' \cdot S(y - y') \frac{dh_{ij}(y')}{dy'}. \quad (8.155)$$

The lower limit of integration in the right hand side can be extended to zero, since the initial conditions are imposed deep in the superhorizon regime.

The numerical solution to Eq. (8.155) has the following properties [55]. Before the horizon entry, the mode is given by

$$h_{jk} = h_{jk(i)} \cdot \frac{\sin y}{y},$$

i.e., the right hand side of Eq. (8.155) is irrelevant. At large y , i.e., deep inside the horizon, the right hand side of Eq. (8.155) is again negligible, and

$$h_{jk} \longrightarrow h_{jk(i)} \cdot A \cdot \frac{\sin(y + \alpha)}{y}.$$

The suppression factor A and phase shift α are determined by dynamics near the horizon crossing, i.e., at $y \sim 1$. One has, approximately [56],

$$\alpha \simeq 0, \quad A \simeq 1 - \frac{5}{9} R_{\nu,0} \simeq 0.8.$$

Thus, the amplitude of gravity waves entering the horizon at radiation domination is suppressed by about 20% due to neutrino. This yields the suppression of CMB observables (temperature anisotropy and polarization) at the level of 36%. This is valid for the present wavelengths in the interval

$$1 \text{ kpc} \sim 2\pi\eta_d \lesssim \lambda \lesssim 2\pi\eta_{eq} \sim 800 \text{ Mpc}.$$

It corresponds to CMB harmonics $l \gtrsim 100$. Gravity waves entering the horizon at matter domination are suppressed too, but the suppression factor is smaller, less than 10%.

Problem 8.24. Find the equation for gravity waves at matter domination, with neutrino back reaction included. Find the solution for h_{ij} in the small wavelength limit.

We note that similar effect exists for photons. The suppression of gravity waves due to photons occurs after recombination only, so this effect is small because of the small photon fraction in the energy density. Furthermore, the effect of photon suppression is potentially interesting in a narrow interval of wavelengths around $\lambda = 2\pi a_0/k \sim 1$ Gpc: these are the modes that enter the horizon soon after recombination, when the photon fraction in the energy density is still sizeable. The amplitude of these modes gets suppressed by about 5%, so this effect is small for CMB observables.

8.5 Photons and Baryons at Recombination Epoch

In this Section we consider the suppression of perturbations in the baryon-photon plasma at recombination. This suppression is additional to that discussed in Section 6.2 within the tight coupling approximation. The phenomena studied here are important, in particular, from the viewpoint of CMB temperature anisotropy and polarization, as well as baryon acoustic oscillations. It is this viewpoint that we have in mind throughout this Section.

As we have discussed in Chapter I.6, photons interact mostly with electrons at the epoch we consider. The Compton scattering off protons is suppressed by the ratio of the electron and proton masses squared, $m_e^2/m_p^2 \sim 10^{-6}$, and hence it is irrelevant. On the other hand, electrons and protons are tightly coupled due to the Coulomb interaction, so they behave as single medium.

In the first place, photons experience free streaming similar to that discussed in Section 8.4. The relevant length scale here is the photon mean free path λ_γ . Below this scale, photons behave in the same way as massless neutrinos studied in Section 8.4. We note that the free streaming effect exists for the electron-baryon component as well. Clearly, the latter effect occurs at very short scales, since tightly coupled baryons and electrons are non-relativistic, see Problem 8.9. We do not consider baryon-electron free streaming in what follows.

The photon mean free path at time η is estimated as follows,

$$\lambda_\gamma \sim \frac{1}{\sigma_T n_e(\eta)},$$

where

$$\sigma_T = 0.67 \cdot 10^{-24} \text{ cm}^2 \quad (8.156)$$

is the Thomson cross section of photon scattering off free electron, n_e is the free electron number density. The number densities of free electrons and protons coincide, and just before recombination the latter is about 0.75 of the total baryon number density; the rest of baryons (and electrons) reside in helium atoms which are produced some time before recombination of hydrogen (see

Chapters I.6 and I.8). Thus, just before recombination, the number density of free electrons is

$$n_e(z > z_r) = 0.75 \frac{\rho_B(z)}{m_p} = 8.4 \cdot 10^{-6} \Omega_B h^2 \cdot (1 + z_r)^3 \text{ cm}^{-3} \simeq 260 \text{ cm}^{-3}. \quad (8.157)$$

This gives for the photon mean free path just before recombination

$$\lambda_\gamma(z > z_r) \sim 6 \cdot 10^4 \cdot \frac{1}{\Omega_B h^2 (1 + z_r)^3} \text{ Mpc.}$$

Hence, the present wavelength of perturbations which are damped due to free streaming by recombination is estimated as

$$\lambda_0 \lesssim (1 + z_r) \cdot \lambda_\gamma(z > z_r) \simeq 2.1 \text{ Mpc.} \quad (8.158)$$

We see that free streaming is relevant for rather short wavelengths.

There exists another effect which is of importance for baryon-photon component. It is called the Silk damping [58]. We consider it in Section 8.5.2.

8.5.1 *Thickness of last scattering sphere*

Before turning to the Silk damping, let us discuss another effect that occurs at recombination and also leads to the suppression of CMB temperature anisotropy at small angular scales. At the same time, this effect is crucial for CMB polarization. It is due to the finiteness of the time interval during which photons decouple: as we have seen in Chapter I.6, recombination proceeds in short but finite time. In other words, there is the time-dependent optical depth at last scattering. The general definition of the optical depth is

$$\tau(t_1, t_2) = \int_{t_2}^{t_1} \frac{dt}{\lambda_\gamma(t)}.$$

In our case, the equivalent expression is given by

$$\tau(\eta_1, \eta_2) = \int_{\eta_2}^{\eta_1} d\eta \cdot a(\eta) \cdot \sigma_T \cdot n_e(\eta). \quad (8.159)$$

The function $e^{-\tau(\eta_1, \eta_2)}$ is the probability that a photon have not scattered off an electron in the time interval (η_2, η_1) . Accordingly, the fraction of photons in the total photon flux at time η_1 that scattered last time in the time interval $(\eta_2, \eta_2 + d\eta_2)$ is equal to $V(\eta_1, \eta_2)d\eta_2$ where

$$V(\eta_1, \eta_2) = e^{-\tau(\eta_1, \eta_2)} \cdot \frac{d\tau(\eta_2)}{d\eta_2} \quad (8.160)$$

is the visibility function. From the viewpoint of CMB, the relevant choice is $\eta_1 = \eta_0$, since all CMB photons are detected today. It follows from (8.160) that

$$\int_0^{\eta_0} d\eta V(\eta_0, \eta) = 1,$$

This should be the case, since the total probability that a CMB photon scatters before coming to an observer equals one. The function $V(\eta_0, \eta)$ has a maximum at the last scattering epoch at the moment of time $\eta = \eta_r$, such that $\tau''(\eta_r) = \tau'^2(\eta_r)$. Clearly, this moment is after the beginning of recombination but before complete photon decoupling: early on, the optical depth $\tau(\eta_0, \eta)$ is large (photons intensely scatter off electrons) and the visibility function is exponentially small, while at late times τ' is close to zero (photons do not scatter), and $V(\eta_0, \eta)$ is again small. The time η_r is naturally called the most probable time of last scattering, or simply the time of last scattering. We use this definition throughout this book. Note that this definition is similar, though not exactly equivalent, to the definition given in Section I.6.2.

As we have seen in Section I.6.2, the electron number density decreases in time at the last scattering epoch as follows,

$$n_e = F(T) \cdot e^{-\frac{\Delta_H}{4T}}, \quad (8.161)$$

where $F(T)$ is a slowly varying function of temperature, and $\Delta_H = 13.6 \text{ eV}$ is the binding energy of hydrogen atom in the ground state.⁸ The relevant optical depth is

$$\tau(\eta_0, \eta) = \int_{\eta}^{\eta_0} d\eta \, a n_e \sigma_T = \int_t^{t_0} dt \, n_e \sigma_T.$$

This is the integral of the exponentially decreasing function; it is saturated at the lower limit of integration, and performing the change of integration variable from t to $1/T$ we obtain

$$\tau(\eta_0, \eta) = n_e(\eta) \cdot \sigma_T \cdot \frac{4T}{\Delta_H H}.$$

Hence, the visibility function is

$$\begin{aligned} V(\eta_0, \eta) &= a n_e(\eta) \cdot \sigma_T \cdot \exp\left(-n_e(\eta) \cdot \sigma_T \cdot \frac{4T}{\Delta_H H}\right) \\ &= a \frac{\Delta_H H}{4T} \exp(-y + \log y), \end{aligned} \quad (8.162)$$

where

$$y = n_e(\eta) \cdot \sigma_T \cdot \frac{4T}{\Delta_H H}.$$

The function (8.162) has a maximum at $y = y_r = 1$, or, in terms of the electron number density,

$$n_e(\eta_r) = \frac{\Delta_H H}{4T \sigma_T}. \quad (8.163)$$

⁸Recall that this behavior is due to the fact that recombination at its late stage proceeds *out of thermal equilibrium*. In thermal equilibrium, the electron number density would decrease faster, $n_e^{eq} \propto \exp(-\Delta_H/2T)$. This peculiarity is important, in particular, for CMB polarization.

The Hubble parameter at recombination is

$$H(T_r) = 100 \frac{\text{km}}{\text{s} \cdot \text{Mpc}} \cdot h \sqrt{\Omega_M(1+z_r)^3 + \Omega_{rad}(1+z_r)^4} \simeq 5.2 \text{ Mpc}^{-1}, \quad (8.164)$$

This gives numerically

$$n_e(\eta_r) \simeq 33 \text{ cm}^{-3}. \quad (8.165)$$

This is considerably smaller than the number density just before recombination, see (8.157). As we noticed in Section I.6.2, photon last scattering occurs when the electron number density has decreased considerably due to recombination.

According to (8.162), the visibility function has a sharp maximum. We approximate V near η_r by the Gaussian function, and obtain in terms of redshift

$$V(0, z) = V(0, z_r) e^{-\frac{(z-z_r)^2}{2\Delta z^2}},$$

where

$$\frac{1}{\Delta z} = \left| \frac{\partial y}{\partial z} \right|_{y=y_r=1} = \frac{\Delta_H}{4T_r z_r}$$

Hence, the half-width of the visibility function is

$$\Delta z = \frac{4T_r}{\Delta_H} \cdot z_r \simeq 84,$$

or in terms of conformal time

$$\frac{\Delta \eta_r}{\eta_r} = \frac{2T_r}{\Delta_H} \simeq 0.04 \quad (8.166)$$

(recombination occurs at matter domination, when $z \propto a^{-1} \propto \eta^{-2}$). The half-time is obtained by using $|\Delta z/z| = H\Delta t$,

$$\Delta t = \frac{4T_r}{\Delta_H H(t_r)} \simeq 0.08 H^{-1}(t_r).$$

This gives the present value of the total width of the last scattering sphere, $2\Delta t(1+z_r) \simeq 32 \text{ Mpc}$. As we show in Section 9.3, the fact that this width is finite leads to wash-out of CMB temperature anisotropy at small angular scales, $\Delta\theta \lesssim \pi\Delta\eta_r/\eta_0$. The smallness of this width is one of the reasons for the pronounced structure of the CMB temperature angular spectrum. On the other hand, fairly long mean free path of photon at last scattering, $\lambda_\gamma(t_r) = [\sigma_T n_e(t_r)]^{-1} \simeq \Delta t$, is responsible for the entire effect of CMB polarization at all but the largest angular scales, see Chapter 10.

Numerical solution to the Boltzmann equation for the system of photons, electrons, protons and hydrogen atoms gives the visibility function shown in Fig. 8.6. The visibility function can be approximated by asymmetric Gaussian function

$$V(\eta) = V(\eta_r) \cdot e^{-\frac{(\eta-\eta_r)^2}{2\Delta\eta^2}}. \quad (8.167)$$

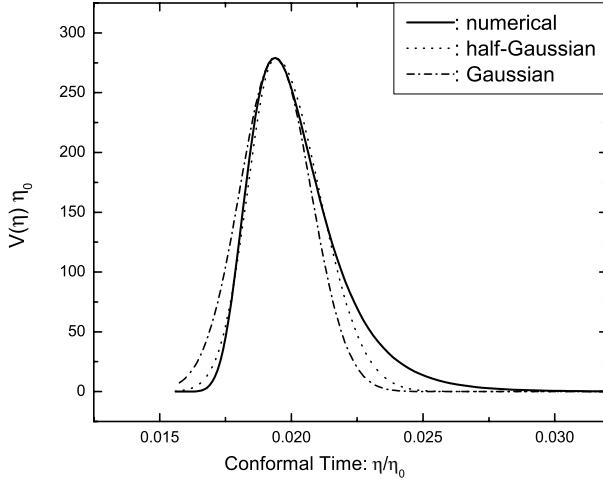


Fig. 8.6 The visibility function at the photon last scattering epoch [57].

For $\Omega_\Lambda = 0.73$, $\Omega_B = 0.044$ the maximum is at $V(\eta_r) \cdot \eta_0 = 279$, and the width is

$$\Delta\eta = \Delta\eta_1 \cdot \theta(\eta_r - \eta) + \Delta\eta_2 \cdot \theta(\eta - \eta_r), \quad \Delta\eta_1 = 0.00110\eta_0, \quad \Delta\eta_2 = 0.00176\eta_0,$$

where $\theta(x)$ is step function. Our estimate (8.166) is in reasonable agreement with the numerical result, in view of the numerical value $\eta_0/\eta_r \simeq 50$, see Section 2.1.2.

8.5.2 Silk damping

We now turn to the Silk damping [58]. It occurs because the energy transfer in a photon-electron collision is small (see Chapter I.6). Hence, photon collides with electrons many times when traveling the distance larger than the photon mean free path, but its energy remains almost the same. Similarly to free streaming, this effect smoothes out temperature inhomogeneities. We emphasize that the Silk damping is absent for dark matter. Therefore, it does not destroy the gravitational potentials (6.3), (6.15) produced by dark matter. Neither it affects the non-oscillating contributions to the photon density perturbations (first terms in (6.39), (6.44)), since they are generated by the dark matter gravitational potentials. The Silk damping suppresses only *oscillations* in δ_γ (second terms in (6.39), (6.44)).

For preliminary estimate of the Silk scale we consider a photon that randomly changes its direction of motion but does not change its energy in the processes of scattering off electrons. Then the relevant length scale is the photon diffusion length in the Hubble time. In this random walk approximation, the Silk scale λ_S is estimated as the geometrical mean of the photon mean free path λ_γ and the Hubble length l_H (or Hubble time $t_H \sim H^{-1}$): the number of collisions in the Hubble time is of order t_H/λ_γ , while the distance between collision points is λ_γ , hence the diffusion

length is

$$\lambda_S \sim \lambda_\gamma \sqrt{\frac{t_H}{\lambda_\gamma}} = \sqrt{\lambda_\gamma \cdot l_H}.$$

This gives for the Silk length

$$\lambda_S(\eta) \sim \sqrt{\frac{1}{\sigma_T \cdot n_e(\eta) \cdot H(\eta)}}. \quad (8.168)$$

Making use of (8.157) and (8.164) we find that the Silk length just before recombination is of order $\lambda_S(\eta < \eta_r) \sim 0.02$ Mpc, and the corresponding present length is estimated as

$$(1 + z_r)\lambda_S(\eta < \eta_r) \sim 21 \text{ Mpc}. \quad (8.169)$$

We will see later in this Section that this underestimates the Silk length just before recombination by about a factor of 1.5.

In fact, the Silk damping operates up to the time of photon last scattering $\eta \approx \eta_r$. The number density of free electrons decreases considerably by that time, and despite the short duration of the last scattering epoch, the Silk length increases somewhat. It is difficult to calculate the latter effect analytically; in what follows, we use the estimate

$$\lambda_{S,0} \simeq 50 \text{ Mpc}. \quad (8.170)$$

We support this estimate later on, see (8.189). The present value of the Silk momentum is

$$\frac{k_S}{a_0} = \frac{2\pi}{\lambda_{S,0}} \simeq 0.1 \text{ Mpc}^{-1}. \quad (8.171)$$

Hence, at photon last scattering, the oscillatory part of photon density perturbations is damped for present wavelengths smaller than $\lambda_{S,0}$.

The accurate treatment of the Silk damping is made by using the Boltzmann equation together with the Einstein equations. An important effect absent, e.g., in the case of massless neutrinos studied in Section 8.4, is the interaction of photons and electrons. It results in the non-vanishing collision integral (8.34) in the right hand side of the Boltzmann equation. To the leading order in the electromagnetic coupling α , the scattering amplitude squared for unpolarized photon⁹ has the following form in electron rest frame (see Ref. [31] and Eq. (8.25)),

$$|\mathcal{M}|^2 = \frac{\alpha^2}{2} \left(\frac{\omega'_{rf}}{\omega_{rf}} + \frac{\omega_{rf}}{\omega'_{rf}} - \sin^2 \theta_{rf} \right) \delta(P_{\gamma,rf} + P_{e,rf} - P'_{\gamma,rf} - P'_{e,rf}), \quad (8.172)$$

⁹Generally speaking, polarization effects are considerable, see also Chapter 10. They are not, however, dramatically important for the Silk damping. Some numerical factors get modified due to polarization, though; we quote the result in the appropriate place.

where θ_{rf} is photon scattering angle ω_{rf} and ω'_{rf} are frequencies of incoming and outgoing photon, respectively; other notations are standard. In the cosmic frame, the incident electron and photon have 3-momenta \mathbf{P}_e and \mathbf{P}_γ , respectively. The formulas for this frame are obtained by applying the Lorentz boost with the γ -factor $\gamma = (1 - \mathbf{P}_e^2/m_e^2)^{-1/2}$. The photon frequencies in the rest frame and cosmic frame are related as follows,

$$\omega_{rf} = \left(\omega - \mathbf{P}_\gamma \frac{\mathbf{P}_e}{m_e} \right) \gamma. \quad (8.173)$$

Electrons are non-relativistic, and they can be treated as ideal fluid with the number density n_e and velocity \mathbf{v}_e . These are given in terms of the electron distribution function as follows,

$$n_e = \int d^3 P_e f_e(\mathbf{P}_e), \quad \mathbf{v}_e = \frac{1}{m_e n_e} \int d^3 P_e \mathbf{P}_e f_e(\mathbf{P}_e).$$

We are interested in the collision integral to the zeroth and first orders in the velocity \mathbf{v}_e . The zeroth order and linear parts of the matrix element are obtained by inserting (8.173) into (8.172),

$$|\mathcal{M}|^2 \propto 1 + \cos^2 \theta - 2(1 - \cos \theta) \cos \theta \cdot \left(\frac{\mathbf{P}_\gamma}{\omega} + \frac{\mathbf{P}'_\gamma}{\omega'} \right) \mathbf{v}_e. \quad (8.174)$$

Problem 8.25. Obtain the formula (8.174).

Other terms in the collision integral (8.34) are also expanded in \mathbf{v}_e to the linear order. In particular,

$$\frac{1}{P_e^0 \cdot P_e'^0} = \frac{1}{m_e^2} \left(1 - \frac{(\mathbf{P}_\gamma - \mathbf{P}'_\gamma) \mathbf{v}_e}{m_e} - \frac{(\mathbf{P}_\gamma - \mathbf{P}'_\gamma)^2}{2m_e^2} \right). \quad (8.175)$$

We now make use of (8.174) and (8.175) in the collision integral (8.34) and perform the integration over the initial and final spatial momenta of electron. Half of these integrations is trivial because of the δ -function in (8.172), and the other half yields the expressions that reduce to n_e and v_e . As a result, the collision integral to the zeroth and first orders in the velocity of the baryon-electron fluid is given by

$$\begin{aligned} C[f_\gamma] &= \frac{3}{4} (1 + \Phi) a n_e \sigma_T \cdot \int \frac{d\Omega'}{4\pi} \int \frac{\omega' d\omega'}{\omega} \\ &\times \left[\left(1 + \cos^2 \theta + 2(1 - \cos \theta) \cos \theta \cdot \left(\frac{\mathbf{v}_e \mathbf{P}_\gamma}{\omega} + \frac{\mathbf{v}_e \mathbf{P}'_\gamma}{\omega'} \right) \right) \delta(\omega - \omega') \right. \\ &\left. + (1 + \cos^2 \theta) (\mathbf{v}_e \mathbf{P}'_\gamma - \mathbf{v}_e \mathbf{P}_\gamma) \frac{\partial \delta(\omega - \omega')}{\partial \omega'} \right] [f_\gamma(x, \mathbf{P}'_\gamma) - f_\gamma(x, \mathbf{P}_\gamma)]. \end{aligned}$$

We see that the equality between the matrix elements squared of the direct and inverse processes leads to the dependence of the result on *the difference* of the distribution functions at photon momenta before and after scattering: both inflow and outflow of photons are accounted for.

Upon integration over the frequency ω' we obtain, up to the linear order in the electron velocity,

$$C[f_\gamma] = -(1 + \Phi) \tau' f_\gamma - \tau' \mathbf{n} \mathbf{v}_e \kappa \frac{\partial f_\gamma}{\partial \kappa} \quad (8.176a)$$

$$+ \frac{3}{16\pi} (1 + \Phi) \tau' \int d\mathbf{n}' f_\gamma(\kappa, \mathbf{n}') \left(1 + (\mathbf{n}\mathbf{n}')^2\right). \quad (8.176b)$$

We replaced the expression $a n_e \sigma_T$ by τ' (see (8.159)) for brevity, and introduced conformal photon frequency κ . The two terms (8.176a) account for outflow of photons from a given region of the phase space due to collisions with electrons, while the term (8.176b) describes inflow of photons from other regions. We see that to the linear order, the change of photon frequency is negligible: scattering off electrons changes only the direction of photon propagation. We decompose, according to (8.52), the photon distribution function into the unperturbed, homogeneous part $f_\gamma^{(0)}$ and perturbation δf_γ and obtain from (8.56) the linearized Boltzmann equation,

$$\begin{aligned} \delta f'_\gamma + \mathbf{n} \frac{\partial \delta f_\gamma}{\partial \mathbf{x}} - \left(\Psi' + \mathbf{n} \frac{\partial \Phi}{\partial \mathbf{x}} \right) \cdot \frac{\partial \log f_\gamma^{(0)}}{\partial \log \kappa} &= -\tau' \delta f_\gamma - \tau' \mathbf{n} \mathbf{v}_e \kappa \frac{\partial \delta f_\gamma}{\partial \kappa} \\ - \tau' \mathbf{n} \mathbf{v}_e \delta f_\gamma \kappa \frac{\partial \log f_\gamma^{(0)}}{\partial \kappa} + \frac{3}{16\pi} \tau' \int d\mathbf{n}' \delta f_\gamma(\kappa, \mathbf{n}') \left(1 + (\mathbf{n}\mathbf{n}')^2\right). \end{aligned} \quad (8.177)$$

Problem 8.26. Find the explicit form of the Boltzmann equation for photons in the absence of the scalar gravitational potentials but in the presence of tensor perturbations of the metric.

To describe dynamics of photons, we use the continuity and Euler equations modified due to the Compton scattering. The formal trick for obtaining these equations is the multipole decomposition of the Boltzmann equation (8.177). This decomposition has been described in details in Section 8.4.1. Hence, we simply give the result for photons,

$$\delta'_\gamma - \frac{4k^2}{3} v_\gamma + 4\Psi' = 0, \quad (8.178a)$$

$$v'_\gamma + \frac{1}{4} \delta_\gamma + \pi_\gamma + \Phi = -(v_\gamma - v_B) \frac{\tau'}{k}, \quad (8.178b)$$

$$\pi'_\gamma - \frac{4}{15} k^2 v_\gamma + \frac{3}{10} k F_{\gamma,3} = -\frac{9}{10} \tau' \pi_\gamma, \quad (8.178c)$$

$$F'_{\gamma,l} - \frac{l}{2l+1} F_{\gamma,l-1} + \frac{l+1}{2l+1} F_{\gamma,l+1} = -F_{\gamma,l} \frac{\tau'}{k}, \quad l = 3, 4, \dots, \quad (8.178d)$$

where we used the fact that electrons are tightly coupled to baryons, i.e., $v_e = v_B$. Solving the system (8.178) together with the similar system for baryons enables one to find the behavior of perturbations at any moment of time.

Problem 8.27. Derive the system of equations (8.178) from the Boltzmann equation (8.177).

The terms in the right hand side of (8.178) are due to photon interaction with electrons. It is instructive to compare Eqs. (8.178a) and (8.178b) to Eqs. (5.6) and (5.7) with $u_s^2 = w = 1/3$ for photons. We see that the Euler equation gets modified, while the continuity equation remains intact. Similar modification of the Euler equation for baryons (Eq. (5.7) with $u_s^2 = \omega = 0$) is

$$v'_B + H a v_B + \Phi = - (v_B - v_\gamma) \frac{\tau'}{k R_B}, \quad (8.179)$$

where (see Section 6.2)

$$R_B \equiv \frac{3\rho_B}{4\rho_\gamma}.$$

Similarly to photons, the continuity equation for baryons is not modified and has the form (5.6) with $u_s = w = 0$. By comparing that equation with Eq. (8.178a) one observes that the perturbations in baryon-photon plasma are adiabatic, in the sense that $3\delta'_\gamma = 4\delta'_B$.

We now generalize Eqs. (6.32) and (6.34) obtained in the tight coupling approximation. We consider modes of wavelengths exceeding the photon mean free path,

$$k \ll \tau'. \quad (8.180)$$

Shorter waves get damped by free streaming anyway. The estimate (8.169) shows that the relevant modes enter the horizon at radiation domination. These modes contain the oscillating part (6.36) which is of primary interest for us here. Accordingly, we discard the gravitational potentials. To simplify the formulas, we also disregard the terms proportional to the Hubble parameter in the above equations; the effect of these terms is encoded in the damping factor in (6.36). We consider Eqs. (8.178) and (8.179) and study linear effects in k/τ' . It follows from (8.178b) and (8.179) that

$$v_\gamma - v_B = - \frac{R_B}{4(1+R_B)} \frac{1}{\tau'} \cdot \delta_\gamma \quad (8.181)$$

(we have neglected here the anisotropic stress potential π_γ : it follows from (8.178c) that π_γ vanishes to the leading order in k/τ' , and hence does not contribute to (8.181)). We multiply Eq. (8.179) by R_B and add the result to (8.178b). In this way we obtain

$$v'_\gamma + R_B v'_B + \frac{1}{4} \delta_\gamma + \pi_\gamma = 0.$$

We combine this equation with the time derivative of Eq. (8.181) and find the modified Euler equation for photons to the linear order in k/τ' ,

$$(1 + R_B) v'_\gamma + \frac{1}{4} \delta_\gamma + \frac{R_B^2}{1 + R_B} \frac{1}{4\tau'} \delta'_\gamma + \pi_\gamma = 0. \quad (8.182)$$

We have neglected here time derivatives of R_B and τ' as compared to the time derivatives of δ_γ . This is possible since photon perturbations rapidly oscillate, while $(\log R_B)' \sim (\log \tau')' \sim 1/(Ha) \ll k$. We express the derivative of the velocity from (8.182) and insert it into the time derivative of the continuity equation (8.178a). The result is

$$\delta''_\gamma + \frac{R_B^2}{3(1 + R_B)} \frac{k^2}{\tau'} \delta'_\gamma + \frac{k^2}{3(1 + R_B)} \delta_\gamma + \frac{4k^2}{3(1 + R_B)} \pi_\gamma = 0. \quad (8.183)$$

Finally, the anisotropic stress potential is given by

$$\pi_\gamma = \frac{8}{9} \cdot \frac{k^2}{3\tau'} \cdot v_\gamma. \quad (8.184)$$

This formula follows, to the leading order in k/τ' , from Eq. (8.178c) (the higher multipole $F_{\gamma,3}$ is negligible in Eq. (8.178c), since Eq. (8.178) shows the hierarchy between multipoles, $F_{\gamma,l}/F_{\gamma,l-1} \simeq k/\tau'$).

We insert (8.184) into (8.183) and find the photon velocity v_γ from the continuity equation (8.178a). This gives the final equation for the photon density contrast,

$$\delta''_\gamma + \frac{k^2}{3(1 + R_B)\tau'} \left(\frac{8}{9} + \frac{R_B^2}{1 + R_B} \right) \delta'_\gamma + k^2 u_s^2 \delta_\gamma = 0, \quad (8.185)$$

where, like in Section 6.2, sound speed is $u_s^2 = [3(1 + R_B)]^{-1}$. The second term in Eq. (8.185) is precisely the term responsible for the Silk damping of acoustic oscillations. The WKB solution to Eq. (8.185) is

$$\delta_\gamma(\eta) = \frac{A}{(1 + R_B)^{1/4}} \exp \left(-\frac{k^2}{k_S^2(\eta)} \right) \cos \left(k \int^\eta d\tilde{\eta} u_s \right), \quad (8.186)$$

where we restored the damping factor present in (6.36). The amplitude A for adiabatic mode is given by (6.37). The conformal Silk momentum is given by

$$\frac{1}{k_S^2(\eta)} = \frac{1}{6} \int_0^\eta \frac{d\eta'}{(1 + R_B)\tau'} \left(\frac{8}{9} + \frac{R_B^2}{1 + R_B} \right). \quad (8.187)$$

This result agrees with the estimate (8.168) modulo a factor of order 1. Indeed, the fraction of free electrons is nearly 1 before recombination, and the integral (8.187) is estimated as $1/k_S^2 \sim \eta/\tau'$. Hence, perturbations of momenta $k \gtrsim \sqrt{\tau'/\eta}$ are suppressed. Since $H = 2/(a\eta)$ at matter domination, and $\tau' = a\sigma_T n_e$, the estimate for the physical Silk momentum is $q_S = k_S/a \sim \sqrt{H\sigma_T n_e}$, in accordance with (8.168).

Effects due to photon polarization modify the system of equations (8.178). This yields slightly different numerical coefficients. Namely, Eqs. (8.184), (8.185) and (8.187) contain the factor 16/15 instead of 8/9. Hence, damping gets slightly stronger.

To estimate the Silk momentum before recombination, we recall that $\tau' = a(\eta)n_e(\eta)\sigma_T \propto \eta^{-4}$. So, the integral in (8.187) is saturated at the upper limit. Since R_B is fairly small, see (6.41), its dependence on time is not particularly important. We neglect this dependence for an estimate, and obtain

$$\frac{1}{k_S^2(\eta_r)} = \frac{1}{30} \frac{\eta_r}{a(\eta_r)n_e(\eta < \eta_r)\sigma_T(1 + R_B)} \left(\frac{16}{15} + \frac{R_B^2}{1 + R_B} \right).$$

The physical Silk momentum is

$$\frac{a^2(\eta_r)}{k_S^2(\eta_r)} = \frac{1}{15} \frac{1}{H(\eta_r)n_e(\eta < \eta_r)\sigma_T(1 + R_B)} \left(\frac{16}{15} + \frac{R_B^2}{1 + R_B} \right). \quad (8.188)$$

This result refines the estimate (8.168). Making use of (6.41) we find that the wavelength $\lambda_S(\eta_r) = 2\pi/q_S(\eta_r)$ is larger by a factor of 1.5 than the estimate (8.169). We emphasize that the formula (8.188) gives the Silk scale *before the beginning of recombination*.

The approximations used for obtaining the formula (8.186) are not valid at recombination epoch itself. As we recalled in Section 8.5.1, the number density of free electrons rapidly decreases at that epoch. Hence, both the parameter τ' and photon mean free path also change rapidly in time. This invalidates the WKB approximation. Furthermore, Eq. (8.185) itself is no longer valid, since it has been obtained by assuming that τ' slowly varies in time. Finally, as the electron density decreases, the inequality (8.180) gets violated: photons become weakly coupled to electrons and baryons. These effects make it difficult to analyse the Silk damping at recombination. This damping is, however, quite substantial. To see this, let us estimate the contribution to the integral (8.187) due to the recombination epoch. We set the upper limit there equal to the photon last scattering time, $\eta = \eta_r$, and recall that n_e , and hence $\tau'(\eta) = a\sigma_T n_e(\eta)$ exponentially decrease, see (8.161). We change the integration variable to $1/T$ and obtain

$$\frac{1}{k_S^2(\eta_r)} = \frac{1}{k_S^2(\eta < \eta_r)} + \Delta \left(\frac{1}{k_S^2} \right), \quad (8.189)$$

where

$$\Delta \left(\frac{1}{k_S^2} \right) \simeq \frac{1}{6} \frac{1}{a^2 H \sigma_T n_e(\eta_r)(1 + R_B)} \frac{4T_r}{\Delta_H} \left(\frac{16}{15} + \frac{R_B^2}{1 + R_B} \right)$$

is the contribution due to the recombination epoch. Its ratio to the first term in (8.189) is

$$\frac{10T_r}{\Delta_H} \frac{n_e(\eta < \eta_r)}{n_e(\eta_r)} \simeq 1.5,$$

where we used the numerical values (8.157) and (8.165). Hence, $1/k_S^2$ grows at recombination considerably, by a factor of 2.3, according to our estimate, so the Silk wavelength $\lambda_S = 2\pi/k_S$ grows by a factor of about 1.5. This gives the estimate (8.170).

Thus, we arrive at the following picture. The main effect for waves in the baryon-photon medium, whose wavelengths are smaller than photon mean free path by recombination, is free streaming. Their present wavelengths are $\lambda_0 \lesssim 2 \text{ Mpc}$, see (8.158). Waves whose present wavelengths are in the interval $2 \text{ Mpc} \lesssim \lambda_0 \lesssim 50 \text{ Mpc}$ are subject to the Silk damping. The latter is approximately described by the formula (8.186) where $k_S \simeq a_0 \cdot 0.1 \text{ Mpc}^{-1}$. Waves with $\lambda_0 \gg 50 \text{ Mpc}$ are not damped at all, so they are described by the formula (6.36).

Our final comment concerns non-oscillating perturbations in the baryon-photon component, induced by the gravitational potentials of dark matter. They vary in time slowly, the relevant time scale being H^{-1} . On the other hand, the above analysis shows that the damping effect is proportional to δ'_γ . Hence, the perturbations in question are almost unaffected by the Silk damping, i.e., they are correctly described by the first term in (6.39).

Chapter 9

Temperature of Cosmic Microwave Background

As we pointed out several times, CMB carries a lot of information on both the cosmological perturbations at the photon last scattering epoch and properties of the Universe after last scattering. The energy spectrum of CMB photons is very precisely Planckian, with temperature

$$T_0 = 2.726 \pm 0.001 \text{ K}. \quad (9.1)$$

There are, however, variations of this temperature over the celestial sphere, i.e., temperature anisotropy. The strongest anisotropy at the level of $\delta T_0/T_0 \sim 10^{-3}$ has clear dipole structure, and its most probable explanation is the motion of the Earth with respect to CMB reference frame. The rest of the anisotropy is at the level $\delta T_0/T_0 \sim 10^{-4} - 10^{-5}$, and it is this part that is of extraordinary importance for cosmology. Besides the temperature anisotropy, CMB has polarization which is also of great interest. CMB polarization is considered in Chapter 10.

When studying CMB temperature anisotropy and polarization, we mainly consider adiabatic scalar perturbations. Possible contributions of isocurvature modes are discussed in Section 9.5.

9.1 CMB Temperature Anisotropy

Let $T_0(\mathbf{n})$ be the temperature of CMB photons coming from the direction \mathbf{n} on celestial sphere. One introduces the deviation of $T_0(\mathbf{n})$ from the average value (9.1),

$$\delta T_0(\mathbf{n}) \equiv T_0(\mathbf{n}) - T_0.$$

The relative temperature fluctuation is conveniently expanded in spherical harmonics $Y_{lm}(\mathbf{n})$, which constitute the complete orthonormalized set of functions on a unit sphere (see the book [61] and Section F.2 for details),

$$\frac{\delta T_0(\mathbf{n})}{T_0} = \sum_{l=1}^{\infty} \sum_{m=-l}^{m=l} a_{lm} Y_{lm}(\mathbf{n}). \quad (9.2)$$

The coefficients a_{lm} obey the relation

$$a_{l,m}^* = (-1)^m a_{l,-m}, \quad (9.3)$$

which follows from real-valuedness of temperature and the relation $Y_{l,m}^* = (-1)^{-m} Y_{l,-m}$.

Since the spherical harmonics are orthonormalized, there is an integral representation for a_{lm} ,

$$a_{lm} = \int d\mathbf{n} \frac{\delta T_0(\mathbf{n})}{T_0} Y_{lm}^*(\mathbf{n}), \quad (9.4)$$

where the integration is performed over the unit sphere. This formula is used to extract the coefficients a_{lm} from the measured function $\delta T_0(\mathbf{n})$. The further analysis proceeds in terms of a_{lm} . The asymptotic formula (F.41) for the spherical harmonics, valid at $l \gg 1$, shows that the multipole number l corresponds to the angular scale of order π/l , so the magnitude of the coefficients a_{lm} gives an idea of the amplitude of temperature fluctuations at this angular scale.

Since the dipole component with $l = 1$ is, most probably, due to the motion of the Earth with respect to CMB, this component is not used in cosmology. We quote here its value,

$$\delta T_{dipole} = 3.346 \pm 0.017 \text{ mK}. \quad (9.5)$$

It is subtracted from δT_0 , so the expansion (9.2) starts from $l = 2$. The coefficients a_{lm} do not depend on the motion of observer for $l \geq 2$.

The motion of an observer at velocity $v \ll 1$ with respect to the CMB reference frame results in the anisotropy of observed temperature due to the Doppler shift of photon frequency,

$$T_0(\theta) = T_0 \frac{\sqrt{1-v^2}}{1-v \cos \theta} = T_0 \cdot \left(1 + v \cos \theta + \frac{v^2}{2} \cos 2\theta + \dots \right), \quad (9.6)$$

where θ is the angle between the directions of motion and observation. Solar system has velocity $v = 369 \pm 2 \text{ km/s}$ towards Hydra constellation. The direction of this motion in Galactic coordinates is $(263.86^\circ \pm 0.04^\circ, 48.24^\circ \pm 0.10^\circ)$. Both the value of the velocity and its direction are determined precisely from the dipole anisotropy.

For isotropic on average Universe and random field of temperature fluctuations, the coefficients a_{lm} are uncorrelated at different l and m . This statement applies to a hypothetical ensemble of Universes like ours. Averages over such an ensemble have the following property,

$$\langle a_{lm} a_{l'm'}^* \rangle = C_l \cdot \delta_{ll'} \delta_{mm'}, \quad (9.7)$$

where the proportionality coefficient C_l depends only on the total angular momentum l but not on its projection m . Cosmological observations (notably, CMB measurements) show that the primordial perturbations are Gaussian; the Gaussianity is preserved during linear evolution. The coefficients a_{lm} are linear functions of primordial perturbations, so they are also Gaussian random variables,

see Appendix C. Hence, the coefficients C_l completely determine CMB temperature anisotropy.

In reality, there is only one Universe we can study. So, there is only one set of coefficients a_{lm} , and one cannot check the relation (9.7) directly. For higher multipoles which have large number of harmonics with different $m = -l, \dots, l$, one can test whether the statistical properties of the observed set of a_{lm} are consistent with the hypothesis (9.7) and with the Gaussianity of the temperature fluctuations. With existing data, this consistency is quite reasonable.

Assuming the validity of (9.7), one determines the coefficients C_l from observations approximately, by evaluating the average over m ,

$$C_l = \frac{1}{2l+1} \sum_{m=-l}^{m=l} |a_{lm}|^2. \quad (9.8)$$

The relative statistical error in C_l is estimated by¹ $1/\sqrt{l+1/2}$. Under the assumption of the Gaussianity, the measured quantities (9.8) carry all information about CMB temperature anisotropy in our Universe. There is irreducible statistical uncertainty in C_l , which we call cosmic variance, so theoretically predicted values of C_l may well differ from measured ones within the relative margin of $1/\sqrt{l+1/2}$. In principle, cosmic variance precludes measuring the cosmological parameters and primordial spectra with arbitrary precision.

There is ongoing discussion in literature on whether the assumptions of the spatial isotropy and Gaussianity of perturbations are indeed valid. The existing data may possibly show correlations of harmonics with different l , considerable variations of C_l in certain multipole intervals and even the existence of an axis correlated to the Galactic plane. There are claims in literature that even with account of cosmic variance, some of these properties are inconsistent with (9.7) and the standard isotropic cosmological model at confidence level of better than 99% [62, 63]. These claims are controversial, however; detailed arguments in favor of the Gaussianity are given in Ref. [64]. We assume in what follows that the formulas (9.7) and (9.8) are valid for our Universe.

The coefficients C_l determine the two-point correlation function of the temperature fluctuations $\langle \delta T_0(\mathbf{n}_1) \delta T_0(\mathbf{n}_2) \rangle$. For the Gaussian fluctuations, all other correlation functions are expressed in terms of the latter. Making use of the definition (9.2) we write for the two-point function,

$$\begin{aligned} \langle \delta T_0(\mathbf{n}_1) \delta T_0(\mathbf{n}_2) \rangle &= T_0^2 \cdot \sum_l C_l \sum_m Y_{lm}(\mathbf{n}_1) Y_{lm}^*(\mathbf{n}_2) \\ &= T_0^2 \cdot \sum_l \frac{2l+1}{4\pi} C_l \cdot P_l(\mathbf{n}_1 \mathbf{n}_2), \end{aligned} \quad (9.9)$$

where P_l are the Legendre polynomials.

¹The definition (9.8) and the fact that the average values of a_{lm} vanish imply that C_l are distributed according to χ^2_{2l+1} law, so the fluctuation is $(\delta C_l)^2 = 2C_l^2/(2l+1)$.

Problem 9.1. Making use of the definitions (F.39), (F.29), prove the representation (9.9).

The variance of temperature is obtained from (9.9),

$$\langle \delta T_0^2(\mathbf{n}) \rangle = T_0^2 \cdot \sum_l \frac{2l+1}{4\pi} C_l \approx T_0^2 \cdot \int d \log l \cdot \frac{(l+1)l}{2\pi} C_l, \quad (9.10)$$

where the last approximate equality is valid for large l . Hence, the quantity

$$\mathcal{D}_l \equiv T_0^2 \frac{l(l+1)}{2\pi} C_l \quad (9.11)$$

determines the amplitude squared of the temperature fluctuation in a decimal interval of multipoles. It is this quantity that is usually shown in plots.

Some results of the measurements of CMB temperature anisotropy are shown in Fig. 9.1. The distribution of warm and cold regions in celestial sphere is illustrated in Fig. I.1.5.

The data plotted in Fig. 9.1 show that:

- (1) CMB temperature anisotropy is at the level $\delta T_0/T_0 \sim 10^{-4} - 10^{-5}$;
- (2) the angular spectrum has the oscillatory behavior with maxima at $l \sim 200, 500, 800, 1100, 1400$;
- (3) the anisotropy decreases as l increases, starting from $l \sim 500$;
- (4) the oscillations are rapidly damped, as l increases, at $l \gtrsim 1000$.

All these features find their explanations in the Λ CDM cosmological model with adiabatic scalar perturbations whose primordial spectrum is nearly flat and has the amplitude of order 10^{-5} . The oscillations of C_l as function of l

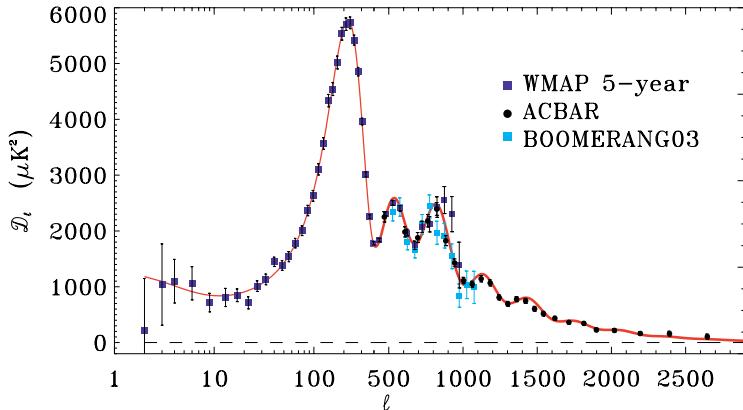


Fig. 9.1 CMB temperature anisotropy data obtained by WMAP, BOOMERANG and ACBAR experiments [65]. The horizontal axis is logarithmic at $l \leq 500$ and linear at higher l . The data points are obtained by binning (averaging over small intervals of angular momenta l). Note the consistency of different experiments in the overlap region. Solid line shows the prediction of the best fit Λ CDM model.

are due to acoustic waves in the baryon-photon plasma at the recombination epoch [16, 17]. The latter phenomenon is quite similar to the Sakharov oscillations, see the beginning of Section 7.1.2. We discuss these oscillations in details in what follows.

Damping of the oscillations in the angular spectrum at $l \gtrsim 1000$ reflects the suppression of the acoustic waves due to the Silk effect. The overall decline in the spectrum is explained mostly by the decrease of the dark matter gravitational potential at high momenta and the finite thickness of the last scattering sphere. At these and especially even smaller angular scales, the temperature anisotropy is additionally smoothed by weak gravitational lensing by galaxies and their clusters.

We now turn to the detailed analysis of CBM temperature angular spectrum and its relation to the cosmological perturbations and photon propagation after last scattering.

9.2 Temperature Anisotropy in Instant Photon Decoupling Approximation

In this Section we consider CMB temperature anisotropy spectrum at large and intermediate angular scales. To this end, it is adequate to make use of the tight coupling approximation in describing the baryon-photon plasma before last scattering and the idealized picture of instantaneous decoupling of CMB photons. Also, we can neglect photon interactions after last scattering, except for the interactions at the epoch of reionization.

9.2.1 General formalism

To simplify the analysis, we note that light-like geodesics, along which photons propagate in the expanding Universe, are invariant under conformal transformations. Indeed, let $x^\mu(s)$ be photon world line, where s is the affine parameter. The action for massless particles is (see Section I.A.7)

$$-\frac{1}{2} \int ds \xi(s) \frac{dx^\mu}{ds} \frac{dx^\nu}{ds} g_{\mu\nu}(x),$$

where $\xi(s)$ is the auxiliary dynamical variable. It is invariant under conformal transformations $g_{\mu\nu} \rightarrow a^2 g_{\mu\nu}$, $ds \rightarrow a^2 ds$. Therefore, geodesics $x^\mu(s)$ in conformal coordinates coincide with geodesics $x^\mu(\lambda)$ in space-time with metric $\gamma_{\mu\nu} = g_{\mu\nu}/a^2(\eta)$ which does not contain the scale factor, see (2.26). We work with metric $\gamma_{\mu\nu}$ in what follows; spatial coordinates, time, momenta and frequencies are then conformal ones. We often do not explicitly mention the latter fact.

Photon world line obeys the geodesic equation

$$\frac{d^2 x^\mu}{d\lambda^2} + \gamma_{\nu\rho}^\mu \frac{dx^\nu}{d\lambda} \frac{dx^\rho}{d\lambda} = 0, \quad (9.12)$$

where $\gamma_{\nu\rho}^\mu$ are the Christoffel symbols calculated with metric $\gamma_{\mu\nu}$. As usual, the photon momentum is

$$P^\mu = \frac{dx^\mu}{d\lambda}. \quad (9.13)$$

We consider it as a function of time along the world line, $\eta = x^0(\lambda)$. Then

$$\frac{dP^\mu}{d\lambda} = \frac{dx^0}{d\lambda} \cdot \frac{dP^\mu}{d\eta} = P^0 \cdot \frac{dP^\mu}{d\eta}.$$

Equation (9.12) with $\mu = 0$ can be viewed as the equation for the time component of the momentum,

$$\frac{dP^0}{d\eta} + \gamma_{\nu\rho}^0 \frac{P^\nu}{P^0} \frac{P^\rho}{P^0} P^0 = 0. \quad (9.14)$$

As an important example, consider scalar perturbations of metric in the spatially flat Universe (cf. (2.66), (2.67)),

$$\delta\gamma_{00} = 2\Phi, \quad \delta\gamma_{ij} = -2\Psi\delta_{ij}.$$

The linearized Christoffel symbols are given in (B.6): $\gamma_{00}^0 = \Phi'$, $\gamma_{oi}^0 = \partial_i\Phi$, $\gamma_{ij}^0 = \Psi'\delta_{ij}$. Hence, linearized Eq. (9.14) has the following form,

$$\frac{dP^0}{d\eta} + P^0 \cdot \left(\Phi' + \Psi' \frac{P^i}{P^0} \frac{P^j}{P^0} \delta_{ij} + 2 \frac{P^i}{P^0} \partial_i\Phi \right) = 0. \quad (9.15)$$

We find from Eq. (9.12) that to the zeroth order in perturbations $P^0 = \text{const}$, $P^i = \text{const}$, so that $n_i = n^i = P^i/P^0$ is the time-independent unit 3-vector along the photon trajectory. With this notation, Eq. (9.15) reads

$$\frac{dP^0}{d\eta} = (\Phi' - \Psi') P^0 - 2(\Phi' + \mathbf{n}\nabla\Phi) P^0. \quad (9.16)$$

The expression in parenthesis in the second term in (9.16) is the total derivative along the photon world line. This follows from the chain of equalities

$$\frac{d\Phi(\eta, \mathbf{x}(\eta))}{d\eta} = \Phi' + \frac{\partial \mathbf{x}}{\partial \eta} \nabla\Phi = \Phi' + \frac{\partial \mathbf{x}}{\partial \eta/\partial \lambda} \nabla\Phi = \Phi' + \frac{\mathbf{P}}{P^0} \nabla\Phi = \Phi' + \mathbf{n}\nabla\Phi.$$

Hence, Eq. (9.16) can be integrated; to the linear order in metric perturbations we have

$$\frac{P^0(\eta'') - P^0(\eta')}{P^0(\eta')} = \int_{\eta'}^{\eta''} (\Phi' - \Psi') d\eta - 2[\Phi(\eta'') - \Phi(\eta')], \quad (9.17)$$

where the integration is performed along the photon trajectory.

Consider now a photon last scattered by cosmic plasma at a point \mathbf{x} . Let Ω be the photon frequency in the rest frame of the plasma. In the reference frame of conformal Newtonian gauge, the plasma at point \mathbf{x} moves with 4-velocity U^μ (recall that we use conformal quantities throughout). For scalar perturbations, we have to

the linear order (see Section 2.2.2) $U^0 = 1 - \Phi$, $U^i = v^i$, where v^i is the 3-vector of the plasma velocity. The components with lower indices are

$$U_0 = 1 + \Phi, \quad U_i = -v^i. \quad (9.18)$$

Let us find the frequency Ω in terms of the momenta P^μ defined in the conformal Newtonian reference frame. In the plasma reference frame, we can choose metric at the point \mathbf{x} to be equal to the Minkowski metric, $\tilde{\gamma}_{\mu\nu} = \eta_{\mu\nu}$; hereafter tilde denotes quantities in the plasma reference frame. Then $\Omega = \tilde{P}^0$. With this choice of metric, the plasma 4-velocity at the moment of photon emission is $\tilde{U}_\mu = (1, 0, 0, 0)$, i.e., $\Omega = \tilde{U}_0 \tilde{P}^0$. This expression can be written in the invariant form,

$$\Omega = U_\mu P^\mu. \quad (9.19)$$

The right hand side of Eq. (9.19) can be calculated in the conformal Newtonian gauge.

Using the expression (9.18) for velocities, we obtain from (9.18) to linear order,

$$\Omega(\eta') = [1 + \Phi(\eta') - \mathbf{n}\mathbf{v}(\eta')]P^0(\eta'),$$

where η' is the time of photon last scattering. Similar formula is valid for the photon frequency at the time of detection η'' , and in that case $\Omega(\eta'')$ is the measured photon frequency while \mathbf{v} is the observer velocity in the conformal Newtonian frame. Hence, we obtain from (9.17) the relative frequency shift for photon emitted in the direction \mathbf{n} at time η' and absorbed at time η'' ,

$$\begin{aligned} & \frac{\Omega(\mathbf{n}, \eta'') - \Omega(\mathbf{n}, \eta')}{\Omega(\mathbf{n}, \eta')} \\ &= \int_{\eta'}^{\eta''} (\Phi' - \Psi') d\eta + \Phi(\eta') - \Phi(\eta'') + \mathbf{n}\mathbf{v}(\eta') - \mathbf{n}\mathbf{v}(\eta''). \end{aligned} \quad (9.20)$$

Note that the frequency shift is proportional to the frequency itself, so the CMB energy spectrum remains thermal after last scattering. The photon temperature, however, depends on the arrival direction. The formula for relative temperature shift is the same as (9.20). We account for the fact that the temperature of emitted photons varies in space at last scattering because of the energy density perturbations, $\delta_\gamma \equiv \delta\rho_\gamma/\rho_\gamma = 4\delta\omega/\omega = 4\delta T/T$. Hence, the final expression for the observable temperature fluctuation is

$$\frac{\delta T}{T}(\mathbf{n}, \eta_0) = \frac{1}{4}\delta_\gamma(\eta_r) + (\Phi(\eta_r) - \Phi(\eta_0)) \quad (9.21a)$$

$$+ \int_{\eta_r}^{\eta_0} (\Phi' - \Psi') d\eta \quad (9.21b)$$

$$+ \mathbf{n}\mathbf{v}(\eta_r) - \mathbf{n}\mathbf{v}(\eta_0). \quad (9.21c)$$

The contribution (9.21a) is called the Sachs–Wolfe (SW) effect [66]. Its second part is the photon relative frequency shift induced by the gravitational potentials at

emission and detection. The monopole component $\Phi(\eta_0)$ is unobservable, since it can be absorbed into the average measured temperature. We set it equal to zero in what follows.

The term (9.21b) also reflects the change of the photon frequency due to gravitational potentials, now depending on time. If photon passes through a region with negative potential whose magnitude grows in time, it first “dives” into the potential well and then gets out from the deeper well. As a result, photon loses its energy. This phenomenon is called the integrated Sachs–Wolfe (ISW) effect.

Finally, the first contribution in (9.21c) is due to the Doppler effect: at the last scattering epoch, baryon-electron-photon medium moves with respect to the conformal Newtonian frame with the velocity $\mathbf{v}(\eta_r)$. The second term in (9.21c) accounts for analogous effect, now due to observer’s motion with the velocity $\mathbf{v}(\eta_0)$. We noticed that the latter effect results in the large dipole anisotropy, which is subtracted from the data since it is irrelevant for cosmology.

In view of this discussion, in what follows we use the formula

$$\frac{\delta T}{T}(\mathbf{n}, \eta_0) = \frac{1}{4}\delta_\gamma(\eta_r) + \Phi(\eta_r) \quad (9.22a)$$

$$+ \int_{\eta_r}^{\eta_0} (\Phi' - \Psi') d\eta \quad (9.22b)$$

$$+ \mathbf{n}\mathbf{v}(\eta_r). \quad (9.22c)$$

We emphasize that the quantities at time η_r refer to the photon emission point, \mathbf{n} is the direction of the photon propagation, and the integral in (9.22b) is evaluated along the photon trajectory.

The result (9.22) does not account for one important phenomenon. Hydrogen in the Universe gets reionized at the epoch of the formation of first stars, $z \sim 10$, and there is non-vanishing probability of photon scattering off free electrons at that epoch. This leads to slight wash out of the anisotropy. We neglect this effect for the time being; it is discussed in Section 9.2.3.

Similar analysis goes through for tensor perturbations h_{ij}^{TT} . In that case, the non-vanishing Christoffel symbols are $\gamma_{ij}^0 = -h'_{ij}/2$ (we temporarily omit the notation TT). We insert them into Eq. (9.14), and obtain the equation for the time component of photon 4-momentum,

$$\frac{dP^0}{d\eta} - \frac{1}{2}P^0 h'_{ij} n_i n_j = 0.$$

To the linear order, the factor P^0 in the second term is independent of time, so we have

$$\frac{P^0(\eta'') - P^0(\eta')}{P^0(\eta')} = \frac{1}{2} \int_{\eta'}^{\eta''} d\eta n_i h'_{ij} n_j. \quad (9.23)$$

Matter velocity is unperturbed for tensor perturbations, $U^0 = 1$, $U^i = 0$, so the emitted and measured photon frequencies coincide with $P^0(\eta')$ and $P^0(\eta'')$,

respectively. Hence, the non-trivial tensor structure implies the strong limitation on the possible contribution to CMB temperature anisotropy: the only effect is ISW,

$$\frac{\delta T}{T}(\mathbf{n}, \eta_0) = \frac{1}{2} \int_{\eta_r}^{\eta_0} d\eta n_i h_{ij}^{TT'} n_j, \quad (9.24)$$

where the integral is again evaluated along the photon trajectory.

Since the gravity wave amplitude decays after the horizon entry, see (3.6), the strongest contribution is expected to come from long tensor waves which enter the horizon after recombination or shortly before it, see Section 9.2.2 for details. Hence, the tensor perturbations are expected to show up in CMB temperature anisotropy at large angular scales only. The bounds on relic gravity waves presented in Section 5.4 are indeed obtained by CMB measurements at large angular scales. We point out here that CMB temperature anisotropy alone is not very sensitive to tensor perturbations for two reasons. First, this anisotropy is a sum of contributions due to the scalar and tensor perturbations; second, there is substantial cosmic variance at low multipoles (large angular scales). Measurements of CMB polarization are more promising from this viewpoint. We consider this aspect in Section 10.

Problem 9.2. Explain, without calculations, why SW and Doppler effects do not have tensor counterparts.

Problem 9.3. Obtain the expression analogous to (9.22), (9.24) for vector perturbations.

Let us come back to the scalar perturbations. Before calculating the anisotropy spectrum, we make a few comments on the ISW effect. The first, general comment is as follows. The terms (9.22a), (9.22c) come from the photon last scattering epoch at which the linearized theory is fully applicable. On the contrary, the ISW term (9.22b) involves the contribution of late times, so the linearized theory may appear to be inadequate. We will see, however, that the *late ISW effect* shows up at large angular scales only. These angular scales correspond to large spatial scales of perturbations, which are still in the linear regime. Hence, all effects studied in this Section are correctly described within the linearized theory.

Second, the gravitational potentials are constant in time at matter domination, $\Phi' = \Psi' = 0$, so the ISW effect is absent. Some contribution comes from the time just after recombination, since the energy density of relativistic matter is still considerable at that time, and the gravitational potentials still evolve (early ISW effect). Another contribution comes from the recent epoch, at which dark energy is important and the gravitational potentials decrease in time. This late ISW effect leads to the enhancement of temperature anisotropy at the largest scales, see Section 9.2.2.

The measurement of the late ISW effect together with the study of the matter distribution in the late Universe is of importance for determining the evolution of the scale factor at low redshifts, $z \lesssim 1$, and hence for obtaining the information on

dark energy. The existing data are consistent with the cosmological constant serving as dark energy, but this result is not very precise. The ISW effect by itself is not very sensitive to the dark energy properties. More promising is the study of *correlations* between CMB temperature and matter distribution at large spatial scales: the ISW effect is enhanced in overdense regions where the gravitational potentials are large. Evidence for this correlation has indeed been found [71, 72].

The absence of the ISW effect at $z \gtrsim 1$ is consistent with observations [73–75]. This confirms the expectation that $\Psi' = \Phi' = 0$ at $z \gtrsim 1$, i.e., that the Universe was indeed matter dominated from the recombination epoch till late times, $z \sim 1$. Interestingly, this fact can be used for constraining particle physics models beyond the Standard Model. As an example, models with exotic particles that decay into relativistic particles (photons and/or neutrinos) and have lifetime $\tau \gtrsim 10^{13}$ s are constrained by the absence of the ISW effect at intermediate angular scales: relativistic particles produced in these decays would modify the cosmological expansion rate between recombination and $z \sim 1$, thus leading to the non-trivial evolution of the gravitational potentials which in turn would generate the ISW effect. Similar exotic particles with lifetime $\tau \lesssim 10^{13}$ s would increase the fraction of relativistic matter after recombination, which would result in delayed freeze out of the gravitational potentials and the enhancement of the *early ISW effect* [76].

Let us now calculate the coefficients C_l . We consider the effect of scalar perturbations here; tensor perturbations are discussed in Section 9.2.2. We choose the origin of the coordinate frame at the point of observation and denote the 3-vector *in the direction of observation* by \mathbf{n} . So, CMB photon momentum is directed along $-\mathbf{n}$. A convenient definition of the observed relative temperature fluctuation is

$$\Theta_0(\mathbf{n}) \equiv \frac{\delta T}{T}(-\mathbf{n}; \eta_0), \quad (9.25)$$

Here $\delta T(\mathbf{n})/T$ is calculated according to (9.22); \mathbf{n} in that formula is the direction of photon propagation, hence the sign in (9.25). It follows from (9.8), (9.4) and the theorem on addition of spherical harmonics that the coefficients C_l have the following integral representation,

$$C_l = \frac{1}{4\pi} \int d\mathbf{n}' d\mathbf{n}'' \Theta_0(\mathbf{n}') \Theta_0(\mathbf{n}'') P_l(\mathbf{n}' \mathbf{n}''). \quad (9.26)$$

where P_l are the Legendre polynomials. It is convenient to obtain more explicit form of this representation. As we have seen in previous Chapters, cosmological perturbations are naturally described in the momentum space. To exploit this fact, we note that the right hand side of (9.22) contains perturbations at points $\mathbf{x} = (\eta_0 - \tilde{\eta}) \mathbf{n}$, where $\tilde{\eta}$ is either the time of last scattering (in (9.22a), (9.22c)) or intermediate time between last scattering and the present epoch (in (9.22b)). Hence, the quantity $\Theta_0(\mathbf{n})$ is the integral over momenta whose integrand depends on momentum only

in the combination \mathbf{kn} ,

$$\Theta_0(\mathbf{n}) = \int d^3k \Theta_0(\mathbf{kn}, k). \quad (9.27)$$

It is therefore convenient to expand Θ_0 in the Legendre polynomials,²

$$\Theta_0(\mathbf{n}) = \sum_l i^l \cdot (2l+1) \int d^3k \tilde{\Theta}_l(\mathbf{k}) \cdot P_l\left(\frac{\mathbf{kn}}{k}\right). \quad (9.28)$$

This expression shows that multipoles $\tilde{\Theta}_l(\mathbf{k})$ encode the information on the contribution of perturbations of wavelength $2\pi/k$ to the temperature anisotropy at angular scale of order π/l .

We now extract the dependence on the primordial spectrum of perturbations by writing

$$\tilde{\Theta}_l(\mathbf{k}) = \Theta_l(k) \cdot \Phi_{(i)}(\mathbf{k}). \quad (9.29)$$

We assume here for definiteness that the perturbations are adiabatic, and parameterize their initial values by the initial gravitational potential $\Phi_{(i)} = -2\mathcal{R}/3$, see (6.3). For isocurvature modes, the formula (9.29) contains $\delta_{CDM,(i)} = \mathcal{S}_{CDM}$ or $\delta_{B,(i)} = \mathcal{S}_B$ instead of $\Phi_{(i)}$. When writing (9.29), we used the fact that the non-trivial dependence on the direction of momentum exists only for $\Phi_{(i)}(\mathbf{k})$, which is in accordance with (9.22). The representation (9.29) is convenient, since it separates the effects due to the primordial perturbations from those related to the subsequent evolution.

Assuming that the primordial perturbations are random and Gaussian, we insert (9.28), (9.29) into (9.26) and make use of (5.39). We thus obtain,

$$\begin{aligned} C_l &= \frac{1}{4\pi} \int \frac{k^2 dk}{(2\pi)^3} P_\Phi(k) \sum_{l', l''} (-i)^{l'} (-i)^{l''} (2l'+1) (2l''+1) \Theta_{l'}(k) \Theta_{l''}(k) \\ &\times \int d\mathbf{n}' \int d\mathbf{n}'' \int d\mathbf{n}_k P_{l'}(\mathbf{n}_k \mathbf{n}'') P_{l''}(-\mathbf{n}_k \mathbf{n}') P_l(\mathbf{n}'' \mathbf{n}'), \end{aligned}$$

where $\mathbf{n}_k \equiv \mathbf{k}/k$. The angular integrals are evaluated by using the integral version of the theorem on addition for the Legendre polynomials,

$$\int d\mathbf{n}'' P_l(\mathbf{nn}'') P_{l'}(\mathbf{n}'\mathbf{n}'') = \delta_{ll'} \frac{4\pi}{2l+1} P_l(\mathbf{nn}').$$

As a result we obtain the general formula

$$C_l = 4\pi \int \frac{dk}{k} \cdot \mathcal{P}_\Phi(k) \cdot \Theta_l^2(k), \quad (9.30)$$

where, as usual, $\mathcal{P}_\Phi(k) = k^3 P_\Phi(k)/(2\pi^2) = 4\mathcal{P}_R/9$. The formula (9.30) shows that the l -th temperature angular harmonic is determined by both the primordial

²We use the fact that the Legendre polynomials constitute a complete set of functions in the interval $[-1, 1]$.

spectrum and the value of the multipole Θ_l with the same l . The properties of the multipoles Θ_l are different in different ranges of l , i.e., at different angular scales. So, we study large, intermediate and small angular scales separately.

9.2.2 Large angular scales

As the first application of the formula (9.30), we consider the contribution to CMB temperature anisotropy coming from perturbations that are superhorizon at recombination. The momenta of these perturbations obey $k\eta_r < 1$. We will see towards the end of this Section that the SW effect (9.22a) dominates for these perturbations. It follows from (5.21) and (6.2) that for these modes

$$\frac{1}{4}\delta_\gamma = \frac{\delta T}{T}(\eta_r) = -\frac{2}{3}\Phi. \quad (9.31)$$

Hence, the two terms in (9.22a) combine into

$$\frac{\delta T}{T}(\mathbf{n}, \eta_0) = \frac{1}{3}\Phi = \frac{3}{10}\Phi_{(i)}, \quad (9.32)$$

where we have recalled the relation (6.3). Note the partial cancellation between the two contributions to the SW effect, $\delta_\gamma/4$ and Φ .

So, the dominant effect due to the large wavelength perturbations is expressed as follows,

$$\Theta_0(\mathbf{n}) = \frac{1}{3}\Phi[\eta_r, (\eta_0 - \eta_r)\mathbf{n}] = \frac{1}{3}\int d^3k\Phi(\mathbf{k})e^{i\mathbf{kn}\cdot(\eta_0-\eta_r)}. \quad (9.33)$$

According to (F.35), plane wave is expanded in the Legendre polynomials with the expansion coefficients given in terms of the spherical Bessel functions. We insert this expansion into (9.33) and compare the result with the expansion (9.28). In this way we obtain

$$\tilde{\Theta}_l(\mathbf{k}) = \frac{1}{3}\Phi(\mathbf{k})j_l[(\eta_0 - \eta)r] = \frac{3}{10}\Phi_{(i)}(\mathbf{k})j_l[(\eta_0 - \eta)r], \quad (9.34)$$

or, in terms of the multipoles independent of the primordial spectrum,

$$\Theta_l(r) = \frac{3}{10}j_l[(\eta_0 - \eta)r] \approx \frac{3}{10}j_l(k\eta_0). \quad (9.35)$$

The spherical Bessel functions $j_l(x)$ are very small at $x < l$ and decrease at $x \gg l$ (see Appendix F). Therefore, it follows from (9.35), (9.30) that at least in the case of the nearly flat primordial spectrum, the perturbations of wavelength $2\pi/k$ contribute mostly to angular harmonics with $l \sim k\eta_0$. This property can be explained as follows. For not too large angular scales, i.e., multipoles with $l \gg 1$, the curvature of the sphere of last scattering is not very important. If it is neglected, the expansion of the SW term in spherical harmonics is basically the Fourier transformation of $[\delta_\gamma/4 + \Phi](\eta_r)$ in the plane at $\eta = \eta_r$, normal to the line of sight. The value of the 2-dimensional ‘‘momentum’’ is equal to $k_{eff} = l/(\eta_0 - \eta_r)$. This

is clear, e.g., from the fact that the sphere of last scattering has coordinate size $(\eta_0 - \eta_r)$; the Laplacian on this sphere is equal to $l(l+1)/(\eta_0 - \eta_r)^2 \simeq l^2/(\eta_0 - \eta_r)^2$, while the Laplacian in our “2-dimensional Fourier representation” is equal to k_{eff}^2 . Since the sphere of last scattering has finite area, the “momentum” \mathbf{k}_{eff} takes discrete values. It is clear that perturbations of 3-momentum \mathbf{k} contribute to the “Fourier harmonic” with $k_{\text{eff}} = l/(\eta_0 - \eta_r)$ only when the projection of momentum \mathbf{k} on the plane orthogonal to the line of sight takes an appropriate value. This happens when $k|\sin \theta| \approx k_{\text{eff}}$, where θ is the angle between the direction of the wave propagation and the line of sight, see Fig. 9.2. It is for this reason that perturbations of momenta $k > l/(\eta_0 - \eta_r) \approx l/\eta_0$ contribute to the SW effect at multipole l . The oscillations of the right hand side of (9.35) are due to the fact that $k \sin \theta$ may not coincide with any of the discrete “momenta” \mathbf{k}_{eff} . Similar consideration applies to the Doppler contribution to CMB temperature anisotropy.

It is clear from (9.35) that in the SW case, the temperature multipoles l are mostly saturated by perturbations of momenta $k \sim l/\eta_0$ whose directions \mathbf{k} are at large angles with the line of sight. Perturbations with higher k give some contributions as well. These are due to modes of perturbations whose momenta are almost parallel to the line of sight. Overall, this behavior is in accordance with the intuitive picture that perturbations of a given wavelength can affect temperature fluctuations at larger angular scales but not at smaller ones.

Let us continue our quantitative analysis. By using (9.35) in (9.30), we obtain

$$C_l = \frac{36\pi}{100} \int \frac{dk}{k} \mathcal{P}_\Phi(k) \cdot j_l^2(k\eta_0). \quad (9.36)$$

This formula is valid only for relatively small l , i.e., large angular scales. Indeed, the expressions (9.33)–(9.35) are valid only for perturbations with low momenta, $k\eta_r < 1$. The condition of validity of (9.36) is that the integral is saturated in this

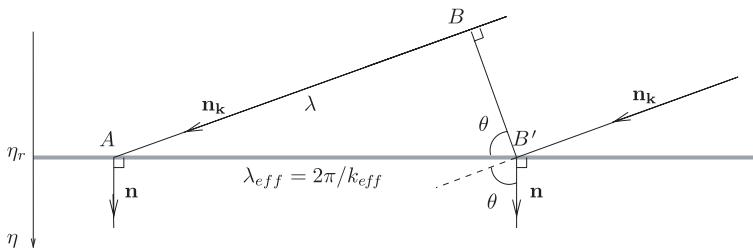


Fig. 9.2 Illustration of the contribution due to perturbation with wavelength $\lambda = 2\pi/k$ propagating along $\mathbf{n}_k \equiv \mathbf{k}/k$, into angular harmonic with $l \simeq (\eta_0 - \eta_r) k_{\text{eff}}$. Photons move from the surface of last scattering (gray line) towards the observer in the direction \mathbf{n} (opposite to the direction of sight), while the vector \mathbf{k} is at angle θ to \mathbf{n} . The phase of the wave of perturbation is the same in the planes normal to \mathbf{n}_k . In particular, the phase is the same at points B and B' . If the distance between the points A and B is equal to λ , the phases at points A , B and B' coincide. Once the relation $\lambda/\sin \theta = \lambda_{\text{eff}}$ is satisfied, the perturbation of momentum k directed along \mathbf{n}_k contributes to the “Fourier harmonic” with $k_{\text{eff}} = 2\pi/\lambda_{\text{eff}}$.

low momentum region. For flat or nearly flat primordial spectrum, $\mathcal{P}_\Phi(k) \approx \text{const}$, the integral in (9.36) is saturated at $k \sim l/\eta_0$. We require that $(l/\eta_0)\eta_r < 1$, and make use of (2.21). In this way we find that the formula (9.36) is valid for

$$l < \frac{\eta_0}{\eta_r} \simeq 50.$$

Problem 9.4. Obtain the expression (9.36) directly from (9.26).

Let us choose the power-law approximation (5.40) for the primordial power spectrum. Then Eq. (9.36) yields

$$\begin{aligned} C_l &= \frac{36\pi A_\Phi}{100(k_*\eta_0)^{n_s-1}} \int_0^\infty d\xi \cdot \xi^{n_s-2} \cdot j_l^2(\xi) \\ &= \frac{9A_\Phi\pi^2}{400(k_*\eta_0)^{n_s-1}} \cdot \frac{2^{n_s}\Gamma(3-n_s)\Gamma(l+\frac{n_s-1}{2})}{\Gamma^2(\frac{3}{2}-\frac{n_s-1}{2})\Gamma(l+2-\frac{n_s-1}{2})}. \end{aligned} \quad (9.37)$$

The measured CMB temperature anisotropy at large angular scales gives the determination of the amplitude of the primordial power spectrum A_Φ . The dependence of C_l on l is in one-to-one correspondence with the tilt n_s of the primordial spectrum.

The coefficients C_l decrease with l ,

$$C_l \propto \frac{1}{(l+1-\frac{n_s-1}{2})(l-\frac{n_s-1}{2})} \cdot \frac{\Gamma(l+\frac{n_s-1}{2})}{\Gamma(l-\frac{n_s-1}{2})} \propto l^{n_s-3}.$$

The blue primordial power spectrum implies larger values of C_l at higher l , and vice versa. This is actually the general result: for n_s close to 1, the dependence of C_l on the tilt for all angular scales is

$$C_l(n_s) = \left(\frac{l}{k_*\eta_0}\right)^{n_s-1} \cdot C_l(n_s=1), \quad (9.38)$$

see also problem 9.6. Here k_* is the fiducial momentum entering (5.40).

For flat primordial power spectrum, the angular spectrum of CMB temperature anisotropy is also flat,

$$C_l = \frac{18\pi}{100} \frac{A_\Phi}{l(l+1)}. \quad (9.39)$$

We recall the definition (9.11) and conclude that the function \mathcal{D}_l is basically independent of l for the SW part. This is illustrated in Fig. 9.3; recall that we consider here low multipoles, $l \lesssim 50$. It is clear from Fig. 9.1 that the real primordial power spectrum is indeed nearly flat.

Regarding the case of the flat primordial power spectrum, we see from Fig. 9.3 that the total angular spectrum \mathcal{D}_l exhibits deviations from the flatness at $l \sim 2-4$ and $l \gtrsim 50$. The reason for them is threefold. Temperature anisotropy at small l is affected by the late ISW effect. The anisotropy at $l \gtrsim 50$ obtains contributions from perturbations entering the horizon at recombination or soon afterwards. Also, the fact that radiation-matter equality and recombination are separated by short

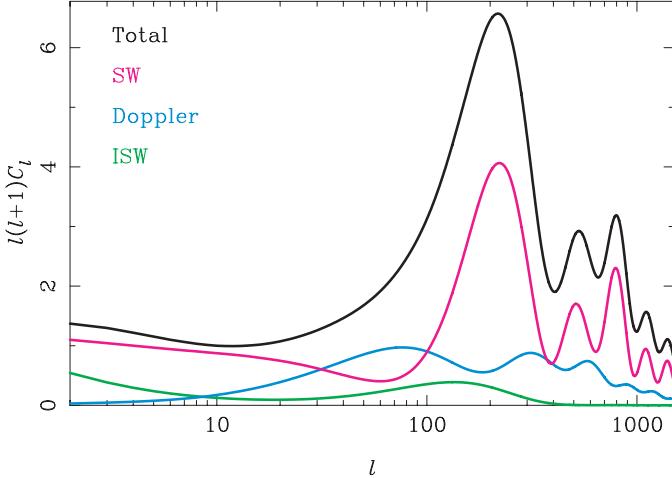


Fig. 9.3 The three contributions to CMB temperature anisotropy and the total angular spectrum [77] calculated for adiabatic perturbations in the spatially flat Λ CDM model with $\Omega_\Lambda = 0.75$, $\Omega_B h^2 = 0.023$, $\Omega_{CDM} h^2 = 0.111$, $h = 0.73$; primordial power spectrum is flat, $n_s = 1$. See Fig. 17.5 for color version.

conformal time interval affects all modes. Thus, the anisotropy at various angular scales is sensitive to changes in the expansion regime.

The first hint towards the existence of CMB temperature anisotropy came from “Relict” experiment [78]. The first reliable measurement of the anisotropy was made by the experiment at COBE satellite (COsmic Background Explorer) [79]. Angular resolution of COBE was rather poor by the present standards, and the range of angular harmonics studied was limited, $l \lesssim 20$. COBE result, nevertheless, showed that the angular spectrum was approximately flat, and determined the quadrupole component

$$\langle Q^2 \rangle = T_0^2 \frac{5}{4\pi} C_2,$$

see the first formula in (9.10) for the numerical coefficient. For flat spectrum, other multipoles are expressed through this component as follows,

$$C_l = \frac{4\pi \langle Q^2 \rangle}{5T_0^2} \frac{6}{(l+1)l}.$$

COBE measurement gave $\langle Q^2 \rangle = (18 \mu\text{K})^2$, so that at $l \lesssim 50$

$$\mathcal{D}_l \simeq 780 \mu\text{K}^2,$$

which is in good agreement with newer data, see Fig. 9.1. It follows from (9.39) that the COBE determination of the amplitude of the primordial power spectrum is $A_\Phi = 100\mathcal{D}_l/(9T_0^2) \simeq 1.16 \cdot 10^{-9}$. Since $A_\Phi = 4A_{\mathcal{R}}/9$, the power spectrum of \mathcal{R} at the length scales somewhat smaller than the present horizon size has the amplitude

$A_{\mathcal{R}} \simeq 2.6 \cdot 10^{-9}$. The present value of this amplitude giving the best fit to all cosmological data, as well as the measured value of n_s are given in (5.37).

Let us see that the ISW and Doppler effects are small for perturbations which were superhorizon at recombination (except for the late ISW effect which is important at $l \lesssim 4$). We begin with the Doppler effect. It follows from (9.22c) that its contribution to the temperature anisotropy is proportional to the velocity $\mathbf{v} \equiv \mathbf{v}_{B\gamma}$ of the baryon-photon component at recombination. We know, however, that the velocity is small for superhorizon modes, see Section 6.1. Hence, the Doppler contribution to the temperature anisotropy is indeed small at low l . This is clearly seen in Fig 9.3.

The late ISW effect is small at $l \gtrsim 4$ because it is generated at late times when dark energy starts to play a role. The relevant length of photon trajectory is roughly comparable to the present horizon size. Hence, the sizeable contribution to the integral (9.22b) comes only from the perturbations of wavelengths comparable to the latter size, i.e., $k\eta_0 \sim 1$, while the contributions of shorter waves average away when integrated along the photon trajectory. Waves with $k\eta_0 \sim 1$ correspond to the smallest multipoles, $l \sim 1$; therefore, the late ISW effect operates at small l only.

It is instructive to find the ISW contribution to multipoles $\Theta_l(k)$ explicitly. To this end, we neglect the difference between Φ and $(-\Psi)$ at late times and use the momentum representation in (9.22b),

$$\Theta_0^{ISW}(\mathbf{n}) = 2 \int_{\eta_r}^{\eta_0} d\eta \int d^3k \Phi'(\eta, \mathbf{k}) e^{i\mathbf{kn}(\eta_0 - \eta)}. \quad (9.40)$$

We consider here large wavelengths of scalar perturbations, so we use the parameterization (4.24),

$$\Phi(\eta, \mathbf{k}) = g(\eta) \cdot \frac{9}{10} \Phi_{(i)}(\mathbf{k}). \quad (9.41)$$

Recall (see Section 4.4), that the function $g(\eta)$ is independent of momentum.³ We insert (9.41) into (9.40) and make use of the expansion (F.35) of the exponential function. Comparing the result with (9.28), (9.29) we find the contribution of modes of momentum k to multipole l ,

$$\Theta_l^{ISW}(k) = \frac{9}{5} \int_{\eta_r}^{\eta_0} d\eta g'(\eta) j_l[(\eta_0 - \eta)k]. \quad (9.42)$$

In fact, the integration here is performed over the time interval corresponding to redshift $z \lesssim 1$. Since $j_l[(\eta_0 - \eta)k]$ rapidly oscillates in time for $(\eta_0 - \eta)k \gtrsim l$ and vanishes for $(\eta_0 - \eta)k < l$, this integral is indeed small for all but very small l .

³We assume here that dark energy is the cosmological constant, $w_{DE} = -1$. In the general case, dark energy itself gets perturbed, and the behavior of the gravitational potentials depends on the momentum of perturbation. The existing bounds on the dark energy equation of state imply that this effect is very small.

The integral (9.42) cannot be evaluated analytically, so we perform an approximate analysis. In the first place, we note that this integral is saturated at η roughly comparable to η_0 . Therefore, the comparison of (9.42) with (9.36) shows that the SW and ISW effects of a given perturbation contribute to *different multipoles*. This means that the interference between the SW and ISW terms is very small, and the ISW contribution can be calculated independently.

Let us make use of the approximate formula (F.21) valid at $l \gtrsim 5$,

$$\int k^2 dk f(k) j_l(k\eta) j_l(k\eta') \approx \frac{\pi}{2\eta^2} f\left(\frac{l + \frac{1}{2}}{\eta}\right) \cdot \delta(\eta - \eta'), \quad (9.43)$$

where $f(k)$ is an arbitrary smooth function. Inserting (9.42) in (9.30) and using (9.43) to perform the momentum integration and one integration over conformal time, we obtain

$$C_l^{ISW} \simeq 4 \cdot \frac{81}{100} \int_{\eta_r}^{\eta_0} d\eta \left(\frac{g'(\eta)}{\eta_0 - \eta} \right)^2 P_\Phi \left(\frac{l + \frac{1}{2}}{\eta_0 - \eta} \right).$$

We introduce dimensionless integration variable $\xi = \eta/\eta_0$ and write for the nearly flat primordial spectrum (5.40)

$$C_l^{ISW} \simeq 8\pi^2 \cdot \frac{81}{100} \frac{A_\Phi}{(k_* \eta_0)^{n_s-1}} \cdot \left(l + \frac{1}{2} \right)^{n_s-4} \cdot \int_{\eta_r/\eta_0}^1 d\xi \cdot (1-\xi)^{2-n_s} \left(\frac{\partial g}{\partial \xi} \right)^2. \quad (9.44)$$

The integral in (9.44) is saturated at $\xi \sim 1$ and it does not depend on l . Hence, the coefficients C_l behave as $C_l \propto l^{n_s-4}$. In particular, for flat primordial power spectrum we have

$$C_l^{ISW} \simeq 8\pi^2 \cdot \frac{81}{100} \frac{A_\Phi}{(l + \frac{1}{2})^3} \int_{\eta_r/\eta_0}^1 d\xi \cdot (1-\xi) \left(\frac{\partial g}{\partial \xi} \right)^2. \quad (9.45)$$

We see that the late ISW contributions to \mathcal{D}_l decays as l increases, in accordance with Fig. 9.3. Making use of the numerical solution for the function $g(z)$ presented in Section 4.4 one can evaluate the integral (9.45) numerically. For the fiducial values of the cosmological parameters,

$$\int_{\eta_r/\eta_0}^1 d\xi \cdot (1-\xi) \left(\frac{\partial g}{\partial \xi} \right)^2 \approx 1.9 \cdot 10^{-2}.$$

This gives

$$C_l^{ISW} \simeq 1.2 \cdot \frac{A_\Phi}{(l + \frac{1}{2})^3}. \quad (9.46)$$

Problem 9.5. Find the coefficients C_l numerically, without using the approximation (9.43). Compare the result with (9.46).

By comparing our result (9.46) with (9.39), we come to the following numerical estimate of the ratio of the SW and late ISW effects,

$$\frac{C_l^{SW}}{C_l^{ISW}} \simeq 0.46 \cdot \frac{(l + \frac{1}{2})^3}{l(l+1)}.$$

We see that for the lowest, and only lowest multipoles these contributions are comparable. This agrees with the numerical result shown in Fig. 9.3.

The approximate formula (9.44) shows that the ISW effect is fairly insensitive to the details of the cosmological evolution: this is an integral effect and it is relevant for a

few multipoles only. Hence, it can at best be used for determining only one dark energy parameter, Ω_Λ . As we have already noticed, correlations of CMB temperature with galaxy distribution are a lot more promising for studying dark energy.

For completeness, let us find the contribution to (9.24) due to tensor perturbations that enter the horizon after recombination. In general, we have for tensor perturbations

$$h_{ij}^{TT}(\eta, \mathbf{k}) = \sum_{A=+, \times} e_{ij}^{(A)} h^{(A)}(\eta, \mathbf{k}) \quad (9.47a)$$

$$h^{(A)}(\eta, \mathbf{k}) = \mathfrak{h}(k, \eta) \cdot h_{(i)}^{(A)}(\mathbf{k}), \quad (9.47b)$$

where the subscript (A) denotes the polarization, $h_{(i)}^{(A)}$ is the primordial amplitude, \mathfrak{h} is the transfer function. We recall (see Section 3.2) that the modes entering the horizon at matter domination obey⁴

$$\mathfrak{h}(k, \eta) = \frac{3j_1(k\eta)}{k\eta}. \quad (9.48)$$

Like in Section 5.4 we assume that $h_{(i)}^{(A)}$ is the Gaussian random field with the two-point correlator (5.44) and power-law dependence of the power spectrum (5.45). We also assume that the tensor tilt n_T is close to zero.

We notice that Eq. (9.24) contains partial derivative with respect to conformal time. Using (9.47a) in (9.24) and performing the Fourier transformation in spatial coordinates, we obtain

$$\Theta_0(\mathbf{n}) = \frac{1}{2} \int d^3k \int_{\eta_r}^{\eta_0} d\eta e^{i(\eta_0 - \eta)k\mathbf{n}\mathbf{n}_k} \sum_A \frac{\partial h^{(A)}}{\partial \eta} \cdot n_i n_m e_{im}^{(A)}(\mathbf{n}_k), \quad (9.49)$$

where, as before, $\mathbf{n}_k \equiv \mathbf{k}/k$. We expand the exponential function in (9.49) in the Legendre polynomials using (F.35), and write for the relative temperature fluctuation

$$\begin{aligned} \Theta_0(\mathbf{n}) &= \frac{1}{2} \int d^3k \sum_{A=+, \times} n_p n_q e_{pq}^{(A)}(\mathbf{n}_k) \cdot h_{(i)}^{(A)}(\mathbf{k}) \\ &\quad \times \sum_{l'=0}^{\infty} (2l' + 1) i^{l'} \cdot P_{l'}(\mathbf{n}\mathbf{n}_k) \cdot \int_{\eta_r}^{\eta_0} d\eta \frac{\partial \mathfrak{h}(k, \eta)}{\partial \eta} j_{l'}[(\eta_0 - \eta)k]. \end{aligned}$$

The coefficients C_l can now be cast in the form

$$C_l = \frac{1}{2l+1} \int d^3k \sum_{m=-l}^l \langle |a_{lm}(\mathbf{k})|^2 \rangle, \quad (9.50)$$

⁴We do not consider here the effects due to the recent accelerated expansion of the Universe. These effects lead to small additional suppression.

where we have in mind that one integration over momenta is trivially performed in the end of calculation because of the δ -function in (5.44). Here

$$\begin{aligned} a_{lm}(\mathbf{k}) &= \frac{1}{2} \int d\mathbf{n} Y_{lm}^*(\mathbf{n}) \sum_{A=+,\times} n_p n_q e_{pq}^{(A)}(\mathbf{n}_k) \cdot h_{(i)}^{(A)}(\mathbf{k}) \\ &\quad \times \sum_{l'=0}^{\infty} (2l'+1) i^{l'} \cdot P_{l'}(\mathbf{n}\mathbf{n}_k) \cdot \int_{\eta_r}^{\eta_0} d\eta \frac{\partial h(k, \eta)}{\partial \eta} j_{l'}[(\eta_0 - \eta) k]. \end{aligned} \quad (9.51)$$

The integrand in (9.50) is independent of the direction of momentum \mathbf{k} , so it can be calculated in any coordinate frame. It is convenient to use the spherical frame (θ, ϕ) with the azimuthal axis directed along \mathbf{n}_k . The convolutions with the polarization tensors in this frame are

$$n_l n_m e_{lm}^{(+)}(\mathbf{n}_k) = \sin^2 \theta \cdot \cos 2\phi, \quad n_l n_m e_{lm}^{(\times)}(\mathbf{n}_k) = \sin^2 \theta \cdot \sin 2\phi. \quad (9.52)$$

We see that the integrand in (9.51) depends explicitly on the azimuthal angle ϕ . Hence, the coefficients a_{lm} are non-trivial, namely

$$\begin{aligned} a_{l,\pm 2}(\mathbf{k}) &= \pi i^l \cdot \sqrt{\frac{2l+1}{4\pi} \frac{(l+2)!}{(l-2)!}} \left(h_{(i)}^{(+)}(\mathbf{k}) \mp i h_{(i)}^{(\times)}(\mathbf{k}) \right) \int_{\eta_r}^{\eta_0} d\eta \frac{\partial h}{\partial \eta} \\ &\quad \times \left(\frac{j_{l-2}[(\eta_0 - \eta) k]}{(2l-1)(2l+1)} + \frac{2j_l[(\eta_0 - \eta) k]}{(2l-1)(2l+3)} + \frac{j_{l+2}[(\eta_0 - \eta) k]}{(2l+1)(2l+3)} \right), \end{aligned} \quad (9.53)$$

other coefficients vanish. The recurrence relation (F.4) for the spherical Bessel functions shows that the sum in parenthesis in (9.53) is equal to

$$\frac{j_l[(\eta_0 - \eta) k]}{(\eta_0 - \eta)^2 k^2}. \quad (9.54)$$

Hence, we obtain finally

$$C_l = \frac{9\pi}{2} \frac{(l+2)!}{(l-2)!} \int \frac{dk}{k} \mathcal{P}_T(k) \cdot \left(\int_{\eta_r}^{\eta_0} d\eta \frac{\partial h(k, \eta)}{\partial \eta} \frac{j_l[(\eta_0 - \eta) k]}{(\eta_0 - \eta)^2 k^2} \right)^2. \quad (9.55)$$

Let us use the result (9.55) to discuss the behavior of the temperature multipole C_l as a function of l . For modes entering the horizon after recombination and thus dominating at $l \lesssim 50$, the derivative $\partial h/\partial \eta$ is small at $\eta = \eta_r$. It becomes sizeable later, at $\eta \sim k^{-1}$. Therefore, the integrand in parenthesis in (9.55) is non-negligible at $k(\eta_0 - \eta) \sim l$ and at the same time $k\eta \sim 1$. In this regime $j_l[(\eta_0 - \eta) k] \sim l^{-1}$, so the integral behaves as l^{-3} . The factor with factorials in (9.55) grows as l^4 . In the case of flat spectrum, $\mathcal{P}_T \approx \text{const}$, the remaining integration over momentum k does not yield extra dependence on l . As a result, the multipoles C_l decay as l^{-2} . Hence, the function $\mathcal{D}_l \propto l^2 C_l$ is approximately independent of l in the interval $2 \lesssim l < 50$. This behavior is similar to the SW effect of the scalar perturbations, as illustrated in Fig. 9.4. The gravity wave contribution rapidly decays at $l \gtrsim 100$ as l grows, since it comes from the modes entering the horizon before recombination,

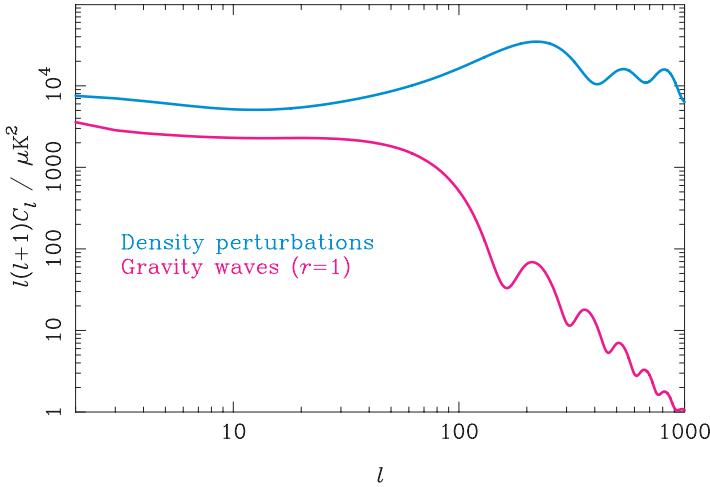


Fig. 9.4 Contributions of scalar and tensor perturbations (upper and lower curves, respectively) to the angular spectrum of CMB temperature anisotropy [77]. For illustration purpose, the primordial scalar and tensor spectra are flat, $n_s = 1$, $n_T = 0$, and have equal amplitude, $\mathcal{P}_R = \mathcal{P}_T$, i.e., $r = 1$ (see Section 5.4). This value of r is, of course, unrealistic.

and the amplitude of these modes at and after recombination decreases with k . We note, though, that for blue primordial tensor spectrum, i.e., for large positive n_T , the contribution of short waves is enhanced, and hence C_l may not be small at large l .

Let us find C_l approximately at $l \gg 1$. It is clear that the ratio η/η_0 is a small parameter in the integral in parenthesis in (9.55). We use the asymptotic formula (F.15) for the function $j_l[k(\eta_0 - \eta)]$, which can be written as follows,

$$j_l[k(\eta_0 - \eta)] = \frac{1}{l\sqrt{u}(u^2 - 1)^{1/4}} \cos \left\{ \left(l + \frac{1}{2} \right) \phi \left[u \cdot \left(1 - \frac{\eta}{\eta_0} \right) \right] \right\}, \quad (9.56)$$

where

$$u = \frac{k\eta_0}{l + 1/2}, \quad \phi(x) = \sqrt{x^2 - 1} - \arccos \left(\frac{1}{x} \right) - \frac{\pi}{4}. \quad (9.57)$$

When writing the slowly varying factor in (9.56) we have set $(l + 1/2) \simeq l$ and used the fact that $\eta/\eta_0 \ll 1$. The ratio η/η_0 cannot be neglected in the argument of cosine, since the phase contains the large factor $(l + 1/2)$. In other words, the combination $(l + 1/2) \cdot u \cdot (\eta/\eta_0) \equiv k\eta$ is of order 1. This gives rise to strong dependence of the phase on η . On the other hand, the next term in the expansion of the phase in η is small, $(l + 1/2) \cdot u^2 \cdot (\eta/\eta_0)^2 = (k\eta)^2/(l + 1/2) = O(l^{-1})$. Hence, the phase is approximated as follows,

$$\left(l + \frac{1}{2} \right) \phi \left[u \left(1 - \frac{\eta}{\eta_0} \right) \right] = \left(l + \frac{1}{2} \right) \phi(u) - l \frac{\eta}{\eta_0} u \phi'(u) + O \left(\frac{1}{l} \right),$$

where $u\phi'(u) = \sqrt{u^2 - 1}$. Therefore, we have for the spherical Bessel function

$$\begin{aligned} j_l[k(\eta_0 - \eta)] &= \frac{1}{l\sqrt{u}(u^2 - 1)^{1/4}} \left\{ \cos \left[\left(l + \frac{1}{2} \right) \phi(u) \right] \cos \left(l \frac{\eta}{\eta_0} \sqrt{u^2 - 1} \right) \right. \\ &\quad \left. + \sin \left[\left(l + \frac{1}{2} \right) \phi(u) \right] \sin \left(l \frac{\eta}{\eta_0} \sqrt{u^2 - 1} \right) \right\}, \end{aligned} \quad (9.58)$$

so the integral in parenthesis in (9.55) takes the following form,

$$\frac{1}{l^3 u^{5/2} (u^2 - 1)^{1/4}} \left\{ \cos \left[\left(l + \frac{1}{2} \right) \phi(u) \right] I_c + \sin \left[\left(l + \frac{1}{2} \right) \phi(u) \right] I_s \right\},$$

where

$$I_c = \int_{\eta_r}^{\infty} d\eta \frac{\partial \mathfrak{h}(k, \eta)}{d\eta} \cos \left(l \frac{\eta}{\eta_0} \sqrt{u^2 - 1} \right) \quad (9.59a)$$

$$I_s = \int_{\eta_r}^{\infty} d\eta \frac{\partial \mathfrak{h}(k, \eta)}{d\eta} \sin \left(l \frac{\eta}{\eta_0} \sqrt{u^2 - 1} \right). \quad (9.59b)$$

We extended the integration to infinity having in mind that $ul \gg 1$. Now, when integrated in (9.55), rapid oscillations in k (in u , in other words) get averaged,

$$\begin{aligned} \cos^2 \left[\left(l + \frac{1}{2} \right) \phi(u) \right] &\rightarrow \frac{1}{2}, \quad \sin^2 \left[\left(l + \frac{1}{2} \right) \phi(u) \right] \rightarrow \frac{1}{2}, \\ \cos \left[\left(l + \frac{1}{2} \right) \phi(u) \right] \cdot \sin \left[\left(l + \frac{1}{2} \right) \phi(u) \right] &\rightarrow 0, \end{aligned} \quad (9.60)$$

so we obtain at $l \gg 1$ for the nearly flat primordial spectrum

$$C_l = \frac{\pi A_T}{4} \frac{1}{(\eta_0 k_*)^{n_T} l^{2-n_T}} \int_1^{\infty} \frac{du}{u^{6-n_T} \sqrt{u^2 - 1}} (I_c^2 + I_s^2), \quad (9.61)$$

where we used the fact that the integration over k is effectively performed in the region $k\eta_0 > l$. Note that this representation is valid for modes entering the horizon both at matter domination and radiation domination. Corrections to the result (9.61) are of order l^{-1} .

For low multipoles we discuss here, $l \ll \eta_0/\eta_r$, the lower limit of integration in (9.59) can be set equal to zero. Making use of (9.48), we find that the integrals I_c and I_s are independent of l at these l , so the angular spectrum behaves as $l(l+1)C_l \propto l^{n_T}$. For the flat primordial spectrum, $n_T = 0$, the anisotropy spectrum is also flat at $l \lesssim 50$, in accordance with the above discussion. In fact, the integrals (9.59) have explicit analytic forms,

$$\begin{aligned} I_c &= \int_0^{\infty} d\zeta \frac{\partial}{\partial \zeta} \left[\frac{3j_1(\zeta)}{\zeta} \right] \cos \left(\zeta \frac{\sqrt{u^2 - 1}}{u} \right) = \frac{1}{2} - \frac{3}{2u^2} - \frac{3\sqrt{u^2 - 1}}{4u^3} \log \frac{u - \sqrt{u^2 - 1}}{u + \sqrt{u^2 - 1}}, \\ I_s &= \int_0^{\infty} d\zeta \frac{\partial}{\partial \zeta} \left[\frac{3j_1(\zeta)}{\zeta} \right] \sin \left(\zeta \frac{\sqrt{u^2 - 1}}{u} \right) = -\frac{3\pi}{4} \frac{\sqrt{u^2 - 1}}{u^3}. \end{aligned}$$

Then the integral (9.61) can be evaluated numerically, and for multipoles $1 \ll l \lesssim 50$ we find for flat primordial spectrum

$$l(l+1)C_l \simeq 0.39A_T.$$

The ratio of the SW contribution of scalar perturbations to the tensor contribution is $C_l^{SW}/C_l^T \simeq 1.46 A_\Phi/A_T = 3.28 A_R/A_T = 3.28/r$. This is consistent with the result of the complete numerical calculation shown in Fig. 9.4.

At $l > \eta_0/\eta_r \sim 50$ the function $\mathcal{D}_l = T_0^2 l(l+1) C_l/(2\pi)$ not only decreases as l grows, but also oscillates, as seen in Fig. 9.4. The formal reason for these oscillations is the oscillatory character of integrals (9.59). We do not consider this phenomenon here; it is analogous to the oscillations in \mathcal{D}_l caused by scalar perturbations. The latter are discussed in details in Section 9.2.3. The behavior of tensor contribution at large l is studied in problem 9.10.

9.2.3 Intermediate angular scales

We now turn to the effect of the adiabatic scalar perturbations whose wavelength is smaller, but not very much smaller than the horizon at recombination, i.e., modes with $k\eta_r \gtrsim 1$. “Not very much smaller” here means that the tight coupling approximation for baryon-photon plasma, which is used in this Section, is still valid. In terms of angular harmonics, this implies

$$l \lesssim 1000.$$

The results of this Section are of relevance for smaller angular scales, $l \gtrsim 1000$, as well: the effects beyond the tight coupling approximation lead to the suppression of multipoles C_l and damping of oscillations, but they affect weakly other properties of the angular spectrum, such as the positions of the peaks. Small angles are considered in Section 9.3.

There are two particularly relevant momentum scales for perturbations studied here, $r_s^{-1}(\eta_r)$ and $r_s^{-1}(\eta_{eq})$, where

$$r_s(\eta) = \int_0^\eta \frac{d\eta}{\sqrt{3(1 + R_B(\eta))}}$$

is the sound horizon at time η . Modes of these momenta enter the sound horizon at recombination and radiation-matter equality, respectively. As we have seen and will see again soon, the corresponding multipoles are $l^{(r)} = \eta_0 r_s^{-1}(\eta_r)$ and $l^{(eq)} = \eta_0 r_s^{-1}(\eta_{eq})$. Therefore, it is useful to find accurate estimates for their values; this is particularly important for $l^{(r)}$.

The sound horizon size $r_s(\eta_r)$ has been calculated in Section 7.1.2; the expression for it is given in (7.33) and numerical value in (7.36). Hence, we obtain

$$l^{(r)} = r_s^{-1}(\eta_r) \eta_0 = \frac{\sqrt{3}}{I_1} \frac{\eta_0}{\eta_r}, \quad (9.62)$$

where the ratio η_0/η_r is given by (2.20) and I_1 is defined in (7.34). Note that $l^{(r)}$ is inversely proportional to the present angular size of the sound horizon at recombination. Numerically, with fiducial values of the cosmological parameters, we get

$$l^{(r)} = 97. \quad (9.63)$$

We do not need the very precise estimate for $l^{(eq)}$, simply because the transition from radiation to matter domination is an extended process whose duration is at least the Hubble time. It suffices to set $r_s(\eta_{eq}) \simeq \eta_{eq}/\sqrt{3}$, which is valid at accuracy of about 10%. Hence,

$$l^{(eq)} \simeq \frac{\sqrt{3}\eta_0}{\eta_{eq}} \sim 210,$$

where we have used (2.23) for the numerical estimate.

So, let us consider perturbations that enter the sound horizon before recombination. The corresponding angular harmonics have

$$l \gg l^{(r)}. \quad (9.64)$$

Our results will be also valid *at qualitative level* for smaller multipoles, $\eta_0/\eta_r \lesssim l \lesssim l^{(r)}$, i.e., $50 \lesssim l \lesssim 100$. However, the latter interval is difficult to study analytically, since the approximations used in Sections 6.1 and 6.2 do not apply to the relevant wavelengths. In particular, our estimates will only be qualitative for the region of the first peak in the angular spectrum. Still, the analytical description, even limited, is useful, since it enables one to better understand the dependence of the features in the angular spectrum on the cosmological parameters.

Let us discuss the SW, Doppler and ISW terms in (9.22) separately, leaving aside the interference between them for the time being. Shortly before the last scattering epoch, the combination entering the SW effect (9.22a) is given by (see (6.43) and (6.44))

$$\frac{1}{4}\delta_\gamma(\eta) + \Phi(\eta) = \left[A(k, \eta) \cos \left(k \int_0^\eta \frac{d\eta'}{\sqrt{3(1 + R_B(\eta'))}} \right) - B(k, \eta) \right] \cdot \Phi_{(i)}, \quad (9.65)$$

where $A(k, \eta)$ and

$$B(k, \eta) = R_B(\eta)F(k, \eta) \quad (9.66)$$

are positive functions that vary slowly in time. Their expressions valid at large k are (see (6.15), (6.39))

$$A = \frac{3}{2} \frac{1}{[1 + R_B(\eta_r)]^{1/4}} \quad (9.67)$$

$$B(k, \eta) = 81I^2R_B(\eta) \frac{\Omega_{CDM}}{\Omega_M} (1 + z_{eq}) \frac{\log(0.2k\eta_{eq})}{(k\eta_0)^2}. \quad (9.68)$$

We recall that the first term in (9.65) is due to the acoustic oscillations in the baryon-photon plasma, while the second term is due to the gravitational potential produced by dark matter and equals $(-R_B\Phi)$. Note that the partial cancellation between the non-oscillating term in $\delta_\gamma/4$ (see (6.35)) and Φ results in the fact that the second term is proportional to the baryon density (more precisely, to $R_B = 3\rho_B/(4\rho_\gamma)$).

We see that the combination responsible for the SW effect oscillates in time over slowly varying negative average value $-B\Phi_{(i)} = -R_B\Phi$ (when discussing signs, we

assume for definiteness that $\Phi_{(i)}$ is positive). These oscillations have definite phase. They yield the oscillations in the angular spectrum clearly visible in Fig. 9.3. Indeed, in the approximation of the instantaneous photon decoupling, the SW term is given by the value of the function (9.65) at $\eta = \eta_r$. Since the oscillation frequency depends on momentum k , the value of this function at $\eta = \eta_r$ oscillates as a function of momentum,

$$\frac{1}{4}\delta_\gamma(\eta_r) + \Phi(\eta_r) = \{A \cos[kr_s(\eta_r)] - B(k)\}\Phi_{(i)}, \quad (9.69)$$

where $A = A(\eta_r)$, $B(k) = B(k, \eta_r)$. As we already noticed, there is the approximate correspondence between the multipole number and momentum, $l \leftrightarrow k\eta_0$, hence the oscillatory behavior of C_l as function of l . We will see that there are oscillations of the Doppler contribution too, as well of the total angular spectrum, see Fig. 9.3.

The formula for the SW contribution to the multipoles $\tilde{\Theta}_l(\mathbf{k})$ is analogous to (9.34). Namely,

$$\tilde{\Theta}_l^{SW}(\mathbf{k}) = \left(\Phi(\eta_r) + \frac{1}{4}\delta_\gamma(\eta_r) \right) j_l[k(\eta_0 - \eta_r)]. \quad (9.70)$$

Since the SW contribution to multipoles C_l (neglecting the interference with other contributions) is determined by the quantity (9.70) *squared*, the positions of the *maxima* in C_l are determined by the positions of *both maxima and minima* of cosine in (9.69). These maxima and minima are at $k \simeq \pi n/r_s$, where $n = 1, 2, \dots$. Accordingly, the SW effect produces peaks in the angular spectrum near $l_n^{SW} \sim n\pi l^{(r)}$, see details below. Thus, the positions of the acoustic peaks in the angular spectrum fit the expected relation $\pi/l_n^{SW} \sim \Delta\theta_s/n$, where

$$\Delta\theta_s = \frac{I_1}{\sqrt{3}} \cdot \frac{\eta_r}{\eta_0}$$

is the angle at which the sound horizon at recombination is seen today.

Figure 9.3 shows that modulo overall decline of the angular spectrum at increasing l , the peaks with odd n have a tendency to be higher than the peaks with even n at moderate n . This is due to the non-oscillating term in (9.69): both oscillating and non-oscillating terms are negative at odd n , and there is the constructive interference between them, while they have opposite signs at even n , and the interference is destructive. To see that this effect is indeed numerically important, let us make use of the crude estimate (6.42). It gives for the intermediate momenta that $A(\eta_r) \sim [1/3 + R_B(\eta_r)]$, $B \sim R_B(\eta_r)$, so these coefficients are of one and the same order of magnitude. Of course, one cannot trust the estimate (6.42) at quantitative level, but it shows that there is indeed interference of two comparable terms at intermediate momenta. This effect is strong for a few first peaks, and afterwards it disappears because of the decay of the non-oscillating part, $B(k) \propto k^{-2}$.

Let us consider in more details the positions of maxima and minima of the SW contribution, still neglecting the interference with other effects. Making use of (9.30), (9.69)

and (9.70) we write

$$C_l = 4\pi \int \frac{dk}{k} \mathcal{P}_\Phi(k) \cdot [A \cos kr_s - B(k)]^2 j_l^2[k(\eta_0 - \eta_r)], \quad (9.71)$$

where $r_s = r_s(\eta_r)$. We will see that the integrand in (9.71) is large in a region whose size is at least $\Delta k \sim 1/r_s$, so that $\eta_0 \Delta k \gg l^{1/3}$. At $k\eta_0 \gtrsim l$, the spherical Bessel function oscillates with the period of at most $l^{1/3}$, while at $k\eta_0 < l$ it is nearly zero, see Appendix F. Hence, the region $k < l/\eta_0$ gives no contribution to the integral, while at $k > l/\eta_0$ we can use the asymptotic formula (F.15) and average cosine squared by setting it equal to 1/2. As a result we obtain, e.g., for the flat primordial spectrum,

$$C_l^{SW} = 2\pi A_\Phi \int_{l/\eta_0}^\infty \frac{dk}{k} \cdot \frac{1}{k\eta_0 \sqrt{k^2 \eta_0^2 - l^2}} (A \cos kr_s - B)^2, \quad (9.72)$$

where we neglected corrections of order $1/l$ and η_r/η_0 . We obtain after obvious algebra

$$\begin{aligned} C_l^{SW} = \frac{2\pi}{l^2} A_\Phi & \left\{ \frac{1}{2} c_1 A^2(l) + \frac{8}{15} c_2 B^2(l) \right. \\ & \left. + \int_1^\infty \frac{du}{u^2 \sqrt{u^2 - 1}} \left[\frac{1}{2} A^2 \cos \left(2 \frac{lr_s}{\eta_0} u \right) - 2AB \cos \left(\frac{lr_s}{\eta_0} u \right) \right] \right\}, \end{aligned} \quad (9.73)$$

where $A(l) = A(k = l/\eta_0)$, $B(l) = B(k = l/\eta_0)$, $u = k\eta_0/l$, and constants c_1 and c_2 are of order 1 (they are determined by the dependence of functions $A(k, \eta_r)$ and $B(k, \eta_r)$ on k ; according to (9.67) and (9.68), one has $c_1 = c_2 = 1$ at large l). The first two terms here monotonically decrease with l , while the third, integral term yields the oscillations of C_l^{SW} . To estimate the latter effect, we consider lr_s/η_0 as large parameter, i.e., we take

$$\frac{lr_s}{\eta_0} = \frac{l}{l^{(r)}} \gg 1. \quad (9.74)$$

Then the integrals with the oscillating functions in (9.73) are saturated in a small region near $u = 1$. The size of this region is of order $\Delta u \sim l^{(r)}/l$, i.e., $\Delta k \sim 1/r_s$. The factor u^{-2} can be replaced by 1 in this region, and $\sqrt{u^2 - 1} \simeq \sqrt{2}\sqrt{u - 1}$. In this way we arrive at the known integral

$$\int_1^\infty \frac{du}{\sqrt{u - 1}} \cos \alpha u = \sqrt{\frac{\pi}{\alpha}} \cos \left(\alpha + \frac{\pi}{4} \right).$$

Hence, the oscillating integral is estimated as

$$\begin{aligned} \int_1^\infty \frac{du}{u^2 \sqrt{u^2 - 1}} & \left[\frac{1}{2} A^2 \cos \left(2 \frac{l}{l^{(r)}} u \right) - 2AB \cos \left(\frac{l}{l^{(r)}} u \right) \right] \\ & \simeq \sqrt{\frac{\pi l^{(r)}}{2l}} \left[\frac{1}{2\sqrt{2}} A^2(l) \cos \left(2 \frac{l}{l^{(r)}} + \frac{\pi}{4} \right) - 2A(l)B(l) \cos \left(\frac{l}{l^{(r)}} + \frac{\pi}{4} \right) \right]. \end{aligned} \quad (9.75)$$

In the first place, we see that the amplitude of oscillations decreases as l increases, although not very fast. Now, the first and the second terms have maxima at

$$l_n^{SW} = \left(\frac{7\pi}{8} + \pi(n-1) \right) l^{(r)}, \quad n = 1, 2, \dots \quad (9.76)$$

and

$$l = \left(\frac{3\pi}{4} + 2\pi n' \right) l^{(r)}, \quad n' = 0, 1, 2, \dots,$$

respectively; the amplitude of second term decreases at large l as l^{-2} , while the amplitude of the first term tends to a constant, see (9.67) and (9.68). For odd n , both terms have nearby maxima, while the maxima of the first term with even n are near the minima of the second term. Hence, multipoles C_l have maxima near $l = l_n^{SW}$; these maxima are enhanced due to the second term at odd n and suppressed at even n . This is precisely the effect of the interference between the two contributions in (9.69).

The approximation (9.74) is poor in the region of the first peak. Furthermore, the shape of the first peak is strongly affected by the dependence of $B(k, \eta_r)$ on momentum. Nevertheless, we can make a qualitative conclusion that the position of the first peak is somewhere between $3\pi l^{(r)}/4 \simeq 230$ and $7\pi l^{(r)}/8 \simeq 265$. This conclusion is supported by the numerical analysis, see Fig. 9.3. At large l , the first term in (9.75) dominates, and the positions of the peaks are given by (9.76); numerically

$$l_n^{SW} \simeq 265 + 305n.$$

This is indeed seen in both the numerical result and experimental data.

We note that the region $u \approx 1$ that dominates in the integral (9.75) corresponds to momenta k close to l/η_0 . This means that the main contribution to the oscillatory picture comes from perturbations of momenta almost perpendicular to the line of sight. More precisely, the integral in (9.75) is saturated in the region $(u - 1) \sim l^{(r)}/l \ll 1$, which corresponds to $\mathbf{n}_k \cdot \mathbf{n} = \cos \theta \sim l^{(r)}/l \ll 1$, where θ is the angle between the momentum \mathbf{k} and the direction of observation \mathbf{n} .

We now turn to the Doppler effect (9.22c). For modes studied in this Section, velocity potential is determined by the photon density contrast, $kv_B = 3/(4k) \cdot \delta'_\gamma$, see (6.40). Making use of Eq. (6.44) and neglecting the time-derivatives of $A(\eta)$ and $R_B F(\eta)$, we obtain for $k > k_s^{(eq)}$ that

$$kv_B(\eta) = -3u_s \Phi_{(i)} A(k, \eta) \cdot \sin \left(k \int_0^\eta \frac{d\eta'}{\sqrt{3(1 + R_B(\eta'))}} \right), \quad (9.77)$$

where $3u_s = \sqrt{3}/[1 + R_B(\eta_R)]$. The oscillations in the velocity potential also have definite phase.

The calculation of the Doppler contribution to the multipoles $\tilde{\Theta}_l(\mathbf{k})$ is similar to that performed in Section 9.2.2. It gives

$$\tilde{\Theta}_l^D(\mathbf{k}) = -kv_B(\eta_r) \cdot j'_l[k(\eta_0 - \eta_r)] \quad (9.78)$$

where we recalled our convention (9.25) concerning the sign of \mathbf{n} ; prime here denotes the derivative of the spherical Bessel function with respect to its argument. Since the phase of oscillations (9.77) is shifted by $\pi/2$ with respect to the phase in (9.65), the maxima of the Doppler contribution to C_l are shifted with respect to the maxima of the SW effect. Unlike in (9.65), the expression (9.77) oscillates about zero average value, so the Doppler oscillations are unmodulated.

Again considering the flat primordial spectrum for definiteness, and performing the calculation similar to that leading to (9.72), we obtain the Doppler contribution to C_l in the absence of the interference with other effects,

$$C_l^D = 2\pi [3u_s(\eta_r)]^2 A_\Phi \int_l^\infty \frac{dk}{k} \cdot \frac{\sqrt{k^2 \eta_0^2 - l^2}}{(k \eta_0)^3} A^2 \sin^2[kr_s(\eta_r)].$$

We have used the asymptotic formula (F.22) for the derivative of the spherical Bessel function. The expression analogous to (9.73) is

$$C_l^D = \frac{2\pi}{l^2} A_\Phi \left\{ \frac{3u_s^2}{2} c_3 A^2 - \int_1^\infty \frac{du \sqrt{u^2 - 1}}{u^4} \frac{(3u_s)^2}{2} A^2 \cos \left(2 \frac{lr_s}{\eta_0} u \right) \right\}, \quad (9.79)$$

where c_3 is of order 1, and $c_3 = 1$ at large l . It is again useful to consider the regime (9.74). The integrand in the second term in (9.79) is small at $u \approx 1$, so the oscillating part is suppressed by the parameter $l^{(r)}/l$. We integrate the oscillatory integral by parts, setting $A \approx \text{const}$, and obtain to the leading order in $l^{(r)}/l$

$$\begin{aligned} - \int_1^\infty \frac{du \sqrt{u^2 - 1}}{u^4} \frac{(3u_s)^2}{2} A^2 \cos \left(2 \frac{lr_s}{\eta_0} u \right) &\simeq \frac{(3u_s)^2}{2} A^2 \frac{l^{(r)}}{2l} \int_1^\infty \frac{du}{u^3 \sqrt{u^2 - 1}} \sin \left(2 \frac{l}{l^{(r)}} u \right) \\ &\simeq \frac{(3u_s)^2 l^{(r)}}{2l} \frac{A^2}{4} \sqrt{\frac{\pi l^{(r)}}{l}} \cos \left(2 \frac{l}{l^{(r)}} - \frac{\pi}{4} \right). \end{aligned}$$

The amplitude of oscillations is suppressed in comparison with the first term in (9.75) by the factor $(3u_s)^2 l^{(r)} / (2l) \simeq 1.0 \cdot (l^{(r)}/l)$, which is small even for the first peaks. The positions of the peaks are shifted as compared to (9.76),

$$l^D \simeq \left(\frac{\pi}{8} + \pi n \right) l^{(r)} \simeq 40 + 305n.$$

This result is in agreement with the numerical data shown in Fig. 9.3.

Notably, the original magnitudes of the Doppler and SW effects, (9.77) and (9.65), are of the same order of magnitude (the Doppler term is even enhanced by a factor of $3u_s \simeq 2$). Nevertheless, the oscillatory pattern of the anisotropy spectrum is dominated by the SW effect. The reason is that the Doppler contribution of a perturbation of momentum \mathbf{k} is proportional to the projection of the velocity \mathbf{v} on the line of sight, i.e., to $\mathbf{n}\mathbf{n}_k = \cos\theta$. This suppresses the entire Doppler contribution. It is also important that the oscillating SW contribution comes mainly from perturbations of momenta perpendicular to the line of sight. Similar effect is absent for the Doppler effect for the above reason. The Doppler part in a multipole C_l comes from rather wide range of absolute values and directions of momenta, so the oscillations get partially washed out.

Problem 9.6. Show that for $(n_s - 1) \ll 1$ the dependence on n_s of the SW and Doppler effects in the anisotropy is indeed given by the formula (9.38).

Problem 9.7. Show that the primordial spectrum (5.36) with weakly running tilt, $dn_s/d\log k \ll 1$, yields, instead of (9.38), the following behavior,

$$C_l[n_s(k)] = \left(\frac{l}{k_* \eta_0} \right)^{n_s - 1 + \frac{dn_s}{d\log k} \log \frac{l}{k_* \eta_0}} \cdot C_l(n_s = 1). \quad (9.80)$$

Let us now turn to the early ISW effect (9.22b). It is clear from Fig. 9.3 that it is considerable near the first peak only. The situation here is similar to what

happens with the late ISW effect. Namely, the early ISW effect is sizeable just after recombination, when the Universe is not quite at matter domination yet. At that time, perturbations with $k\eta_r \ll 1$ do not evolve, while perturbations with $k\eta_r \gg 1$ are averaged out when integrated along the photon trajectory. Hence, the non-negligible contribution to the early ISW effect is due to perturbations with $k\eta_r \sim 1$, which precisely corresponds to the region of the first acoustic peak. The early ISW contribution to C_l is suppressed by the parameter

$$\left[\frac{\rho_{rad}}{\rho_M}(\eta_r) \right]^2 = \left(\frac{1+z_r}{1+z_{eq}} \right)^2 \simeq 0.12.$$

Indeed, the amplitude of the early ISW effect is non-zero only because $\rho_{rad} \neq 0$. In other words, this amplitude is proportional to $\rho_{rad}/\rho_M(\eta_r)$, and the contribution to C_l is proportional to this ratio squared.

We now recall that the total anisotropy is given by the integral over momenta of $|\tilde{\Theta}_l^{SW}(\mathbf{k}) + \tilde{\Theta}_l^{ISW}(\mathbf{k}) + \tilde{\Theta}_l^D(\mathbf{k})|^2$. Hence, there is the interference between different terms. The interference between the SW and Doppler effects is very small. Indeed, it follows from (9.70) and (9.78) that the interference term in C_l is the integral over momenta of a slowly varying function multiplied by $j_l[k(\eta_0 - \eta_r)] \cdot j'_l[k(\eta_0 - \eta_r)]$. The latter function rapidly oscillates near zero value at momenta we discuss, so this integral is suppressed at least by the factor $(k\eta_0)^{-1} \sim l^{-1}$. On the other hand, there is the interference of the ISW effect with both the Doppler and SW effects near the first peak. Roughly speaking, since the early ISW effect is produced at $\eta \sim \eta_r$, its contribution to the multipoles is estimated as

$$\begin{aligned} \tilde{\Theta}_l^{ISW}(\mathbf{k}) &= 2 \int_{\eta_r}^{\eta_*} \Phi' d\eta j_l[k(\eta_0 - \eta)] \simeq 2 j_l[k(\eta_0 - \eta_r)] \int_{\eta_r}^{\eta_*} \Phi' d\eta \\ &= 2(\Phi_{MD} - \Phi(\eta_r)) j_l[k(\eta_0 - \eta_r)], \end{aligned}$$

where Φ_{MD} is the gravitational potential late at matter domination and η_* is an arbitrary moment of time at that epoch. By comparing this expression with (9.70) we see that the ISW and SW effects are roughly proportional to each other, and both have negative sign near the first peak (since the potential Φ decreases in time at the epoch we consider). Hence, there is the constructive interference between these effects.

Let us sketch more accurate analysis of the early ISW effect. We use the formula (9.58) and write

$$\begin{aligned} \tilde{\Theta}_l^{ISW}(\mathbf{k}) &= \frac{1}{l\sqrt{u}(u^2 - 1)^{1/4}} \\ &\times \left\{ \cos \left[\left(l + \frac{1}{2} \right) \phi(u) \right] I_c^{ISW} + \sin \left[\left(l + \frac{1}{2} \right) \phi(u) \right] I_s^{ISW} \right\}, \quad (9.81) \end{aligned}$$

where

$$I_c^{ISW} = 2 \int_{\eta_r}^{\eta_0} d\eta \Phi' \cos \left(l \frac{\eta}{\eta_0} \sqrt{u^2 - 1} \right),$$

$$I_s^{ISW} = 2 \int_{\eta_r}^{\eta_0} d\eta \Phi' \sin \left(l \frac{\eta}{\eta_0} \sqrt{u^2 - 1} \right),$$

where we use the notations introduced in 9.2.2. We see that the oscillating terms with I_c^{ISW} and I_s^{ISW} in (9.81) have the same phases as $j_l[k(\eta_0 - \eta_r)]$ and $j'_l[k(\eta_0 - \eta_r)]$, respectively. Therefore, the terms with I_c^{ISW} and I_s^{ISW} interfere with the SW effect and Doppler effect. The decay of the ISW effect at higher l is also seen from the above formula.

Problem 9.8. *Making use of (9.81), show that there is indeed the constructive interference of the ISW and SW effects near the first peak.*

Problem 9.9. *Find the behavior of the ISW effect at $l \gg l^{(r)}$.*

Problem 9.10. *Find the contribution of tensor modes to C_l at $l > \eta_0/\eta_{eq} \sim 100$, assuming that the primordial tensor perturbations are Gaussian and have flat power spectrum. Find both monotonic and oscillating part of this contribution. Hints: Use Eqs. (3.10), (9.59) and (9.61); use the results of Section 2.1.2 to find the evolution of the scale factor. Take into account that the integrals in (9.59) are saturated at the lower limit of integration.*

There is yet another phenomenon in the real Universe that is important for CMB temperature anisotropy (and also for CMB polarization) at intermediate and small angular scales. This is scattering of CMB photons after recombination [18, 80]. Just after the last scattering epoch, the Universe is transparent to photons. Later on, however, hydrogen in the Universe gets ionized again, and the Universe becomes less transparent. This occurs at the epoch of the formation of first stars, at redshift $z = z_{rei} \sim 10$. Baryon cooling accompanying the star formation proceeds via the emission of ultraviolet light which ionizes hydrogen. Some relic photons scatter off free electrons at that time and later. From the viewpoint of CMB temperature anisotropy, this epoch corresponds to very large angular scales, plagued by the cosmic variance. So, the *details* of the reionization are unimportant for us, and we use the approximation of instantaneous and homogeneous reionization. This approximation implies that starting from redshift $z = z_{rei}$, there is a non-vanishing probability of photon rescattering without frequency shift⁵ but with random change of the direction of propagation. The probability to “live through” that epoch without rescattering is $e^{-\tau_{rei}}$, where

$$\tau_{rei} = \sigma_T \int_{t_{rei}}^{t_0} dt \cdot n_e$$

⁵Since the number of photons of given frequency does not change, the energy spectrum of CMB remains Planckian.

is the optical depth in the reionized Universe, n_e is the number density of free electrons and the integration is performed along photon world line. In our approximation, the two parameters, optical depth τ_{rei} and redshift z_{rei} determine completely the modification of the angular spectrum of CMB at the reionization epoch.

The effect of reionization is described on the basis of the following simple observation. Photons that rescatter at $z = z_{rei}$ at given place in the Universe have last scattered at recombination anywhere on a sphere of coordinate radius $(\eta_{rei} - \eta_r) \simeq \eta_{rei}$. Since the direction of photon propagation changes randomly, photons coming to an observer from the given direction gather from the whole of that sphere. Hence, the contribution of rescattered photons to anisotropy at angular scales smaller than $\Delta\theta_{rei} = \eta_{rei}/\eta_0$ is washed out, and the remaining anisotropy is due to photons that have not rescattered. On the other hand, the anisotropy at angular scales larger than $\Delta\theta_{rei}$ remains intact, since the regions of the coordinate size η_{rei} at recombination are not resolved at these angular scales. Generally speaking, there is also the Doppler effect due to the motion of the baryon-electron plasma at the reionization epoch. This effect is very small, however, because the velocities of perturbations in baryon-electron fluid are very small at large scales. At small distance scales, structure formation is important and the electron velocities are higher, but these scales do not contribute to CMB anisotropy because of the large thickness of the rescattering sphere.

This observation immediately gives

$$a_{lm} = a_{lm}^{(0)}, \quad l < l_{rei} \quad (9.82a)$$

$$a_{lm} = e^{-\tau_{rei}} a_{lm}^{(0)}, \quad l > l_{rei}, \quad (9.82b)$$

where $a_{lm}^{(0)}$ are the anisotropy coefficients in the absence of reionization, and

$$l_{rei} = \frac{\pi}{\Delta\theta_{rei}} = \frac{\pi\eta_0}{\eta_{rei}}, \quad (9.83)$$

We see that the entire effect of reionization reduces to the suppression of temperature multipoles C_l with $l > l_{rei}$ by the factor $e^{-2\tau_{rei}}$.

Until now we have considered the parameters τ_{rei} and z_{rei} as independent. The parameter τ_{rei} determines the suppression factor $e^{-2\tau_{rei}}$, while the parameter z_{rei} determines the value of l_{rei} above which this suppression occurs. To find the numerical value of l_{rei} we use the estimate (9.83), recall the relation (2.16) and write for matter domination

$$\eta_{rei} = \frac{2}{a_0 H_0 \sqrt{1 + z_{rei}}}.$$

We obtain

$$l_{rei} \simeq \pi \sqrt{1 + z_{rei}} I(\Omega_M) \simeq 9.2, \quad (9.84)$$

where we set $z_{rei} = 10$ and used the result (2.17). Hence, reionization is relevant for all multipoles except for the lowest ones. This is illustrated in Fig. 9.5.

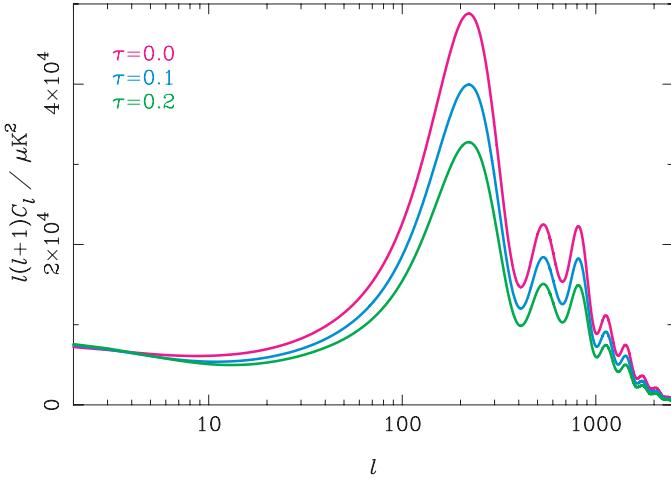


Fig. 9.5 Numerical result for the effect of reionization on the angular spectrum of CMB temperature anisotropy [77]. Lower curves correspond to larger τ .

In fact, the parameters z_{rei} and τ_{rei} are not completely independent. The hydrogen density in the Universe is known, and it limits the maximum possible density of free electrons. It is hard to obtain the precise determination of z_{rei} from CMB measurements, especially in view of the cosmic variance at low multipoles. In practice, one assumes that hydrogen gets completely ionized at $z = z_{rei}$. This uniquely relates the optical depth to z_{rei} , namely,

$$\tau_{rei} = \sigma_T \int_{t_{rei}}^{t_0} dt \cdot 0.75 n_B(t), \quad (9.85)$$

where we recalled that about a quarter of baryons are in helium atoms which are not reionized due to much higher ionization potential. We note that this relation is used in the calculation shown in Fig. 9.5.

The integral in (9.85) is straightforward to calculate, since it is saturated at the lower limit of integration, $z \sim z_{rei}$. At that time, the Hubble parameter is $H = H_0 \sqrt{\Omega_M(1+z)^3}$. Hence

$$\tau_{rei} = \sigma_T \int_0^{z_{rei}} \frac{dz}{1+z} 0.75 \eta_B n_{\gamma,0} (1+z)^3 \frac{1}{H(z)} = 0.5 \frac{\eta_B \sigma_T n_{\gamma,0}}{H_0 \sqrt{\Omega_M}} (1+z_{rei})^{3/2}.$$

We recall that $\sigma_T = 0.67 \cdot 10^{-24} \text{ cm}^2$, and obtain for $h = 0.705$ and $\Omega_M = 0.27$

$$\tau_{rei} = 2.1 \cdot 10^{-3} (1+z_{rei})^{3/2}, \quad (9.86)$$

which gives $\tau_{rei} \simeq 0.08$ for $z_{rei} = 10$. The optical depth extracted from the data is equal to [2] $\tau_{rei} = 0.084 \pm 0.016$, which corresponds to $z_{rei} = 10.8 \pm 1.4$. It is worth noting that the expression (9.85) is actually the upper bound on the parameter τ_{rei} at given z_{rei} ; in other words, measuring τ_{rei} and using then Eq. (9.86) gives a lower

bound on z_{rei} . Still, the estimate based on (9.85) gives a good idea of the redshift of the reionization epoch.

Remarkably, the optical depth turns out to be small enough, so that the CMB temperature anisotropy survives at small angular scales. On the other hand, reionization is not entirely negligible. This is one of numerous coincidences we encounter in cosmology.

9.3 Small Angular Scales

The approximation of tight coupling of photons and baryons and instantaneous photon decoupling is not adequate for studying CMB temperature anisotropy at small angular scales, i.e., at $l \gtrsim 1000$. There are at least four important effects that must be accounted for. Two of them are due to processes occurring at recombination and last scattering epoch.

The first of them, the Silk damping, leads to the suppression of the *acoustic oscillations* in baryon-photon plasma. We have discussed it in Section 8.5.2. The physics behind it is as follows. Towards recombination, the number density of electrons is small, and the mean free path of a photon is large. Photons scatter off electrons without changing their energy. Hence, energetic photons from over-dense regions travel fairly long distances and transfer energy to underdense regions, thus washing out the density fluctuations. The suppression factor for the acoustic oscillations of perturbation of momentum k is estimated as e^{-k^2/k_S^2} , where $k_S \equiv k_S(\eta_r) \sim a_0 \cdot 0.1 \text{ Mpc}^{-1}$, see (8.171). The terms in C_l which are proportional to the oscillating parts in δ_γ and v_γ squared (e.g., the terms with A^2 in (9.73)) are suppressed by e^{-2k^2/k_S^2} , so that the suppression is relevant already at

$$l \gtrsim l_S = \frac{k_S \eta_0}{\sqrt{2}} \sim 1000. \quad (9.87)$$

The non-oscillating part of perturbations is not affected by the Silk damping, but it decreases as k^{-2} at large k , see (9.68), so the entire angular anisotropy rapidly decreases with l at $l \gtrsim 1000$, see Fig. 9.1.

Let us make two comments concerning the Silk damping. First, it follows from (8.186) that this phenomenon is approximately described by introducing into the formulas of Section 9.2.3 the damping factor $e^{-\frac{k^2}{k_S^2}}$ multiplying the oscillating parts of perturbations. The main effect is that the amplitude A in (9.73) and (9.79) is replaced by $A e^{-\frac{l^2}{2k_S^2}}$. Second, lower multipoles are also somewhat suppressed, since perturbations of momenta $k \gtrsim k_S$ give some contributions to multipoles with $l < l_S$.

Problem 9.11. Estimate the corrections due to the Silk effect to formulas (9.73) and (9.79) at $l \ll l_S$ to the leading order in l/l_S . Hint: Use the leading logarithmic approximation wherever necessary.

Problem 9.12. Within the approximation described above, $A(k) \rightarrow e^{-k^2/k_s^2} A(k)$, find the SW and Doppler contributions at $l \gg l_S$, including the dependence of the pre-exponential factors on l/l_S . Hint: Recall that the integrals like (9.75) are saturated at $(u - 1) \sim l^{(r)}/l$.

The second effect is due to the fact that photons do not decouple simultaneously: the sphere of last scattering has finite width $2\Delta z_r \simeq 170$ (see Section 8.5.1). This also leads to the suppression of the angular spectrum at small angular scales. This effect is approximately described as follows. Since photon decoupling is not simultaneous, photons last scatter at somewhat different times η_i . Hence, the functions $\Phi(\eta_r)$, $\delta_\gamma(\eta_r)$, $v_\gamma(\eta_r)$ are replaced by convolutions with the visibility function (8.160). The latter can be approximated by the Gaussian in η_i with the center at $\eta_i = \eta_r$ and width $\Delta\eta_r$. Consider, as an example, the SW effect. The contribution of perturbation of momentum \mathbf{k} to the temperature anisotropy in the direction \mathbf{n} (see Eq. (9.27)) is now given by

$$\Theta_0(\mathbf{kn}, k) = \int d\eta_i N e^{-\frac{(\eta_i - \eta_r)^2}{2\Delta\eta_r^2}} e^{i\mathbf{kn}(\eta_0 - \eta_i)} \{A \cos[kr_s + ku_s(\eta_i - \eta_r)] - B\} \Phi_{(i)},$$

where N normalizes the Gaussian to 1, $r_s = r_s(\eta_r)$, the slowly varying functions A and B are the same as in (9.69), and we can set $A = A(\eta_r)$, $B = B(\eta_r)$. Upon integration, we obtain

$$\begin{aligned} \Theta_0(\mathbf{kn}, k) &= e^{i\mathbf{kn}(\eta_0 - \eta_r)} \left\{ \frac{A}{2} \left[e^{-\frac{\Delta\eta_r^2}{2}(\mathbf{kn} - ku_s)^2} e^{ikr_s} \right. \right. \\ &\quad \left. \left. + e^{-\frac{\Delta\eta_r^2}{2}(\mathbf{kn} + ku_s)^2} e^{-ikr_s} \right] - B e^{-\frac{\Delta\eta_r^2}{2}(\mathbf{kn})^2} \right\} \Phi_{(i)}. \end{aligned}$$

We see that the effect we discuss generally leads to the exponential suppression. However, the suppression depends on the direction of the propagation of the perturbation wave. In particular, the suppression is absent in the last term for waves propagating normally to the line of sight. The reason is clear: perturbations proportional to $B(\eta)$ depend on time weakly, so waves whose momenta are normal to the line of sight look the same along the line of sight at any time in the interval $\Delta\eta_r$. The suppression in the second and third terms is also absent for special directions of wave propagation; this occurs when the phase of a wave is constant along the line of sight because of the cancellation between spatial and temporal phase shifts. Therefore, contributions of some perturbation waves are not exponentially suppressed, and the overall suppression of C_l is power law, and in fact rather weak. On the other hand, the oscillating terms (9.75) are due to waves with $\mathbf{n}_\mathbf{k}\mathbf{n} \sim l^{(r)}/l \ll 1$, so the oscillations in the angular spectrum C_l are exponentially suppressed by the factor

$$C_l^{(osc)} \propto e^{-(ku_s \Delta\eta_r)^2} = e^{-\frac{l^2}{\Delta t^2}}, \quad (9.88)$$

where

$$\Delta l = \frac{\eta_0}{u_s \Delta \eta_r} = \sqrt{3[1 + R_B(\eta_r)]} \frac{\eta_0}{\eta_r} \frac{\eta_r}{\Delta \eta_r} \simeq 2800,$$

and we used (8.166) for numerical estimate. Hence, the effect due to the finite thickness of the sphere of last scattering suppresses particularly strongly the oscillations in the angular spectrum at $l \gtrsim 2800$. This suppression is additional to the Silk effect.

The effect of finite width of the sphere of last scattering on the non-oscillating part of the angular spectrum can be described within the approach based on Eq. (9.58). Consider, e.g., the SW effect. The phenomenon we discuss modifies the formula (9.70) as follows,

$$\tilde{\Theta}_l^{SW}(\mathbf{k}) = \Phi_{(i)} \int d\eta_i N e^{-\frac{(\eta_i - \eta_r)^2}{2\Delta\eta_r^2}} \{A \cos[kr_s + ku_s(\eta_i - \eta_r)] - B\} j_l[k(\eta_0 - \eta_i)].$$

We make use of (9.58), with $\eta = \eta_i$, integrate over η_i and obtain for the term proportional to B ,

$$\tilde{\Theta}_l^{SW,B}(\mathbf{k}) = -\Phi_{(i)} B \cdot \frac{1}{l\sqrt{u}(u^2 - 1)^{1/4}} e^{-\frac{l^2}{2l_*^2}(u^2 - 1)} \cos \left[\left(l + \frac{1}{2} \right) \phi(u) - l \frac{\eta_r}{\eta_0} \sqrt{u^2 - 1} \right],$$

where $u = k\eta_0/l$ and

$$l_* = \frac{\eta_0}{\Delta\eta_r} \sim 1330.$$

The corresponding contribution to C_l is found by averaging fast oscillations according to (9.60). For flat primordial spectrum this gives

$$C_l^{SW,B} = \frac{2\pi A_\Phi}{l^2} \int_1^\infty \frac{du}{u} B^2 e^{-\frac{l^2}{l_*^2}(u^2 - 1)} \frac{1}{u\sqrt{u^2 - 1}}.$$

At $l \gg l_*$, the integral is saturated in the region $u \approx 1$, and we get

$$C_l^{SW,B} = \frac{\pi^{3/2} A_\Phi}{l^2} \frac{l_*}{l} B^2 (k = l/\eta_0).$$

Hence, the effect of finite thickness of the sphere of last scattering is the suppression of high multipoles, linear in the parameter l_*/l . Similar result holds for the non-oscillating part in C_l proportional to A^2 .

Problem 9.13. At $l \gg l_*$, find the non-oscillating part of the SW contribution to C_l proportional to A^2 .

Problem 9.14. Show that the oscillating part of the SW effect in C_l is suppressed at $l \gtrsim l_*$ by the factor given in (9.88).

Finally, there are two other effects that modify the anisotropy spectrum at small angular scales. They are due to structures in the late Universe.

The first effect is weak gravitational lensing. The structures — galaxies, clusters of galaxies — produce gravitational potentials which deflect CMB photons. The deflection angles are small, so this effect is relevant at small angular scales. This is shown in Fig. 9.6. At the same time, weak lensing generates the secondary

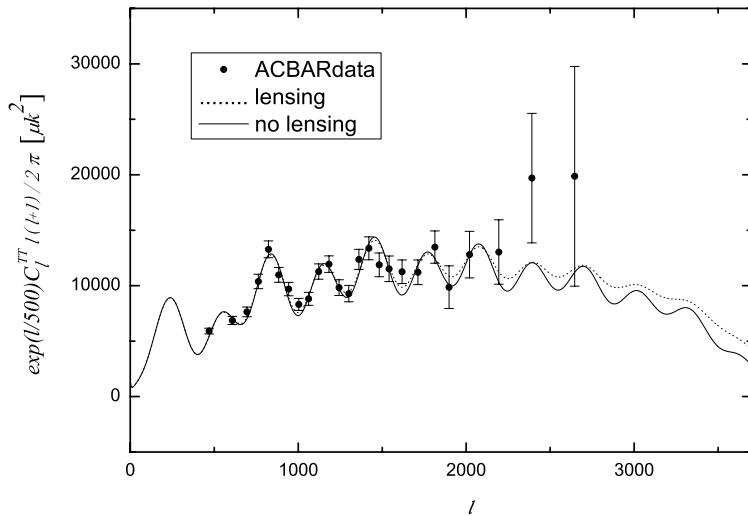


Fig. 9.6 The effect of weak gravitational lensing on CMB temperature anisotropy spectrum [81]. The experimental data are from ACBAR experiment. Note the scale at vertical axis that artificially enhances higher multipoles to compensate for the Silk effect.

anisotropy. The study of weak lensing is of importance for the analysis of the structure formation in the Universe.

The second phenomenon is called the Sunyaev–Zeldovich (SZ) effect [82]. Clusters of galaxies are rich in hot ionized gas. CMB photons passing through a cluster have finite probability to scatter off hot electrons. The photons are heated up, and the secondary CMB anisotropy is generated. Also, the Planck spectrum is distorted in the directions towards clusters: photons leak from the low frequency part to the higher frequency part of the spectrum. The SZ effect is obviously correlated with positions of clusters, and it is also relevant for the structure formation studies.

9.4 Anisotropy Spectrum and Cosmological Parameters

Let us briefly discuss the dependence of the angular spectrum on the cosmological parameters. This issue is very important, since CMB temperature anisotropy is one of the main sources of information about these parameters. Furthermore, it is the most accurate source for some of them.

The cosmological parameters are naturally divided into two categories. The first one characterizes the primordial cosmological perturbations, and the second has to do with the Universe at relatively late cosmological epoch. The former includes the amplitude and tilt of the adiabatic scalar perturbations, $A_{\mathcal{R}}$ (or $A_{\Phi} = (4/9)A_{\mathcal{R}}$) and n_s , the amplitude of tensor perturbations A_T , as well as more subtle characteristics like running scalar tilt $dn_s/d\log k$, tensor tilt n_T , etc. The dependence of

the angular spectrum on these parameters is relatively simple: the amplitudes A_R and A_T determine the overall magnitudes of scalar and tensor contributions to the angular spectrum C_l , while the dependence on the tilts n_s and n_T , as well as on $dn_s/d\log k$ is to a reasonable approximation given by the formula (9.80). Extremely important issues are also the possible non-Gaussianity of primordial perturbations and the admixture of isocurvature modes. The parameters of primordial perturbations are of major interest from the viewpoint of the theory of the very early Universe, as they tell a lot about the mechanism of their generation before the Hot Big Bang epoch.

In the framework of the Λ CDM model, the second set consists of the present value of the Hubble parameter H_0 , relative energy densities Ω_B , Ω_{CDM} , Ω_Λ , Ω_{curv} , Ω_ν , obeying $\Omega_B + \Omega_{CDM} + \Omega_\Lambda + \Omega_{curv} + \Omega_\nu = 1$ (modulo photon contribution to the present energy density; note that Ω_ν is often replaced by the sum of neutrino masses) and the optical depth at reionization τ_{rei} . It is also of great interest to understand whether dark energy is the cosmological constant or it has non-trivial effective equation of state $p_\Lambda = w_\Lambda \rho_\Lambda$, where w_Λ is different from -1 and, generally, depends on time. Note that the redshift z_r of the photon last scattering epoch depends on cosmological parameters very weakly, see Section I.6.2. Barring the possible existence of exotic light particles, the parameter $\Omega_{rad} h^2$ defined in Section 2.1.2 is also well-known. We note in this regard, that the CMB data are used to place bounds on light exotica, which are formulated in terms of bounds on the effective number of relativistic degrees of freedom at and around the recombination epoch.

The total number of cosmological parameters is quite large, and measuring the CMB temperature anisotropy alone is insufficient to determine all of them unambiguously. Indeed, as we discuss in this Section, the properties of the angular spectrum C_l are approximately degenerate in some of the parameters. This phenomenon is similar to the approximate degeneracy in parameters we encountered in Sections I.4.6 and 7.1.2 when discussing SNe Ia and BAO, respectively. The way to remove this degeneracy is to analyse the complete set of the cosmological data obtained by different methods; we use in this and accompanying book the values of parameters derived precisely in this way. Details and references can be found, e.g., in Ref. [3].

Clearly, accurate analysis of the dependence of the angular anisotropy spectrum on cosmological parameters and the determination of these parameters from observations requires numerical work. The relevant codes can be found, e.g., in Ref. [83]. Still, it is of importance to understand this dependence at the qualitative level. This is precisely the purpose of our discussion here.

We begin our discussion with the general normalization of the angular spectrum. As we have seen in Section 9.2.2, the low multipole region is sensitive mostly to the amplitudes A_R and A_T , and is approximately degenerate in these two parameters. At the same time, tensor perturbations give small contribution to the intermediate

multipoles, see Fig. 9.4. If other parameters were known, the general proportion between the two spectral regions would enable one to determine A_R and A_T separately. However, the power in the intermediate part of the spectrum depends on the parameter τ_{rei} , see Fig. 9.5, and, notably, on the scalar tilt n_s : the excess in the low multipole region may be due to either tensor modes or “red” scalar spectrum, $n_s < 1$, or both. The degeneracy with τ_{rei} and n_s is partially removed by the measurements of CMB polarization, temperature anisotropy at small angular scales ($l \gtrsim 1000$) and other data. In particular, the power at small angular scales is smaller for “red” spectrum. However, the spectrum at small angles is subject to its own degeneracies, as it depends on the Silk value l_S which itself depends on the cosmological parameters. Hence, despite of the high precision of CMB temperature anisotropy measurements, the bound on the tensor amplitude is not particularly strong, see (5.46), and the scalar amplitude A_R and tilt n_s are known today within 4% and 1.5% margins, respectively, see (5.37).

The *positions* of peaks in the angular spectrum C_l are very sensitive to the spatial curvature and to lesser extent to the dark energy and dark matter densities. We discuss this point in Section I.6.4. The characteristic angular scale is set by the angle at which the sound horizon at recombination is seen today. This angle depends strongly on the spatial geometry, see the left panel in Fig. 9.7: positive spatial curvature (i.e. $\Omega_{curv} < 0$) increases this angle and hence shifts the peak positions to smaller l . The interplay between Ω_Λ and Ω_M is more subtle: both the sound horizon size at recombination and the distance photons travel since then depend on these parameters. The former behaves roughly as $r_s \propto \Omega_M^{-1/2}$, see (7.33) and (2.18), while the latter has the form $\eta_0 \propto \Omega_M^{-1/2} I(\Omega_\Lambda/\Omega_M)$, where the function I mildly decreases as Ω_Λ/Ω_M grows, see Eq. (2.16). The net effect is that the angular scale mildly increases as Ω_Λ grows (at a given Ω_{curv}), and hence the positions of the peaks shift towards smaller l . This is illustrated in the right panel of Fig. 9.7. We note here, that the spatial curvature and dark energy have similar effects. This property is illustrated in Fig. 9.7. This degeneracy is removed

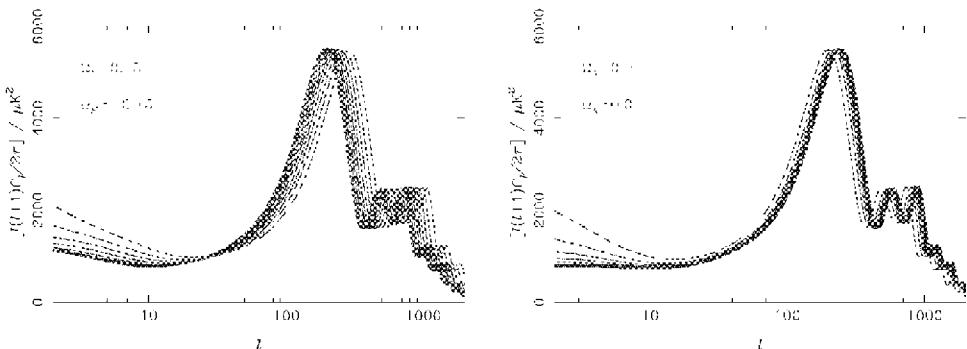


Fig. 9.7 The effect of the spatial curvature (left) and the cosmological constant (right) on the CMB temperature angular spectrum [77]. See Fig. 17.6 for color version.

by using SNe Ia data, see Fig. I.4.6, and/or BAO results, see Fig. 7.6. The dark energy density is known today with good accuracy, and the bound on spatial curvature is quite strong, see Section 2.1.2. In fact, the sound horizon at recombination r_s depends also on the parameter $\Omega_B h^2$, see (7.6). This comes from the dependence of the sound speed on $R_B \propto \rho_B$. Hence, the peak positions, like other properties of the anisotropy spectrum, are complex combinations of the cosmological parameters.

Let us now briefly discuss the peculiarity of the angular spectrum, namely, enhanced odd peaks and suppressed even ones, over and beyond the general decrease at large l . As we discussed in Section 9.2.3, this peculiarity is due to the second, non-oscillating term in (9.65). According to (9.66), this term depends strongly on the baryon and dark matter densities. The main dependence on the baryon density, i.e., on $\Omega_B h^2$, comes from $R_B \propto \rho_B$, so the effect is enhanced at larger $\Omega_B h^2$. This is illustrated in Fig. 9.8. We note that the dependence on the baryon density is particularly strong in the region of the first peak. The function $B(k, \eta_r)$ is large at relatively small momenta, it is proportional to R_B and rapidly varies with momentum, so its effect in the first peak region is pronounced.

The value of $\Omega_M h^2$ determines to large extent the early ISW effect in CMB temperature anisotropy, which is substantial in the first peak region (see Fig. 9.3). As we discussed in Section 9.2.3, this is due to the dependence of the cosmological expansion rate after recombination on $\Omega_M h^2$ (for given Ω_{rad}).

Finally, the high multipoles are sensitive to the Silk scale and optical depth at reionization. As we have already mentioned, the former is, in turn, a rather complicated combination of the cosmological parameters, while the optical depth is considered as an independent parameter.

Thus, the angular spectrum of CMB temperature C_l encodes information on both primordial cosmological perturbations and parameters characterizing the

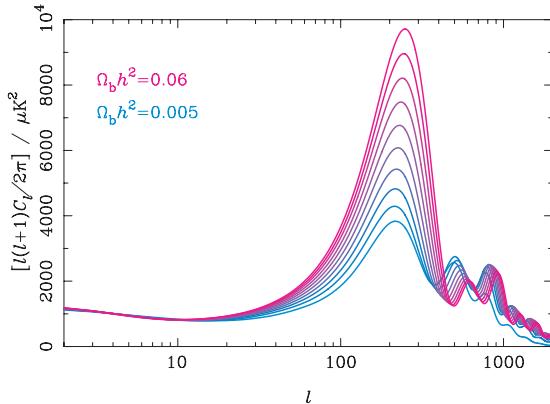


Fig. 9.8 The effect of baryon density on CMB temperature anisotropy [77]. Upper curves correspond to larger baryon fraction.

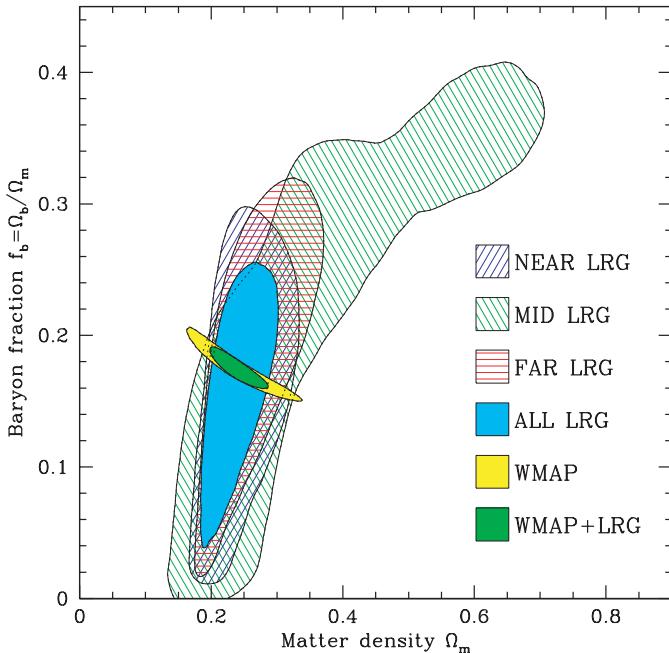


Fig. 9.9 Regions in the space of parameters $(\Omega_M, \Omega_B/\Omega_M)$, allowed at 95% C.L. by data on large scale structure (LRG) and on CMB temperature anisotropy (WMAP), Ref. [53]. See Fig. 17.7 for color version.

content of the Universe and its relatively late evolution. As an example, among other cosmological observations, CMB temperature anisotropy measurements have the highest sensitivity⁶ to Ω_B , see Fig. 9.9. We give other examples in the course of our presentation in this book. The advantage of this source of cosmological information is that the corresponding theoretical analysis is based predominantly on linearized theory and thus it is very robust. On the other hand, the angular spectrum is approximately degenerate in cosmological parameters. The latter fact emphasizes the role of other results of observational cosmology, some of which we discuss in this book.

9.5 Temperature Anisotropy Generated by Isocurvature Modes

Our main emphasis in this book is on adiabatic perturbations, since they dominate in our Universe. However, as we discussed in Section 5.1, the existence of even small admixture of isocurvature modes would be of great importance for cosmology. So, it is worth studying the possible effect of these modes on CMB temperature anisotropy.

⁶In particular, CMB sensitivity noticeably exceeds the sensitivity of BBN to the baryon fraction, see Fig. I.8.3.

As we have seen in Section 6.5, baryon and CDM isocurvature modes are practically indistinguishable. Hence, we consider the CDM mode for definiteness.⁷ The results of Section 6.4 show that it is very different from the adiabatic mode. At small k corresponding to low multipoles, the isocurvature mode, like the adiabatic one, is time-independent, see (6.49) and (6.50). One has for these momenta (see (6.50))

$$\frac{1}{4}\delta_\gamma = \Phi. \quad (9.89)$$

This means that gravitational wells of superhorizon size, where $\Phi < 0$, are *poor* in photons, in contrast to the adiabatic mode. We find from (9.89) that the SW effect for superhorizon isocurvature perturbations gives the following contribution to CMB temperature anisotropy,

$$\Theta_0(\mathbf{n}) = \frac{1}{4}\delta(\eta_r) + \Phi(\eta_r) = 2\Phi(\eta_r),$$

where $\Phi(\eta_r)$ is expressed in terms of the initial value \mathcal{S} via (6.49) and (6.46). On the other hand, unlike for the adiabatic mode, perturbations of shorter wavelengths, that enter the horizon at radiation domination, decrease with momentum, see (6.58). Hence, the admixture of the isocurvature mode would modify the general proportion between the power at small/intermediate l and large l .

Now, we have for the isocurvature mode inside the sound horizon (see (6.59)),

$$\frac{1}{4}\delta_\gamma(\eta_r) + \Phi(\eta_r) = -R_B(\eta_r)\Phi - A(k, \eta) \cdot \delta_{CDM,(i)} \cdot \sin[kr_s + \varphi(k)] \quad (9.90a)$$

$$kv_B = \frac{3}{4k}\delta'_\gamma = -3u_s A(k, \eta) \cdot \delta_{CDM,(i)} \cdot \cos[kr_s + \varphi(k)]. \quad (9.90b)$$

At large k the parameters are (see (6.58))

$$\varphi = 0, \quad A(k, \eta_r) = \frac{I\sqrt{3}}{(1+R_B)^{1/4}} \frac{\sqrt{1+z_{eq}}}{k\eta_0}, \quad \Phi = -9I^2 \frac{1+z_{eq}}{(k\eta_0)^2} \delta_{CDM,(i)}, \quad (9.91)$$

where we have set $\Omega_{CDM}/\Omega_M = 1$ to simplify formulas. It is important that Φ has the opposite sign to $\delta_{CDM,(i)}$. We see that the phase of the oscillations in (9.90) is shifted with respect to the adiabatic mode; at large k this phase shift equals $\pi/2$. In the latter regime, the SW and Doppler effects “are swapped” as compared to the adiabatic mode. Hence, the peaks in the SW effect are at $l \simeq 35 + 295n$ at large l , whereas the peaks in the Doppler contribution occur at $l \simeq 260 + 295n$. Furthermore, the first and second terms in (9.90a) have opposite signs, so the even peaks in C_l are enhanced and odd ones are suppressed, again as in contrast to the adiabatic mode. Finally, both terms in (9.90a) decrease as k grows, so the anisotropy spectrum C_l decays rapidly at large l .

⁷Let us remind the reader that we do not consider neutrino isocurvature modes in this book.

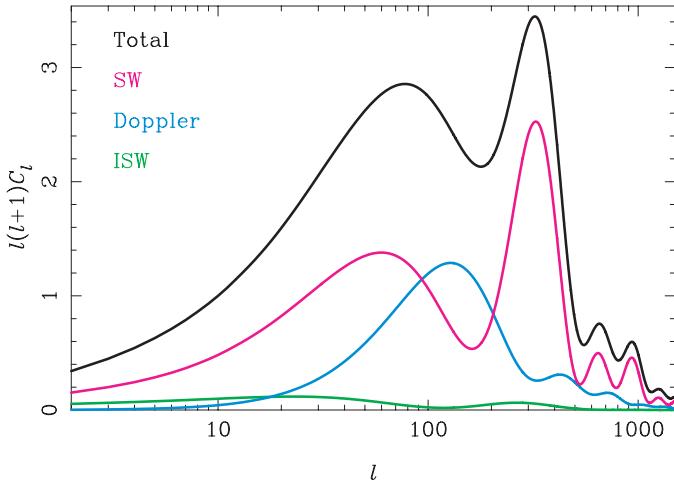


Fig. 9.10 CMB temperature anisotropy spectrum generated by isocurvature mode [77], see Fig. 17.8 for color version. The primordial power spectrum is flat.

The temperature anisotropy spectrum generated by the isocurvature mode is shown in Fig. 9.10. It is *qualitatively* different from the observed spectrum, see Fig. 9.1. On the contrary, the prediction based on adiabatic mode, also shown in Fig. 9.1, is in good agreement with observations. Hence, the isocurvature mode, if any, has much smaller amplitude than the adiabatic one. The observational bound on the admixture of the isocurvature mode is presented in Section 5.4; it is the CMB temperature anisotropy data that are mostly used for obtaining this bound.

This page is intentionally left blank

Chapter 10

*CMB Polarization

Cosmic microwave background is polarized. In this Chapter we first discuss the physics behind CMB polarization and introduce the relevant formalism. We then consider the generation of the polarization by scalar and tensor perturbations and obtain the analytical estimates for the polarization observables within an approximation, somewhat similar to the tight coupling and instantaneous photon decoupling approximations used in Section 9.2. Finally, we present contemporary (at the time this book is written) data on CMB polarization and briefly discuss their relevance for cosmology.

10.1 Sources of CMB Polarization

At the end of recombination, the Compton scattering of photons off free electrons terminates. It is the Compton scattering that is responsible for CMB polarization [84, 18, 85]. The resulting polarization is *linear*. The underlying reason for this effect is that the differential cross section of the Compton scattering off electrons at rest depends on the polarizations of incoming and outgoing photons [31],

$$\frac{d\sigma}{d\Omega} = \frac{3\sigma_T}{8\pi} \cdot |\epsilon'^* \epsilon|^2, \quad (10.1)$$

where σ_T is the Thomson cross section, ϵ' and ϵ are the unit 3-vectors of polarization of incoming and outgoing photons, respectively. The interpretation of this formula is as follows. The directions of the propagation of the incoming and outgoing photons, \mathbf{n}' and \mathbf{n} , determine the scattering plane, see Fig. 10.1. The scattering cross section of a photon with the polarization vector ϵ'_\perp normal to the scattering plane is equal to $3\sigma_T/8\pi$, and the polarization vector of the outgoing photon remains normal to that plane. A photon with polarization vector $\epsilon'_{||}$ lying in the scattering plane interacts with smaller cross section $(3\sigma_T/8\pi) \cdot \cos^2 \theta$, where θ is the scattering angle (the angle between the vectors \mathbf{n}' and \mathbf{n}), while the polarization vector of the outgoing photon still lies in the scattering plane.

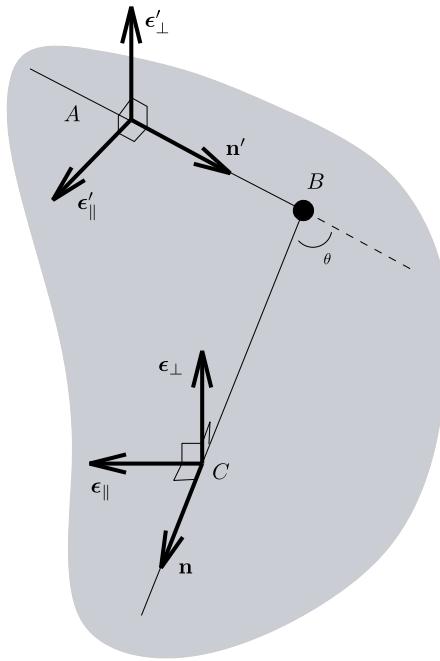


Fig. 10.1 Scattering of a photon (trajectory of motion ABC) off electron (B). Straight lines AB and BC define the scattering plane.

A comment is in order. Recall that the direction of the polarization vector of an electromagnetic plane wave coincides with the direction of the electric field vector \mathbf{E} , while the intensity (number of photons) is proportional to \mathbf{E}^2 . The above discussion then implies that in terms of scattering of electromagnetic waves, the electric field of the outgoing wave \mathbf{E} is proportional to the projection of the electric field \mathbf{E}' of the incoming wave onto the plane normal to vector \mathbf{n} ,

$$\mathbf{E} = \text{const} \cdot [\mathbf{E}' - \mathbf{n} \cdot (\mathbf{n}\mathbf{E}')]. \quad (10.2)$$

This relation is instrumental for what follows.

Consider now unpolarized electromagnetic radiation propagating along the direction \mathbf{n}' and scattered in the direction \mathbf{n} . It is clear from the above discussion that scattered wave is partially polarized in the direction normal to the scattering plane. If the radiation coming from different directions \mathbf{n}' is different, an observer who detects radiation propagating along a given direction \mathbf{n} observes partially polarized radiation. This is precisely what happens in the Universe: before the very last scattering, the photon flux incident on an electron is anisotropic because there are density, velocity and metric perturbations in the Universe at the last scattering epoch. Clearly, the resulting CMB polarization is linear, and the degree of polarization cannot exceed $\delta T/T$.

Let us make a few points concerning this mechanism. First, the incident photon flux before the very last scattering is anisotropic to the extent that the cosmic medium is inhomogeneous at the length scale of the order of the photon mean free path. Hence, the generation of polarization occurs exclusively because the baryon-photon medium is *not tightly coupled*: CMB polarization vanishes in the limit of zero photon mean free path.

Second, the photon coordinate mean free path is of the order of the thickness of the last scattering sphere $\Delta\eta_r$. Perturbations of low conformal momenta vary very little at this length scale. Say, the suppression factor for perturbations relevant for multipoles with $l \sim 50$, i.e., modes with $k \sim \eta_r^{-1}$, is $\Delta\eta_r/\eta_r$. This gives a crude estimate for the degree of polarization, $(\Delta\eta_r/\eta_r) \cdot (\delta T/T) \sim 1 \mu\text{K}/T_0$. This is indeed the right order of magnitude of the observed effect.

Third, the electromagnetic radiation that scatters the last time is polarized by itself due to the previous scattering processes. We ignore this effect in our semi-quantitative estimates, i.e., we assume that the radiation is unpolarized before the very last scattering, and all of the polarization is generated in the very last scattering event. This approximation is justified by the fact that the photon mean free path increases rapidly towards the end of recombination, photons travel shorter distances between the previous scattering events, so the polarization generated in those events is relatively weak. We emphasize, however, that there is no parameterically small factor behind this approximation, unlike in the case of CMB temperature anisotropy: in the latter case, similar approximation of instantaneous photon decoupling is justified at large and intermediate angular scales by the fact that the parameters $\Delta\eta_r \cdot k \sim l \cdot (\Delta\eta_r/\eta_0)$ and k_S/k are small. For CMB polarization, corrections to our approximation are sizeable (dozen per cent and even more) for all angular scales.

Fourth, CMB polarization is generated by both scalar and tensor perturbations. Scalar and tensor contributions to the CMB temperature anisotropy and polarization differ in various respects [86–89, 57]. Most notably, the *B-mode* of polarization (see Section 10.2) is generated at large angular scales by tensor perturbations only [88, 89]. This gives an opportunity to discriminate between the scalar and tensor perturbations, and hence detect the tensor modes.

Fifth, since CMB polarization and temperature anisotropy have common source — scalar and tensor perturbations — the patterns of polarization and temperature must be correlated on the celestial sphere. We discuss this correlation in this Chapter.

Now, like in the case of the temperature anisotropy, CMB polarization is suppressed at small angular scales by the Silk damping and finite thickness of the last scattering sphere. The analysis of these effects is similar to what we have done in Section 9.3, and we do not perform it in this book. Hence, the results in this Chapter on the magnitude of the CMB polarization are valid for low multipoles, $l \lesssim 1000$.

Finally, besides the polarization generated at the recombination epoch, there is the secondary polarization produced at collisions of CMB photons with free electrons at the reionization epoch [85, 90]. The latter effect is proportional to the scattering probability at reionization, which is parameterized by the optical depth τ_{rei} . Since this secondary polarization is generated at $z \sim 10$, the effect exists at large angular scales only. In fact, CMB polarization is even more sensitive to τ_{rei} than the temperature anisotropy, mostly because the former is less degenerate in cosmological parameters. We note in passing that the secondary polarization is also produced due to gravitational lensing by structures.¹

10.2 Polarization Tensor. *E*- and *B*-modes

Let us introduce the parameters characterizing the polarization of electromagnetic radiation propagating in a given direction \mathbf{n} . Let \mathbf{s} be a unit vector normal to the vector \mathbf{n} . An observer with a polarimeter oriented along \mathbf{s} measures the intensity $I(\mathbf{s})$ (temperature $T(\mathbf{s})$ in the CMB case). It is proportional to the projection squared of the electric field on \mathbf{s} , averaged over the period of a wave, $\langle |\mathbf{E} \cdot \mathbf{s}|^2 \rangle = s_a \langle E_a E_b^* \rangle s_b$, where E_a with $a = 1, 2$ are components of the electric field in the plane normal to \mathbf{n} , and s_a are components of \mathbf{s} . Hence, the polarization properties are in general encoded in the Hermitean tensor

$$I_{ab} = \langle E_a E_b^* \rangle,$$

so that $I(\mathbf{s}) \propto s_a I_{ab} s_b$. The total intensity $I = \langle E_a E_a^* \rangle = \langle |E_1|^2 \rangle + \langle |E_2|^2 \rangle$ is an invariant that does not have direct relevance to polarization. Let us introduce the dimensionless *polarization tensor*

$$P_{ab} = \frac{I_{ab}}{I}.$$

This is a Hermitean tensor with unit trace. It is completely determined by three real parameters. For unpolarized radiation, one has $P_{ab} = \delta_{ab}/2$ and $\det P_{ab} = 1/4$, while completely linearly polarized radiation has $P_{ab} = E_a E_b^*$, where \mathbf{E} is a well defined 2-dimensional vector, so that $\det P_{ab} = 0$. In general, $0 \leq \det P_{ab} \leq 1/4$. Hence, one introduces the scalar *degree of polarization* by writing

$$\mathcal{P} = \sqrt{1 - 4 \cdot \det P}.$$

It varies from zero (unpolarized radiation) to 1 (completely polarized radiation).

For circularly polarized wave, one has $E_2 = \pm i E_1$, $P_{11} = P_{22} = 1/2$, and $P_{12} = -P_{21} = \pm i$ is pure imaginary. On the contrary, electric field can be chosen

¹There are also mechanisms that distort CMB polarization. An example is the Faraday rotation occurring when electromagnetic radiation passes through ionized gas in magnetic field. We do not discuss this and other mechanisms that lead to the distortion of CMB polarization; their effects are small at angular scales we consider.

real for linearly polarized radiation, and the tensor P_{ab} is real and symmetric. It is the latter case that is of interest for our purposes. Thus, linearly polarized radiation is described by the real symmetric tensor P_{ab} with unit trace. It is determined by two independent parameters.

Let $\tilde{s}_a^{(1)}$ and $s_a^{(2)}$ be the orthonormalized pair of the eigenvectors of the matrix P_{ab} . Since P_{ab} has unit trace, it is written as

$$P_{ab} = \lambda_p s_a^{(1)} s_b^{(1)} + (1 - \lambda_p) s_a^{(2)} s_b^{(2)},$$

where $0 \leq \lambda_p \leq 1$. The degree of polarization is $\mathcal{P} = |1 - 2\lambda_p|$. We see that partially polarized wave can be viewed as *an incoherent sum* of two waves of intensities $I \propto \lambda_p$ and $I \propto (1 - \lambda_p)$, which are linearly polarized in the mutually orthogonal directions $\mathbf{s}^{(1)}$ and $\mathbf{s}^{(2)}$.

One can decompose partially polarized radiation in an alternative way, by extracting completely unpolarized (n) and completely polarized (p) parts. As an example, for $\lambda_p < 1/2$ one writes

$$I_{ab} = [\lambda_p \delta_{ab} + (1 - 2\lambda_p) s_a^{(2)} s_b^{(2)}] \cdot I \equiv \frac{1}{2} \delta_{ab} \cdot I^{(n)} + E_a^{(p)} E_b^{(p)},$$

where $\mathbf{E}^{(p)}$ is a well-defined vector. The intensities of unpolarized and polarized parts are equal to $(1 - \mathcal{P})I$ and $\mathcal{P}I$, while the direction of polarization coincides with the direction of vector $\mathbf{E}^{(p)} \propto \mathbf{s}^{(2)}$. The latter observation enables one to represent the observational data in terms of short dashes on a sphere directed along the CMB polarization vector $\mathbf{E}^{(p)}$. One version of such a representation is shown in Figs. 10.2, 17.9. We note that maps of this sort generally have singularities. Their classification and properties can be found, e.g., in the book [60].

In optics, polarization is traditionally described in terms of the Stokes parameters U, Q, V . They are related to the matrix P_{ab} as follows,

$$P_{ab} = \frac{1}{2} \begin{pmatrix} 1+Q & U-iV \\ U+iV & 1-Q \end{pmatrix},$$

with $\mathcal{P} = \sqrt{U^2 + V^2 + Q^2}$. If any of the Stokes parameters is non-zero, the radiation is (partially) polarized. Circular polarization corresponds to $Q = U = 0, V \neq 0$. On the contrary, $V = 0$ for linear polarization. The latter properties are independent of the coordinate frame: the quantities V^2 and $Q^2 + U^2$ are separately invariant under rotations of the plane normal to the direction of radiation propagation.

Problem 10.1. Prove the last statement above. Hint: Make use of the decomposition of the polarization tensor into symmetric and antisymmetric parts.

For linear polarization one has

$$Q = \frac{\langle E_1^2 \rangle - \langle E_2^2 \rangle}{I}, \quad U = \frac{2\langle E_1 E_2 \rangle}{I}, \quad V = 0. \quad (10.3)$$

The parameter Q describes the linear polarization along one of the coordinate axes, while U gives the polarization along the line rotated by the angle $\pi/4$ from the

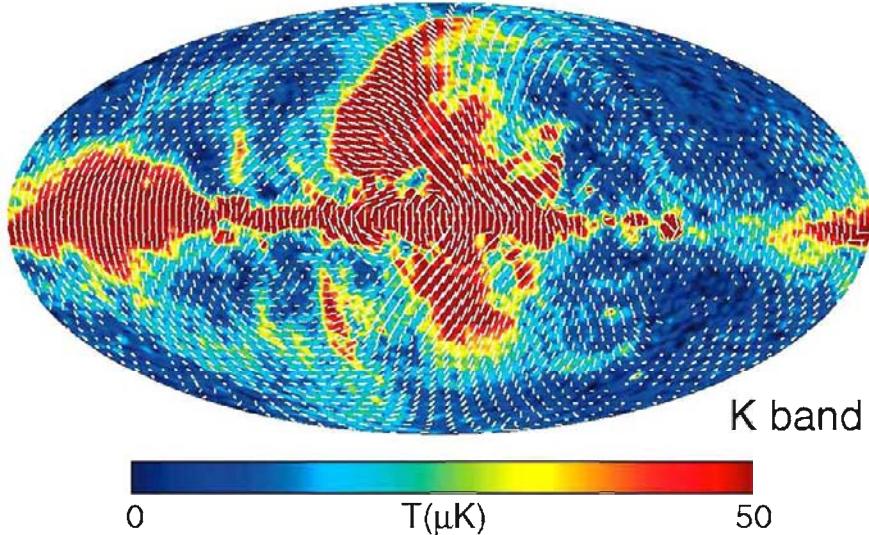


Fig. 10.2 CMB polarization map obtained by the WMAP experiment [68], see Fig 17.9 for color version. Color encodes the degree of polarization in K-band wavelength (23 GHz), $\mathcal{P} \cdot T_0$, white short dashes show the direction of polarization, their length is logarithmically dependent on the magnitude of \mathcal{P} .

coordinate axes. Indeed, in the plane normal to the propagation direction, one introduces the coordinate frames (x, y) and (x', y') rotated by angle 45° with respect to each other, and denotes the intensity of radiation measured with a polarimeter oriented along the x -axis by I_x , and similarly for other axes. Then

$$Q = \frac{I_x - I_y}{I}, \quad U = \frac{I_{x'} - I_{y'}}{I}.$$

Clearly, the parameters Q and U are not invariant under rotations of the coordinate frame. The invariant parameter is the degree of polarization $\mathcal{P} = \sqrt{Q^2 + U^2}$. The ratio $U/Q \equiv \tan 2\alpha$ determines the angle α between the polarization vector and x -axis.

It is often convenient to use, instead of the tensor P_{ab} , the traceless tensor

$$\mathcal{P}_{ab} = P_{ab} - \frac{1}{2}g_{ab} = \frac{1}{\langle E_a E^a \rangle} \left(\langle E_a E_b \rangle - \frac{1}{2} \langle E_c E^c \rangle g_{ab} \right). \quad (10.4)$$

We have accounted here for the fact that we are interested in the polarization tensor on the celestial sphere and introduced the metric g_{ab} on this sphere. The tensor \mathcal{P}_{ab} vanishes for unpolarized radiation. The degree of polarization is given in its terms by $\mathcal{P} = \sqrt{-4 \det \mathcal{P}_a^b}$.

The Stokes parameters U and Q are inconvenient for our purposes, since they are coordinate-dependent. The only coordinate-independent parameter introduced so far is \mathcal{P} ; no other local invariants can be constructed from the elements of the symmetric tensor \mathcal{P}_{ab} . However, the tensor \mathcal{P}_{ab} defined on the 2-sphere can be represented in terms of scalar and pseudoscalar potentials \mathcal{P}_E and \mathcal{P}_B ,

$$\mathcal{P}_{ab} = -\{\nabla_a \nabla_b\} \mathcal{P}_E - \{\epsilon_a^c \nabla_b \nabla_c\} \mathcal{P}_B, \quad (10.5)$$

where ∇_a is the covariant derivative on the sphere, ϵ_{ab} is the antisymmetric tensor (the latter involves the factor \sqrt{g} , see Section I.A.1), braces denote symmetric traceless part, i.e.,

$$\{\nabla_a \nabla_b\} = \frac{1}{2} (\nabla_a \nabla_b + \nabla_b \nabla_a - g_{ab} \Delta), \quad \{\epsilon_a^c \nabla_b \nabla_c\} = \frac{1}{2} (\epsilon_a^c \nabla_b \nabla_c + \epsilon_b^c \nabla_a \nabla_c).$$

Here $\Delta = g_{ab} \nabla^a \nabla^b$ is the Laplacian on the sphere. The inverse transformation is

$$-\Delta(\Delta + 2) \mathcal{P}_E = 2\{\nabla^a \nabla^b\} \mathcal{P}_{ab}, \quad (10.6a)$$

$$-\Delta(\Delta + 2) \mathcal{P}_B = 2\{\epsilon_a^c \nabla^c \nabla^b\} \mathcal{P}_{ab}. \quad (10.6b)$$

The signs in the definition of the potentials (10.5) are chosen according to the tradition; this choice corresponds to the sign convention in (10.7) and (10.8).

Problem 10.2. Derive the inverse transformation formulas (10.6).

The formulas (10.5) and (10.6) relate the two functions on the sphere, U and Q , parameterizing the tensor \mathcal{P}_{ab} in a certain coordinate frame, to the two functions \mathcal{P}_E and \mathcal{P}_B , which are invariant under coordinate transformations. Accordingly, one introduces the notions of *E-mode* ($\mathcal{P}_E \neq 0, \mathcal{P}_B = 0$) and *B-mode* ($\mathcal{P}_E = 0, \mathcal{P}_B \neq 0$). This separation into modes is important in the cosmological context [92, 93], since scalar perturbations generate only *E-mode* (modulo lensing effects), while tensor perturbations generate both *E-* and *B-modes*, see Sections 10.3.2 and 10.3.3. The detection of the *B-mode*, and hence tensor perturbations, is the major goal of CMB polarization studies.

Let us illustrate the possibility to represent the symmetric traceless tensor in the form (10.5). Let the manifold on which this tensor is defined be plane rather than sphere. We perform the Fourier transformation on this plane, $\mathcal{P}_{ab} = \mathcal{P}_{ab}(q_c)$. For given momentum q_c one can define two linear independent vectors q_c and $\epsilon_{cd} q_d$ and decompose any tensor in their bilinear combinations,

$$\mathcal{P}_{ab} = q_a q_b \mathcal{P}_1 + \epsilon_{ac} q_b q_c \mathcal{P}_2 + \epsilon_{bc} q_a q_c \mathcal{P}_3 + \epsilon_{ac} \epsilon_{bd} q_c q_d \mathcal{P}_4.$$

Since \mathcal{P}_{ab} is symmetric, we have

$$\mathcal{P}_2 = \mathcal{P}_3 \equiv \mathcal{P}_B/2, \quad (10.7)$$

while zero trace implies

$$\mathcal{P}_1 = -\mathcal{P}_4 \equiv \mathcal{P}_E/2. \quad (10.8)$$

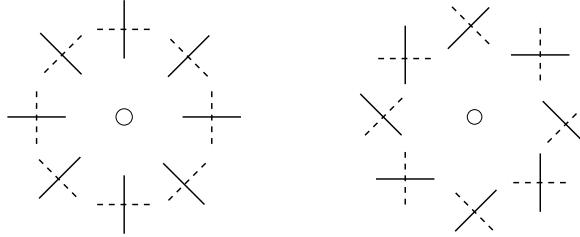


Fig. 10.3 E - and B -modes (left and right panel, respectively) in the natural representation of polarization. Negative and positive E - and B -modes are shown by solid and dashed lines, respectively.

Thus, the tensor \mathcal{P}_{ab} has precisely the form (10.5) with the obvious replacement of ∇_a by $i q_a$. Note that

$$q_a \mathcal{P}_{ab} = -\frac{q^2}{2} (q_b \mathcal{P}_E + \epsilon_{bc} q_c \mathcal{P}_B).$$

Hence, the potentials \mathcal{P}_E and \mathcal{P}_B correspond to the decomposition of the divergence of the tensor into gradient and curl parts, respectively. This is the origin of the nomenclature: the gradient and curl modes are called E - and B -modes, in analogy to electric and magnetic fields.

The E - and B -modes are illustrated in Fig. 10.3. Namely, consider a region near a maximum or minimum of the potential \mathcal{P}_E (left panel) or \mathcal{P}_B (right panel), and assume that the potential is symmetric under rotations around the point of extremum (the position of extremum is shown by small circle). The properties of polarization are represented in a way similar to that used in Figs. 10.2, 17.9: lines show the direction of polarization. The direction of lines along (perpendicular to) the radial direction corresponds to negative (positive) E -mode; this is shown in the left panel. The rotation of lines clockwise (counterclockwise) in the right panel corresponds to decreasing (increasing) potential P_B . The reflection with respect to a straight line crossing the origin does not change the left picture (parity-even E -mode) and reverses the picture in the right panel (parity-odd B -mode).

It is convenient to expand the potentials of E - and B -modes in spherical harmonics on the celestial sphere,

$$\begin{aligned} \mathcal{P}_E(\mathbf{n}) &= \sqrt{2} \sum_{lm} \sqrt{\frac{(l-2)!}{(l+2)!}} a_{lm}^E Y_{lm}(\mathbf{n}), \\ \mathcal{P}_B(\mathbf{n}) &= \sqrt{2} \sum_{lm} \sqrt{\frac{(l-2)!}{(l+2)!}} a_{lm}^B Y_{lm}(\mathbf{n}). \end{aligned} \tag{10.9}$$

The choice of the normalization here will become clear later on. The polarization tensor is determined by the second derivatives of the potentials, so the monopole and dipole harmonics in (10.9) are irrelevant. We use the relations (10.6) and recall that the spherical harmonics Y_{lm} are eigenfunctions of the spherical Laplacian with the eigenvalues $[-l(l+1)]$, so that $\Delta(\Delta+2)Y_{lm} = [(l+2)!(l-2)!]Y_{lm}$.

Then we find

$$\begin{aligned} a_{lm}^E &= -\sqrt{\frac{2(l-2)!}{(l+2)!}} \int d\mathbf{n} Y_{lm}^*(\mathbf{n}) \{\nabla^a \nabla^b\} \mathcal{P}_{ab}(\mathbf{n}) \\ &= -\int d\mathbf{n} Y_{lm}^{(E)ab*}(\mathbf{n}) \mathcal{P}_{ab}(\mathbf{n}), \end{aligned} \quad (10.10)$$

$$\begin{aligned} a_{lm}^B &= -\sqrt{\frac{2(l-2)!}{(l+2)!}} \int d\mathbf{n} Y_{lm}^*(\mathbf{n}) \{\epsilon_b^c \nabla^a \nabla^b\} \mathcal{P}_{ac}(\mathbf{n}) \\ &= -\int d\mathbf{n} Y_{lm}^{(B)ab*}(\mathbf{n}) \mathcal{P}_{ab}(\mathbf{n}), \end{aligned} \quad (10.11)$$

where we integrated by parts and introduced the notation

$$\begin{aligned} Y_{lm,ab}^{(E)} &= \sqrt{\frac{2(l-2)!}{(l+2)!}} \left(\nabla_a \nabla_b Y_{lm} - \frac{1}{2} g_{ab} \nabla^c \nabla_c Y_{lm} \right), \\ Y_{lm,ab}^{(B)} &= \sqrt{\frac{(l-2)!}{2(l+2)!}} (\nabla_a \nabla_c Y_{lm} \epsilon_b^c + \nabla_c \nabla_b Y_{lm} \epsilon_a^c). \end{aligned}$$

The formulas (10.10), (10.11) enable one to calculate the coefficients $a_{lm}^{E,B}$ for known polarization tensor. The coefficients transform as follows under the spatial reflection,

$$a_{lm}^E \rightarrow (-1)^l a_{lm}^E, \quad a_{lm}^B \rightarrow -(-1)^l a_{lm}^B.$$

Problem 10.3. Show that the functions $Y_{lm,ab}^{(X)}$, $X = E, B$ constitute a complete orthonormalized system for second-rank symmetric traceless tensors on a sphere, so that the expansion of \mathcal{P}_{ab} is

$$\mathcal{P}_{ab}(\mathbf{n}) = -\sum_{l=2}^{\infty} \sum_{m=-l}^l \left(a_{lm}^E Y_{lm,ab}^{(E)}(\mathbf{n}) + a_{lm}^B Y_{lm,ab}^{(B)}(\mathbf{n}) \right).$$

This result gives an alternative definition of the coefficients $a_{lm}^{E,B}$, which is equivalent to (10.9). Hint: Making use of the properties of the spherical harmonics Y_{lm} listed in Section F.2, show that

$$\int d\mathbf{n} Y_{lm,ab}^{(X)*}(\mathbf{n}) Y_{l'm'}^{(X')ab}(\mathbf{n}) = \delta_{XX'} \delta_{ll'} \delta_{mm'},$$

and

$$\sum_{lm} Y_{lm,ab}^{(X)*}(\theta, \phi) Y_{lm}^{(X')ab}(\theta', \phi') = \delta_{XX'} \cdot \delta(\phi - \phi') \delta(\cos \theta - \cos \theta').$$

When discussing CMB polarization, it is traditional to use the spherical coordinate frame. The 2-dimensional metric in this frame is standard,

$$ds^2 = g_{ab} dx^a dx^b = d\theta^2 + \sin^2 \theta d\phi^2. \quad (10.12)$$

This is orthogonal, but not normalized coordinate frame, so the linear polarization tensor (10.4) is expressed in terms of the Stokes parameters as follows,

$$\mathcal{P}_{ab}(\mathbf{n}) = \frac{1}{2} \begin{pmatrix} Q(\mathbf{n}) & -U(\mathbf{n}) \sin \theta \\ -U(\mathbf{n}) \sin \theta & -Q(\mathbf{n}) \sin^2 \theta \end{pmatrix}. \quad (10.13)$$

The explicit form of the tensor spherical harmonics $Y_{lm}^{(E)ab}$ and $Y_{lm}^{(B)ab}$ is conveniently written in terms of the auxiliary functions

$$\begin{aligned} W_{lm}(\mathbf{n}) &\equiv \left(\frac{\partial^2}{\partial \theta^2} - \frac{\cos \theta}{\sin \theta} \frac{\partial}{\partial \theta} + \frac{m^2}{\sin^2 \theta} \right) Y_{lm}(\mathbf{n}) \\ &= \left(2 \frac{\partial^2}{\partial \theta^2} + l(l+1) \right) Y_{lm}(\mathbf{n}), \end{aligned} \quad (10.14)$$

$$X_{lm}(\mathbf{n}) \equiv \frac{2im}{\sin \theta} \left(\frac{\partial}{\partial \theta} - \frac{\cos \theta}{\sin \theta} \right) Y_{lm}(\mathbf{n}). \quad (10.15)$$

One has

$$Y_{lm,ab}^{(E)}(\mathbf{n}) = \sqrt{\frac{(l-2)!}{2(l+2)!}} \begin{pmatrix} W_{lm}(\mathbf{n}) & X_{lm}(\mathbf{n}) \sin \theta \\ X_{lm}(\mathbf{n}) \sin \theta & -W_{lm}(\mathbf{n}) \sin^2 \theta \end{pmatrix}, \quad (10.16)$$

$$Y_{lm,ab}^{(B)}(\mathbf{n}) = \sqrt{\frac{(l-2)!}{2(l+2)!}} \begin{pmatrix} -X_{lm}(\mathbf{n}) & W_{lm}(\mathbf{n}) \sin \theta \\ W_{lm}(\mathbf{n}) \sin \theta & X_{lm}(\mathbf{n}) \sin^2 \theta \end{pmatrix}. \quad (10.17)$$

Problem 10.4. Making use of the definition of the spherical harmonics $Y_{lm}(\mathbf{n})$ (see Appendix F), derive the second relation in (10.14).

To describe the CMB polarization, one can use, instead of the tensor harmonics $Y_{lm,ab}^{(E)}$, $Y_{lm,ab}^{(B)}$, another set of harmonics

$$\pm_2 Y_{lm,ab} = Y_{lm,ab}^{(E)} \mp i Y_{lm,ab}^{(B)} = \pm_2 Y_{lm} \cdot \begin{pmatrix} 1 & \mp i \sin \theta \\ \mp i \sin \theta & -\sin^2 \theta \end{pmatrix}, \quad (10.18)$$

where

$$\pm_2 Y_{lm}(\mathbf{n}) = \sqrt{\frac{(l-2)!}{2(l+2)!}} (W_{lm}(\mathbf{n}) \pm i X_{lm}(\mathbf{n})). \quad (10.19)$$

The tensors (10.18) have the following transformation property under rotation of the tangent plane to the sphere by angle α ,

$${}_s Y_{lm,ab} \rightarrow e^{is\alpha} \cdot {}_s Y_{lm,ab}, \quad s = \pm 2.$$

Hence, ${}_s Y_{lm,ab}$ are objects of helicity s , cf. Section 2.2.3. In other words, one can construct objects of helicity ± 2 from the tensor \mathcal{P}_{ab} by projecting it onto helicity ± 1 vectors $\mathbf{e}^{(\pm)} = \mathbf{e}^{(1)} \pm i \mathbf{e}^{(2)}$ in the tangent plane,

$$\pm_2 \mathcal{P} \equiv \mathcal{P}_{ab} e^{(\pm)a} e^{(\pm)b} = Q \pm i U.$$

The functions ${}_s Y_{lm}(\mathbf{n})$ are called spherical harmonics of helicity (or spin) ± 2 . They obey normalization and completeness conditions

$$\int d\mathbf{n} {}_s Y_{l'm'}^*(\mathbf{n}) \cdot {}_{s'} Y_{lm}(\mathbf{n}) = \delta_{ll'} \delta_{mm'} \delta_{ss'}, \quad (10.20)$$

$$\sum_{lm} {}_s Y_{lm}^*(\theta', \phi') \cdot {}_{s'} Y_{lm}(\theta, \phi) = \delta(\phi - \phi') \cdot \delta(\cos \theta - \cos \theta') \cdot \delta_{ss'}. \quad (10.21)$$

Their property under complex conjugation is

$${}_s Y_{l,m}^* = (-1)^{m+2} \cdot {}_{-s} Y_{l,-m},$$

while under spatial reflection

$${}_s Y_{lm} \rightarrow (-1)^l \cdot {}_{-s} Y_{lm}.$$

Problem 10.5. Derive the formulas (10.18), (10.20) and (10.21).

The CMB temperature anisotropy is described by the angular spectrum $C_l \equiv C_l^{TT}$. Likewise, one introduces the spectra of the two polarization modes and all possible correlators $\langle a_{lm}^E a_{l'm'}^{E*} \rangle$, $\langle a_{lm}^B a_{l'm'}^{B*} \rangle$, $\langle a_{lm}^T a_{l'm'}^{T*} \rangle$, etc. The coefficients a_{lm}^T correspond to the temperature fluctuations; we considered them in Chapter 9 without using the superscript T . The Gaussian fluctuations obey

$$\langle a_{lm}^X a_{l'm'}^{Y*} \rangle = C_l^{XY} \delta_{ll'} \delta_{mm'},$$

where $X, Y = T, E, B$, and the brackets denote the average over an ensemble of Universes. This gives

$$C_l^{XY} = \frac{1}{2l+1} \sum_m \langle a_{lm}^X a_{lm}^{Y*} \rangle. \quad (10.22)$$

Generally, there are 6 correlators, but parity under spatial reflection requires that the correlators C_l^{TB} and C_l^{EB} vanish. The latter property is consistent with the observations [91]. Hence, CMB data can in principle be used to determine the four spectra C_l^{TT} , C_l^{TE} , C_l^{EE} and C_l^{BB} .

Problem 10.6. Let us add to the Lagrangian of free electromagnetic field the term that violates Lorentz-invariance, P and CP , namely, $\Delta \mathcal{L} = b_\mu A_\nu \epsilon^{\mu\nu\lambda\rho} F_{\lambda\rho}$, where b_μ is a constant space-like vector. Show that photon polarization rotates in vacuo in this theory. Show that this leads to mixing between the Stokes parameters. Find corrections to the spectra C_l^{XY} resulting from this mixing, to the leading non-trivial order in b_μ .

Since CMB polarization and temperature anisotropy have one and the same origin, the behavior of the polarization and cross correlation spectra C_l^{EE} , C_l^{BB} , C_l^{TE} is in a well-defined correspondence with the behavior of the temperature spectrum C_l^{TT} . This is illustrated in Fig. 10.4. We consider the physics behind the features in these spectra in the following Sections. Here we only highlight a few properties. The polarization spectrum clearly has oscillatory behavior. For adiabatic

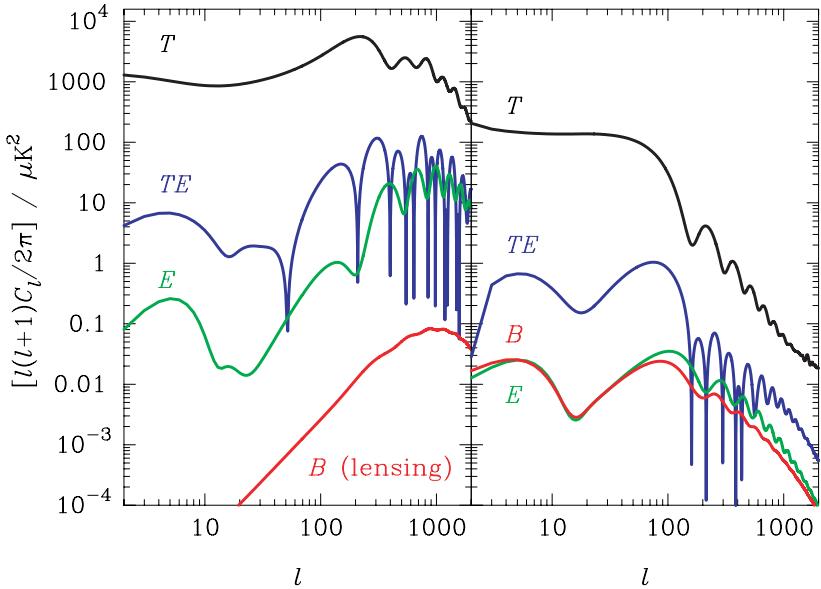


Fig. 10.4 An example of CMB temperature and polarization angular spectra [94], see Fig. 17.10 for color version. Left and right panels show contributions of adiabatic scalar perturbations and tensor perturbations, respectively. The values of cosmological parameters are similar to those used in this book, except for unrealistically large tensor-to-scalar ratio, $r = 0.38$, and optical depth at reionization, $\tau_{rei} = 0.17$.

scalar modes, maxima in C_l^{TT} are at similar multipole numbers l as minima in C_l^{EE} , and vice versa. As we show later on, this is due to the fact that polarization is mostly due to the *velocity perturbations* in the baryon-photon medium, while temperature anisotropy comes mainly from the perturbations in the energy density and gravitational potential. The cross correlation spectrum C_l^{TE} involves both of these contributions, while Fig. 10.4 shows the absolute value of C_l^{TE} , so the ‘‘jumps’’ in the latter are twice more often. The generation of small admixture of B -mode in the case of scalar perturbations is due to gravitational lensing of E -mode photons in the recent Universe. Note also the enhancement of the spectrum at $l \lesssim 10$ in both panels, which is due to the effect of reionization.

10.3 Generation of CMB Polarization

The purpose of this Section is to derive approximate formulas for CMB polarization which show its dependence on the primordial scalar and tensor spectra and on other cosmological parameters. We briefly discuss the relevance of polarization measurements for cosmology in Section 10.4. For polarization produced at the recombination epoch, we use the approximation in which its generation occurs at the very last

scattering event, while the radiation is considered unpolarized before that, see the discussion in Section 10.1.

10.3.1 Generalities

Consider unpolarized radiation coming to the point of the very last scattering along the direction \mathbf{n}' and scattered in the direction \mathbf{n} towards an observer. The electric field of the outgoing radiation obeys Eq. (10.2). Let \mathbf{e}_a be the set of two basis vectors on the celestial sphere (or, equivalently, on the sphere of last scattering) normalized by $\mathbf{e}_a \mathbf{e}_b = g_{ab}$. Then Eq. (10.2) has the form $E_a = C \mathbf{e}_a \mathbf{E}'$, where C is a constant independent of the vector \mathbf{n}' , E_a are components of the electric field of outgoing radiation and \mathbf{E}' is the electric field of incoming radiation. Since incoming radiation is unpolarized, we have to average over the directions of the vector \mathbf{E}' keeping in mind that it is orthogonal to \mathbf{n}' . In the Cartesian basis, the average values of bilinear combinations determining the polarization tensor are

$$\langle E'_i E'_j \rangle = \frac{1}{2} (\delta_{ij} - n'_i n'_j) \langle \mathbf{E}'^2 \rangle, \quad (10.23)$$

Then

$$\langle E_a E_b \rangle = \frac{1}{2} (g_{ab} - \mathbf{n}' \mathbf{e}_a \cdot \mathbf{n}' \mathbf{e}_b) \cdot C^2 \cdot \langle \mathbf{E}'^2 \rangle, \quad (10.24)$$

while the intensity of the outgoing wave is

$$\langle E_a E_b g^{ab} \rangle = \frac{1}{2} (1 + \mathbf{n} \mathbf{n}'^2) \cdot C^2 \cdot \langle \mathbf{E}'^2 \rangle. \quad (10.25)$$

Problem 10.7. Obtain the result (10.25) directly, by using the expression (10.1) for the Compton cross section.

Let us now make use of the formula (10.4) and integrate it over arrival directions \mathbf{n}' for the interesting case when the intensity of incoming radiation $I(\mathbf{n}')$ depends on arrival direction. We also recall that last scattering can occur at different moments of time η ; the fraction of photons detected at the present time η_0 and last scattered in the time interval $(\eta, \eta + d\eta)$ equals $V(\eta_0, \eta) d\eta$, where $V(\eta_0, \eta)$ is the visibility function (8.160). We obtain the following expression for the polarization tensor,

$$\mathcal{P}_{ab}(\mathbf{n}) = \frac{\int_0^{\eta_0} d\eta V(\eta_0, \eta) \int d\mathbf{n}' I(\eta, \mathbf{n}') [\frac{1}{2}(1 - (\mathbf{n} \mathbf{n}')^2) g_{ab} - \mathbf{n}' \mathbf{e}_a \cdot \mathbf{n}' \mathbf{e}_b]}{\int_0^{\eta_0} d\eta V(\eta_0, \eta) \int d\mathbf{n}' I(\eta, \mathbf{n}') [1 + (\mathbf{n} \mathbf{n}')^2]}. \quad (10.26)$$

It is clear from this formula that the polarization tensor \mathcal{P}_{ab} vanishes in the isotropic case, when the intensity $I(\mathbf{n}')$ of incoming radiation is independent of the arrival direction \mathbf{n}' . Hence, polarization is generated only if there is local anisotropy in the intensity (temperature). Furthermore, since the vector \mathbf{n}' enters the right hand side of (10.26) in a bilinear combination, this local anisotropy must have *quadrupole component*.

The anisotropy in the intensity $I(\eta, \mathbf{n}')$ is due to scalar and tensor perturbations. Hence, the numerator of (10.26) is linear in these perturbations, while the denominator is of the zeroth order. We write the intensity in the form $I = C_1 T^4$ and obtain to the zeroth order

$$\int d\mathbf{n}' I(\eta, \mathbf{n}') (1 + (\mathbf{n}\mathbf{n}')^2) = 4C_1 T^4(\eta) \cdot \frac{4\pi}{3}. \quad (10.27)$$

The numerator in (10.26) contains the anisotropic part of the intensity, which is given by

$$\delta I(\eta, \mathbf{n}') = C_1 T^4(\eta) \cdot 4 \frac{\delta T}{T}(\eta, \mathbf{x}; \mathbf{n}'), \quad (10.28)$$

where $(\delta T/T)(\eta, \mathbf{x}; \mathbf{n}')$ is the temperature perturbation of photons arriving along the direction \mathbf{n}' to the position \mathbf{x} where they last scatter at time η . The previous (last before the very last) scattering of these photons occurs at different moments of time η' , and the fraction of photons scattered in the time interval $(\eta', \eta' + d\eta')$ is $V(\eta, \eta')d\eta'$, where $V(\eta, \eta')$ is the visibility function at time η . The latter is again given by (8.160). In complete analogy to (9.22) we write for scalar perturbations

$$\begin{aligned} \frac{\delta T}{T}(\eta, \mathbf{x}; \mathbf{n}') = & \int_0^\eta d\eta' V(\eta, \eta') \int d^3 k \left[\Phi(\eta') + \frac{1}{4} \delta_\gamma(\eta') \right. \\ & \left. + i\mathbf{k}\mathbf{n}' v_B(\eta') \right] e^{i[\mathbf{k}\mathbf{x} - \mathbf{k}\mathbf{n}'(\eta - \eta')]} . \end{aligned} \quad (10.29)$$

The sign of the second term in the exponent is due to the fact that photon propagates along the direction \mathbf{n}' before the very last scattering, so it interacts the previous time at the point $[\mathbf{x} - \mathbf{n}'(\eta - \eta')]$. As we will see later, the Doppler term $\mathbf{k}\mathbf{n}' v_B$ gives the dominant contribution to CMB polarization; the Sachs–Wolfe contribution is suppressed by the parameter $\Delta\eta_r/\eta_r \simeq 0.04$. The integrated Sachs–Wolfe contribution is also small, and we have omitted it in (10.29). In complete analogy to (9.24), we have for tensor perturbations

$$\frac{\delta T}{T}(\eta, \mathbf{x}; \mathbf{n}') = \frac{1}{2} \int_0^\eta d\eta' V(\eta, \eta') \int d^3 k \int_{\eta'}^\eta d\eta'' n'_i h'_{ij}(\eta'') n'_j e^{i[\mathbf{k}\mathbf{x} - \mathbf{k}\mathbf{n}'(\eta - \eta'')]}. \quad (10.30)$$

We set the lower limit of integration over conformal time in (10.29), (10.30) equal to zero. This is a formality, since all integrals with the visibility function are saturated in a short time interval around the time of last scattering.

Using the formulas (10.27), (10.28), we obtain from (10.26) that the CMB polarization tensor is

$$\begin{aligned} \mathcal{P}_{ab}(\eta_0; \mathbf{n}) = & \frac{3}{4\pi} \int d\mathbf{n}' \left[\frac{1}{2}(1 - (\mathbf{n}\mathbf{n}')^2) g_{ab} - \mathbf{n}' \mathbf{e}_a \cdot \mathbf{n}' \mathbf{e}_b \right] \\ & \times \int_0^{\eta_0} d\eta V(\eta_0, \eta) \frac{\delta T}{T}(\eta, \mathbf{x}; \mathbf{n}'), \end{aligned} \quad (10.31)$$

where $\delta T/T(\eta, \mathbf{x}; \mathbf{n}')$ is given by the integrals (10.29) and (10.30) for scalar and tensor perturbations, respectively. Let us make one point here. Until now we denoted by \mathbf{n} the direction of photons propagation after the very last scattering. It is convenient to make the replacement $\mathbf{n} \rightarrow -\mathbf{n}$ in what follows, so that \mathbf{n} is the direction of observation. Then the photon last scatters at point $\mathbf{x} = (\eta_0 - \eta)\mathbf{n}$.

The result (10.31) involves multiple integrals over conformal time. Let us consider them in some details, and expand in the parameters $\Delta\eta_r/\eta_r$, $k\Delta\eta_r$ to the leading non-trivial order (as we have already noticed, the condition $k\Delta\eta_r \ll 1$ means that we do not study multipoles with $l \gtrsim \eta_0/\Delta\eta_r \sim 1000$).

It follows from (10.29) that the contribution of scalar perturbations to (10.31) involves the integral

$$\int_0^{\eta_0} d\eta V(\eta_0, \eta) e^{i\mathbf{k}\mathbf{n}(\eta_0 - \eta)} \int_0^\eta d\eta' V(\eta, \eta') e^{-i\mathbf{k}\mathbf{n}'(\eta - \eta')} f(\eta'), \quad (10.32)$$

where $f(\eta')$ is a slowly varying function. It is saturated in the region

$$\eta', \eta \sim \eta_r, \quad (\eta - \eta_r) \sim (\eta' - \eta_r) \sim \Delta\eta_r. \quad (10.33)$$

Hence, to the leading order in $\Delta\eta_r/\eta_r$, the integral (10.32) has the form

$$f(\eta_r) \cdot e^{i\mathbf{k}\mathbf{n}(\eta_0 - \eta_r)} \cdot \tilde{I},$$

where

$$\tilde{I} = \int_0^{\eta_0} d\eta V(\eta_0, \eta) e^{i\mathbf{k}\mathbf{n}(\eta_r - \eta)} \int_0^\eta d\eta' V(\eta, \eta') e^{-i\mathbf{k}\mathbf{n}'(\eta - \eta')}. \quad (10.34)$$

Making use of the definitions (8.160), (8.159), we change the order of integration in (10.34),

$$\begin{aligned} \tilde{I} &= \int_0^{\eta_0} d\eta' \sigma_{T\text{an}_e}(\eta') \exp \left(- \int_{\eta'}^{\eta_0} \sigma_{T\text{an}_e}(\eta'') d\eta'' \right) e^{i(\mathbf{k}\mathbf{n}'\eta' + \mathbf{k}\mathbf{n}\eta_r)} \\ &\times \int_{\eta'}^{\eta_0} d\eta \sigma_{T\text{an}_e}(\eta) e^{-i(\mathbf{k}\mathbf{n}' + \mathbf{k}\mathbf{n})\eta}. \end{aligned} \quad (10.35)$$

In the time range (10.33), the number density of free electrons exponentially decreases in time, see (8.161), hence $n_e(\eta) = F(T) \cdot e^{-\frac{\Delta H}{4T}} = F(\eta) \cdot e^{-\frac{\eta^2}{2\eta_r\Delta\eta_r}}$, where the function $F(\eta)$ varies in time slowly. To the leading order in $\Delta\eta_r/\eta_r$, the number density of electrons at $(\eta - \eta_r) \ll \eta_r$ can be approximated by

$$n_e(\eta) = n_e(\eta_r) e^{-\frac{\eta - \eta_r}{\Delta\eta_r}}. \quad (10.36)$$

The scale factor varies in time slowly, so we can set in the integrand in (10.35)

$$a(\eta) = a(\eta') = a(\eta_r). \quad (10.37)$$

The inner integral over η in (10.35) is saturated at the lower limit, so to the leading order in $\Delta\eta_r/\eta_r$ we have

$$e^{i\mathbf{kn}'\eta'} \int_{\eta'}^{\eta_0} d\eta \sigma_T a n_e(\eta) e^{-i(\mathbf{kn}' + \mathbf{kn})\eta} = \frac{\sigma_T a n_e(\eta') \cdot \Delta\eta_r}{1 + i(\mathbf{kn} + \mathbf{kn}')\Delta\eta_r} \cdot e^{-i\mathbf{kn}\eta'}.$$

The remaining integral over η' has the form

$$\int_0^{\eta_0} d\eta' e^{-i\mathbf{kn}(\eta' - \eta_r)} \sigma_T^2 a^2 n_e^2(\eta') \exp\left(-\int_{\eta'}^{\eta_0} \sigma_T a n_e(\eta'') d\eta''\right).$$

Making use of (10.36) and (10.37), we evaluate it by performing the change of integration variable to τ (optical depth) which obeys $d\tau = \sigma_T a n_e d\eta$. As a result, we obtain to the zeroth order in $\mathbf{kn}\Delta\eta_r$:

$$\tilde{I} = \frac{1}{1 + i\mathbf{kn}'\Delta\eta_r}. \quad (10.38)$$

Hence, the leading order expression for the integral entering the formula for the scalar contribution to the polarization tensor is

$$\int_0^{\eta_0} d\eta V(\eta_0, \eta) e^{i\mathbf{kn}(\eta_0 - \eta)} \int_0^\eta d\eta' V(\eta, \eta') e^{-i\mathbf{kn}'(\eta - \eta')} f(\eta') = \frac{e^{i\mathbf{kn}(\eta_0 - \eta_r)}}{1 + i\mathbf{kn}'\Delta\eta_r} \cdot f(\eta_r). \quad (10.39)$$

Note that we cannot neglect $\mathbf{kn}'\Delta\eta_r$ in the denominator: it is this term that is responsible for local quadrupole anisotropy and hence polarization, see Section 10.3.2.

Let us now turn to the reionization epoch, $z \sim 10$, when the additional generation of CMB polarization occurs. Since the optical depth at this epoch is small, $\tau_{rei} \simeq 0.08$, the probability that photon scatters twice is negligible. Since the last scattering before reionization occurs at the recombination epoch, the integral (10.32) involves $\eta = \eta_{rei}$ and $\eta' \sim \eta_r$, so that

$$\begin{aligned} & \int_0^{\eta_0} d\eta V(\eta_0, \eta) e^{i\mathbf{kn}(\eta_0 - \eta)} \int_0^\eta d\eta' V(\eta_0, \eta') e^{i\mathbf{kn}'(\eta' - \eta)} f(\eta') \\ &= \tau_{rei} \cdot e^{i\mathbf{kn}(\eta_0 - \eta_{rei})} \cdot e^{-i\mathbf{kn}'(\eta_{rei} - \eta_r)} f(\eta_r). \end{aligned} \quad (10.40)$$

Also, CMB polarization generated at recombination is distorted at the reionization epoch, so the contribution (10.39) has to be multiplied by $e^{-\tau_{rei}}$ for small enough wavelengths. This factor is close to 1, however, and we omit it in what follows.

The contribution of tensor perturbations generated at recombination, Eq. (10.30), yields the following integral in (10.31)

$$\int_0^{\eta_0} d\eta V(\eta_0, \eta) e^{i\mathbf{kn}(\eta_0 - \eta)} \int_0^\eta d\eta' V(\eta, \eta') \int_{\eta'}^\eta d\eta'' f(\eta'') e^{-i\mathbf{kn}'(\eta - \eta'')} \quad (10.41)$$

$$= f(\eta_r) e^{i\mathbf{kn}(\eta_0 - \eta_r)} \int_0^{\eta_0} d\eta V(\eta_0, \eta) e^{i\mathbf{kn}(\eta_r - \eta)} \int_0^\eta d\eta' V(\eta, \eta') \int_{\eta'}^\eta d\eta'' e^{-i\mathbf{kn}'(\eta - \eta'')}, \quad (10.42)$$

where $f(\eta) \propto n'_i h'_{ij}(\eta) n'_j$. It is sufficient to work to the leading (zeroth) order in $\mathbf{k}\mathbf{n}'\Delta\eta_r$, since the quadrupole component of the local anisotropy is present already at this order (the expression (10.30) contains the bilinear combination $n'_i n'_j$ explicitly). The integration over η'' gives

$$\int_{\eta'}^{\eta} d\eta'' e^{-i\mathbf{k}\mathbf{n}'(\eta-\eta'')} = \frac{1}{i\mathbf{k}\mathbf{n}'}(1 - e^{-i\mathbf{k}\mathbf{n}'(\eta-\eta')}).$$

Now, for any η , the visibility function obeys

$$\int_0^{\eta} d\eta' V(\eta, \eta') = 1.$$

The integrals over η and η' in (10.42) are saturated in the region (10.33), so to the zeroth order in $\mathbf{k}\mathbf{n}\Delta\eta_r$ we have

$$\int_0^{\eta_0} d\eta V(\eta_0, \eta) \cdot e^{i\mathbf{k}\mathbf{n}(\eta_r - \eta)} \int_0^{\eta} \frac{d\eta' V(\eta, \eta')}{i\mathbf{k}\mathbf{n}'} = \frac{1}{i\mathbf{k}\mathbf{n}'}.$$

We have already encountered similar integral with the weight $e^{-i\mathbf{k}\mathbf{n}'(\eta-\eta')}$, see Eq. (10.34), and its value is given by (10.38). Hence, we obtain finally

$$\begin{aligned} & \int_0^{\eta_0} d\eta V(\eta_0, \eta) e^{i\mathbf{k}\mathbf{n}(\eta_0 - \eta)} \int_0^{\eta} d\eta' V(\eta, \eta') \int_{\eta'}^{\eta} \partial\eta'' f(\eta'') e^{-i\mathbf{k}\mathbf{n}'(\eta-\eta'')} \\ &= \frac{1}{i\mathbf{k}\mathbf{n}'} \left(1 - \frac{1}{1 + i\mathbf{k}\mathbf{n}'\Delta\eta_r} \right) \cdot f(\eta_r) e^{i\mathbf{k}\mathbf{n}(\eta_0 - \eta_r)} \\ &= \Delta\eta_r \cdot f(\eta_r) e^{i\mathbf{k}\mathbf{n}(\eta_0 - \eta_r)} \end{aligned} \quad (10.43)$$

Problem 10.8. Since Eq. (10.42) contains $\eta - \eta'' \sim \Delta\eta_r$, the last exponential factor in (10.42) can be set equal to 1, so that $\int_{\eta'}^{\eta} d\eta'' e^{i\mathbf{k}\mathbf{n}(\eta-\eta'')} = \eta - \eta'$. Using this fact, rederive the result (10.43).

The tensor contribution generated at reionization has the form similar to (10.41), where the integration over η'' is performed in a fairly large time interval $\eta_{rec} \lesssim \eta \lesssim \eta_{rei}$. Hence, one cannot replace the function $f(\eta'')$ by a constant, and the product $\mathbf{k}\mathbf{n}'\eta''$ cannot be treated as a small parameter. Instead of (10.43) we have now

$$\begin{aligned} & \int_0^{\eta_0} d\eta V(\eta_0, \eta) e^{i\mathbf{k}\mathbf{n}(\eta_0 - \eta)} \int_0^{\eta} d\eta' V(\eta, \eta') \int_{\eta'}^{\eta} d\eta'' f(\eta'') e^{-i\mathbf{k}\mathbf{n}'(\eta-\eta'')} \\ &= \tau_{rei} \cdot e^{i\mathbf{k}\mathbf{n}(\eta_0 - \eta_{rei})} \int_{\eta_r}^{\eta_{rei}} d\eta f(\eta) e^{-i\mathbf{k}\mathbf{n}'(\eta_{rei} - \eta)}. \end{aligned} \quad (10.44)$$

It is convenient to perform further analysis for scalar and tensor perturbations separately.

10.3.2 Scalar perturbations

Let us begin with the scalar contribution to the polarization tensor (10.31) generated at recombination. Using (10.39), we write

$$\mathcal{P}_{ab}(\mathbf{n}) = \int d^3k \mathcal{P}_{ab}(\mathbf{k}, \mathbf{n}), \quad (10.45)$$

where

$$\begin{aligned} \mathcal{P}_{ab}(\mathbf{k}, \mathbf{n}) &= \frac{3}{4\pi} e^{i\mathbf{k}\mathbf{n}(\eta_0 - \eta_r)} \int \frac{d\mathbf{n}'}{1 + i\mathbf{k}\mathbf{n}'\Delta\eta_r} \left[\frac{1}{2}(1 - (\mathbf{n}\mathbf{n}')^2)g_{ab} - \mathbf{n}'\mathbf{e}_a \cdot \mathbf{n}'\mathbf{e}_b \right] \\ &\times \left[\Phi(\mathbf{k}, \eta_r) + \frac{1}{4}\delta_\gamma(\mathbf{k}, \eta_r) + i\mathbf{k}\mathbf{n}'v_B(\mathbf{k}, \eta_r) \right]. \end{aligned} \quad (10.46)$$

The latter quantity has the meaning of the contribution of a perturbation of momentum \mathbf{k} to the polarization tensor. We often omit the argument in the notation for the Fourier components of Φ , δ_γ and v_B .

To proceed further, we employ the trick used in Section 9.2.2 for the analysis of the tensor contribution to CMB temperature anisotropy. Namely, we note that the potentials $\mathcal{P}_E(\mathbf{n})$ and $\mathcal{P}_B(\mathbf{n})$, as well as temperature anisotropy $\delta T(\mathbf{n})$ are scalars under spatial rotations. The multipoles $C_l^{X,Y}$, $X, Y = T, E, B$ are also scalars. Furthermore, the contributions of modes with different momenta add up in C_l^{XY} incoherently. So, when calculating the contribution of a mode of a given momentum \mathbf{k} we can use any coordinate frame, which may be different for different \mathbf{k} .

Let us choose the spherical coordinate frame with the azimuthal axis along the vector \mathbf{k} . In this frame, the coordinates of the vectors \mathbf{n} and \mathbf{n}' are (θ, ϕ) and (θ', ϕ') , respectively. In the 3-dimensional Cartesian coordinates, we have

$$\begin{aligned} \mathbf{n} &= (\sin\theta \cos\phi, \sin\theta \sin\phi, \cos\theta), & \mathbf{e}_\theta &= (-\cos\theta \cos\phi, -\cos\theta \sin\phi, \sin\theta), \\ \mathbf{n}' &= (\sin\theta' \cos\phi', \sin\theta' \sin\phi', \cos\theta'), & \mathbf{e}_\phi &= (\sin\theta \sin\phi, -\sin\theta \cos\phi, 0). \end{aligned}$$

Hence, the bilinear expressions with scalar products entering (10.46) are

$$\begin{aligned} 2(\mathbf{n}\mathbf{n}')^2 &= \sin^2\theta \sin^2\theta' + 2\cos^2\theta \cos^2\theta' + \sin^2\theta \sin^2\theta' \cos 2(\phi - \phi') \\ &+ \sin 2\theta \sin 2\theta' \cos(\phi - \phi'), \end{aligned} \quad (10.47a)$$

$$\begin{aligned} 2(\mathbf{e}_\theta \mathbf{n}')^2 &= \cos^2\theta \sin^2\theta' + 2\sin^2\theta \cos^2\theta' + \sin^2\theta' \cos^2\theta \cos 2(\phi - \phi') \\ &- \sin 2\theta \sin 2\theta' \cos(\phi - \phi'), \end{aligned} \quad (10.47b)$$

$$2(\mathbf{e}_\phi \mathbf{n}')^2 = \sin^2\theta \sin^2\theta' - \sin^2\theta \sin^2\theta' \cos 2(\phi - \phi'), \quad (10.47c)$$

$$2\mathbf{e}_\theta \mathbf{n}' \cdot \mathbf{e}_\phi \mathbf{n}' = \frac{1}{2} \sin 2\theta \sin^2 \theta' \sin 2(\phi - \phi') - \sin^2 \theta \sin 2\theta' \sin(\phi - \phi'). \quad (10.47d)$$

Making use of these expressions in (10.46) and performing the integration over the azimuthal angle ϕ' we find that the terms which depend on ϕ' do not contribute to (10.46). As a result, in our coordinate frame off-diagonal components of the polarization tensor $\mathcal{P}_{\theta\phi}(\mathbf{k}, \mathbf{n})$ and $\mathcal{P}_{\phi\theta}(\mathbf{k}, \mathbf{n})$ vanish, which yields $U(\mathbf{k}, \mathbf{n}) = 0$. The components $\mathcal{P}_{\theta\theta}(\mathbf{k}, \mathbf{n})$ and $\mathcal{P}_{\phi\phi}(\mathbf{k}, \mathbf{n})$ are independent of ϕ . This means, in particular, that the scalar perturbations do not generate B -mode of polarization. Indeed, the tensor $\mathcal{P}_{ab}(\mathbf{k}, \mathbf{n})$ is diagonal and depends only on θ , so the right hand side of Eq. (10.6b) vanishes.

The expression for the component $\mathcal{P}_{\theta\theta}(\mathbf{k}, \mathbf{n})$ involves the combination

$$\begin{aligned} \frac{1}{2} \left(1 - (\mathbf{n}\mathbf{n}')^2\right) g_{\theta\theta} - \mathbf{n}'\mathbf{e}_\theta \cdot \mathbf{n}'\mathbf{e}_\theta &= \frac{\sin^2 \theta}{4} (1 - 3 \cos^2 \theta') \\ &= -\frac{1}{2} \sin^2 \theta \cdot P_2(\cos \theta'), \end{aligned} \quad (10.48)$$

where P_2 is the Legendre polynomial of the second order, see Appendix F. Similarly, $\mathcal{P}_{\phi\phi}(\mathbf{k}, \mathbf{n})$ involves

$$\frac{1}{2} \left(1 - (\mathbf{n}\mathbf{n}')^2\right) g_{\phi\phi} - \mathbf{n}'\mathbf{e}_\phi \cdot \mathbf{n}'\mathbf{e}_\phi = \frac{1}{2} \sin^4 \theta \cdot P_2(\cos \theta'). \quad (10.49)$$

Hence, there is the following relation between the two non-vanishing components,

$$\mathcal{P}_{\phi\phi}(\mathbf{k}, \mathbf{n}) = -\sin^2 \theta \cdot \mathcal{P}_{\theta\theta}(\mathbf{k}, \mathbf{n}).$$

This is in accordance with (10.13). To integrate over θ' , we perform the expansion

$$\frac{1}{1 + i\mathbf{k}\mathbf{n}'\Delta\eta_r} = \sum_{l=0}^{\infty} (-ik\Delta\eta_r \cos \theta')^l. \quad (10.50)$$

Making use of the properties of the integrals with the Legendre polynomials (see Appendix F), we find to the leading non-trivial order in $k\Delta\eta_r$:

$$\mathcal{P}_{\theta\theta}(\mathbf{k}, \mathbf{n}) = -\frac{\mathcal{P}_{\phi\phi}(\mathbf{k}, \mathbf{n})}{\sin^2 \theta} = \frac{k\Delta\eta_r}{5} \sin^2 \theta \cdot \left[\left(\Phi + \frac{1}{4}\delta_\gamma \right) \cdot k\Delta\eta_r - kv_B \right] \cdot e^{i\mathbf{k}\mathbf{n}(\eta_0 - \eta_r)}. \quad (10.51)$$

Like in the temperature anisotropy case, the contribution to the l -th multipole comes mainly from modes of momenta $k\eta_0 \sim l$. Therefore, the polarization at relatively large angular scales we consider in this Section, $l \lesssim \eta_0/\Delta\eta_r \sim 1000$, is mostly due to the Doppler effect.

The reason for the domination of the Doppler effect over the SW effect is as follows. The Doppler effect in the temperature fluctuation (the last term in (10.46)) is locally anisotropic by itself, since it is proportional to $\mathbf{k}\mathbf{n}'$. This local anisotropy

is of dipole structure. It becomes quadrupole due to extra factor $\mathbf{k}\mathbf{n}'\Delta\eta_r$ reflecting the anisotropy in the photon flux from the distance of the order of the photon mean free path and formally originating from the expansion of the denominator in (10.46). Hence the factor $k\Delta\eta_r$ in (10.51). Temperature fluctuations caused by the SW effect do not possess local anisotropy, so the quadrupole local anisotropy appears to the second order in $\mathbf{k}\mathbf{n}'\Delta\eta_r$.

For a perturbation of momentum \mathbf{k} , we define the quantity $a_{lm}^E(\mathbf{k})$ by the formula completely analogous to (10.10), but with $\mathcal{P}_{ab}(\mathbf{k}, \mathbf{n})$ instead of $\mathcal{P}_{ab}(\mathbf{n})$. Then the multipole C_l^{EE} is expressed as follows, cf. (9.50),

$$C_l^{EE} = \frac{1}{2l+1} \int d^3k \sum_m \langle a_{lm}^E(\mathbf{k}) a_{lm}^{E*}(\mathbf{k}) \rangle, \quad (10.52)$$

where we have in mind that one integration over momenta is already performed by making use of the property

$$\langle kv_B(\mathbf{k}, \eta_r)kv_B^*(\mathbf{k}', \eta_r) \rangle = \frac{1}{(2\pi)^3} P_v(k, \eta_r) \delta(\mathbf{k} - \mathbf{k}'),$$

where $P_v(k, \eta_r)$ is the velocity power spectrum at recombination, which can be expressed in terms of the primordial scalar spectrum using the formulas of Chapter 6. As we already pointed out, the integrand in (10.52) can be calculated in any coordinate frame. In the frame introduced above, we obtain by using (10.16) and (10.14) and keeping only the Doppler term in (10.51),

$$\begin{aligned} a_{lm}^E(\mathbf{k}) &= - \int d\mathbf{n} Y_{lm}^{(E)ab*} \mathcal{P}_{ab}(\mathbf{n}) = - \int d\mathbf{n} \left[Y_{lm,\theta\theta}^{(E)*} P_{\theta\theta}(\mathbf{n}) + Y_{lm,\phi\phi}^{(E)*} \frac{P_{\phi\phi}(\mathbf{n})}{\sin^4\theta} \right] \\ &= \frac{2}{5} k^2 v_B \Delta\eta_r \sqrt{\frac{(l-2)!}{2(l+2)!}} \times \int d\mathbf{n} e^{ik(\eta_0 - \eta_r) \cos\theta} \sin^2\theta \\ &\quad \times \left[2 \frac{\partial^2}{\partial\theta^2} + (l+1)l \right] Y_{lm}(\mathbf{n}). \end{aligned} \quad (10.53)$$

The integral here vanishes for $m \neq 0$ since $Y_{lm} \propto e^{im\phi}$. The spherical harmonics with $m = 0$ are (see (F.39)),

$$Y_{l,0} = \sqrt{\frac{2l+1}{4\pi}} P_l(\cos\theta).$$

It follows from (F.23) that the Legendre polynomials obey

$$\frac{d^2}{d\theta^2} P_l(\cos\theta) = \cos\theta \frac{dP_l(\cos\theta)}{d\cos\theta} - l(l+1) P_l(\cos\theta).$$

Therefore,

$$\sin^2\theta \cdot \left[2 \frac{\partial^2}{\partial\theta^2} + (l+1)l \right] P_l(\cos\theta) = (1-x^2) \left[2x \frac{dP_l(x)}{dx} - l(l+1) P_l(x) \right], \quad (10.54)$$

where $x = \cos \theta$. We now use the recurrence relations (F.25), (F.26) and write the right hand side of (10.54) in the following form,

$$\frac{(l+2)!}{(l-2)!} \left(\frac{P_{l-2}(x)}{(2l-1)(2l+1)} - \frac{2P_l(x)}{(2l-1)(2l+3)} + \frac{P_{l+2}(x)}{(2l+1)(2l+3)} \right).$$

We insert this expression into (10.53) together with the expansion (F.35) of the oscillating exponent in the Legendre polynomials, then integrate over θ using orthogonality relation between the Legendre polynomials (F.39) and obtain

$$a_{l,0}^E(\mathbf{k}) = -\frac{4\pi}{5} k v_B k \Delta \eta_r \sqrt{\frac{(l+2)!}{2(l-2)!}} \sqrt{\frac{2l+1}{4\pi}} \cdot i^l \left(\frac{j_{l-2}[k(\eta_0 - \eta_r)]}{(2l-1)(2l+1)} \right. \\ \left. + \frac{2j_l[k(\eta_0 - \eta_r)]}{(2l-1)(2l+3)} + \frac{j_{l+2}[k(\eta_0 - \eta_r)]}{(2l+1)(2l+3)} \right). \quad (10.55)$$

We have already encountered the expression in parenthesis in (10.55), see (9.53); it is given by (9.54). We obtain finally

$$a_{l,m}^E(\mathbf{k}) = -\delta_{m,0} \cdot \frac{4\pi}{5} k^2 v_B(\mathbf{k}) \Delta \eta_r \sqrt{\frac{(l+2)!}{2(l-2)!}} \sqrt{\frac{2l+1}{4\pi}} \cdot i^l \cdot \frac{j_l[(\eta_0 - \eta_r)k]}{(\eta_0 - \eta_r)^2 k^2}. \quad (10.56)$$

This result enables us to calculate the spectra C_l^{EE} and C_l^{TE} .

The E -mode polarization spectrum proper, C_l^{EE} , is obtained from (10.52) and (10.56),

$$C_l^{EE} = \frac{2\pi}{25} \int \frac{d^3 k}{(2\pi)^3} P_v(k, \eta_r) \cdot \frac{(l+2)!}{(l-2)!} (\Delta \eta_r)^2 \left(\frac{j_l[(\eta_0 - \eta_r)k]}{(\eta_0 - \eta_r)^2 k} \right)^2. \quad (10.57)$$

As usual, the main contribution to the integral (10.57) comes from momenta $k \sim l/\eta_0$. The Doppler contribution to the polarization is small at $l \lesssim l^{(r)} \sim 100$, since the relevant modes of perturbations are superhorizon and have small velocity $k v_B$. The spectrum C_l^{EE} oscillates at $l \sim 100 - 1000$, and positions of its maxima and minima are almost the same as for the Doppler contribution to the temperature anisotropy, see problem 10.10. The overall magnitude of C_l^{EE} is suppressed by the factor $(\Delta \eta_r)^2$ whose origin has already been discussed. The dependence of C_l^{EE} on l is weak in this range of multipoles (modulo oscillations), so the quantity

$$\mathcal{D}_l^{EE} = \frac{l(l+1)}{2\pi} C_l^{EE} T_0^2$$

grows like l^2 as l increases. The latter property follows from simple power counting: for nearly flat primordial spectrum of scalar perturbations, we have $d^3 k P_v \propto dk/k$, and at $k \sim l/\eta_0$ one has $j_l \sim 1/l$, so that the integral in (10.57) is independent of l , modulo oscillations. All these properties are in accordance with the numerical analysis shown in Fig. 10.4.

Problem 10.9. Estimate the behavior of the multipoles C_l^{EE} at $l \simeq 30 - 100$.

Problem 10.10. Show that for the adiabatic perturbations, the spectrum C_l^{EE} is estimated at $l \sim 100 - 1000$ as follows,

$$C_l^{EE} = \frac{3\pi}{50} \left(\frac{\Delta\eta_r}{\eta_0} \right)^2 \frac{A_\Phi A^2(l)}{1 + R_B(\eta_r)} \left[\frac{2}{3} c_4 - \frac{1}{2} \sqrt{\frac{\pi l^{(r)}}{l}} \cos \left(2 \frac{l}{l^{(r)}} + \frac{\pi}{4} \right) \right], \quad (10.58)$$

where $A(l)$ is the same function which enters (9.73) and (9.79), c_4 is of order 1 and the explicit expressions at large l are given by (9.67) and $c_4 = 1$.

It follows from (10.58) that the spectrum of the E -mode has maxima near

$$l_n^{EE} = \left[\frac{3\pi}{8} + \pi(n-1) \right] l^{(r)} \simeq 115 + 305(n-1), \quad n = 1, 2, \dots,$$

which is consistent with the numerical results shown in Fig. 10.4. One observes from Fig. 10.4, that the oscillations in C_l^{EE} are damped at large l . This is due to the fact that the amplitude of the oscillatory term in (10.58) decreases with l as $1/\sqrt{l}$. We insert in (10.58) the numerical values

$$\frac{\Delta\eta_r}{\eta_r} \simeq 0.04, \quad \frac{\eta_r}{\eta_0} \simeq 0.02, \quad A_\Phi \simeq 1.09 \cdot 10^{-9},$$

use formula (9.67) for the estimate and obtain

$$C_l^{EE} T_0^2 \sim 7 \cdot 10^{-4} \mu\text{K}^2,$$

This is indeed the correct order of magnitude of E -mode.

We recall that the approximations used in deriving the result (10.58) are valid for $100 \lesssim l \lesssim 1000$. Hence, the right hand side of (10.58) is positive, as should be the case. In this range of multipoles, the amplitude of the oscillating term is of the same order as the monotonic term. Hence, the oscillations in the spectrum C_l^{EE} are rather pronounced, see also Fig 10.4.

Problem 10.11. Find approximately the spectrum C_l^{EE} generated by CDM isocurvature perturbations.

Let us now turn to lower multipoles, $l \ll \eta_0/\eta_r \simeq 30$. The main effect here is due to the reionization epoch. Its smallness is mainly caused by the small optical depth at reionization, $\tau_{rei} \simeq 0.08$.

Since we are talking about large angular scales, the Doppler effect in the temperature fluctuations is negligible. Making use of (10.31) and (10.40), we obtain

$$\begin{aligned} \mathcal{P}_{ab}(\mathbf{k}, \mathbf{n}) &= \frac{3\tau_{rei}}{4\pi} e^{i\mathbf{k}\mathbf{n}(\eta_0-\eta_r)} \int d\mathbf{n}' e^{-i\mathbf{k}\mathbf{n}'(\eta_{rei}-\eta_r)} \left[\Phi(\eta_r) + \frac{1}{4} \delta_\gamma(\eta_r) \right] \\ &\times \left[\frac{1}{2} (1 - (\mathbf{n}\mathbf{n}')^2) g_{ab} - \mathbf{n}' \mathbf{e}_a \cdot \mathbf{n}' \mathbf{e}_b \right]. \end{aligned} \quad (10.59)$$

The integration over the azimuthal angle ϕ' is performed by employing the formulas (10.47), so the non-vanishing components of the polarization tensor are again

determined by the combinations (10.48) and (10.49). The integration over polar angle θ' is again performed using the expansion (F.35) for the exponential factor in the integrand (10.40) and orthogonality property (F.34) of the Legendre polynomials. Also, we can set $(\eta_{rei} - \eta_r) \simeq \eta_{rei}$. In this way we obtain in our reference frame

$$\mathcal{P}_{\theta\theta}(\mathbf{k}, \mathbf{n}) = -\frac{\mathcal{P}_{\phi\phi}(\mathbf{k}, \mathbf{n})}{\sin^2 \theta} = \frac{3\tau_{rei}}{2} \sin^2 \theta \cdot \left(\Phi + \frac{1}{4}\delta_\gamma \right) j_2[k(\eta_{rei} - \eta_r)] e^{i\mathbf{kn}(\eta_0 - \eta_{rei})}. \quad (10.60)$$

In complete analogy to the previous case, the B -mode is not generated.

The main contribution to C_l^{EE} comes again from perturbations with $k \sim l/\eta_0$, so the argument of the spherical Bessel function j_2 in (10.60) is $k\eta_{rei} \sim l\eta_{rei}/\eta_0$. It should not be much larger than 1, since the spherical Bessel functions are small at large argument. Hence, the reionization effect is sizeable at large angular scales only, $l \lesssim l_{rei} = \pi\eta_0/\eta_{rei} \simeq 9$, see (9.83), (9.84) and also Fig. 10.4.

The rest of the calculation is similar to what we have done above. Recalling (9.31) and (9.32), we write the final result,

$$C_l^{EE} = \tau_{rei}^2 \cdot \frac{81\pi}{200} \cdot \frac{(l+2)!}{(l-2)!} \int \frac{dk}{k} \mathcal{P}_\Phi \cdot j_2^2(k\eta_{rei}) \cdot \frac{j_l^2[(\eta_0 - \eta_{rei})k]}{(\eta_0 - \eta_{rei})^4 k^4}. \quad (10.61)$$

It shows that the multipoles C_l^{EE} behave as l^{-4} at $l \gg \pi\eta_0/\eta_{rei}$, so that \mathcal{D}_l^{EE} decay as l^{-2} . Indeed, the integral (10.61) is saturated at $k \sim l/\eta_0$, and in this range of multipoles $k\eta_{rei} \gg 1$. Then $j_2(k\eta_{rei}) \propto (k\eta_{rei})^{-1} \propto l^{-1}$ and $j_l[k(\eta_0 - \eta_{rei})] \propto l^{-1}$ (modulo oscillating factors), hence the above conclusion. The decay of \mathcal{D}_l^{EE} at large l at $l \gtrsim 9$ is clear in Fig. 10.4.

For $l \gg \pi\eta_0/\eta_{rei}$, the integral in (10.61) can be calculated analytically. We make use of the asymptotics (F.10), (F.15) valid for the spherical Bessel functions and replace the rapidly oscillating factors by their averages over periods,

$$j_2^2(k\eta_{rei}) \simeq \frac{1}{2(k\eta_{rei})^2}, \quad j_l^2[(\eta_0 - \eta_{rei})k] \simeq \frac{1}{2(\eta_0 - \eta_{rei})k} \cdot \frac{1}{\sqrt{(\eta_0 - \eta_{rei})^2 k^2 - (l + \frac{1}{2})^2}}.$$

As a result, we obtain for (10.61) and the flat primordial spectrum $\mathcal{P}_\Phi(k) = A_\Phi$,

$$C_l^{EE} = \tau_{rei}^2 \cdot A_\Phi \cdot \frac{81\pi}{800} \cdot \frac{(l+2)!}{(l-2)!} \cdot \frac{1}{(l + \frac{1}{2})^8} \left(\frac{\eta_0}{\eta_{rei}} - 1 \right)^2 \int_1^\infty \frac{du}{u^8 \sqrt{u^2 - 1}} \quad (10.62a)$$

$$\simeq \tau_{rei}^2 \cdot A_\Phi \cdot \frac{81\pi}{1750} \left(\frac{\eta_0}{\eta_{rei}} - 1 \right)^2 \frac{1}{l^4}, \quad (10.62b)$$

where we extended the integration to $k = \infty$, since the integral is saturated at $k \sim l/\eta_0$. The decrease of multipoles $C_l^{EE} \propto l^{-4}$ is explicit now. With $\tau_{rei} \simeq 0.08$ we find numerically

at $l \sim l_{rei}$

$$\mathcal{D}_l^{EE} = \frac{l(l+1)}{2\pi} C_l^{EE} T_0^2 \sim 0.05 \text{ } \mu\text{K}^2 \cdot \left(\frac{10}{l}\right)^2, \quad l \gtrsim 10,$$

which is consistent with numerical result shown in Fig. 10.4.

Problem 10.12. Derive more accurate analytical estimate and show that the spectrum (10.61) has a maximum. Estimate its position.

We now turn to the cross correlation spectrum C_l^{TE} . We begin with the effect of the recombination epoch. To this end, we introduce the quantity $a_{lm}^T(\mathbf{k})$, which is the contribution of perturbations of momentum \mathbf{k} to temperature anisotropy coefficient a_{lm}^T . It follows from (9.4), (9.22) that in our coordinate frame it is given by

$$\begin{aligned} a_{lm}^T(\mathbf{k}) &= 2\pi\delta_{m,0} \cdot \sqrt{\frac{2l+1}{4\pi}} \cdot i^l \\ &\times \left\{ \left(\frac{1}{4}\delta_\gamma + \Phi \right) j_l[(\eta_0 - \eta_r)k] + kv_B j'_l[(\eta_0 - \eta_r)k] \right\}. \end{aligned} \quad (10.63)$$

We omitted the ISW term here, since it is small for the realistic cosmological parameters.

Problem 10.13. Obtain the formula (10.63).

We insert (10.56) and (10.63) into (10.22) and obtain

$$\begin{aligned} C_l^{TE} &= \frac{2\pi}{5} \cdot \sqrt{\frac{(l+2)!}{2(l-2)!}} \Delta\eta_r \cdot \int \frac{d^3k}{(2\pi)^3} \left\{ -\frac{j_l^2[(\eta_0 - \eta_r)k]}{(\eta_0 - \eta_r)^2 k} P_{SW,v}(k, \eta_r) \right. \\ &\quad \left. - \frac{j_l[(\eta_0 - \eta_r)k] j'_l[(\eta_0 - \eta_r)k]}{(\eta_0 - \eta_r)^2 k} P_v(k, \eta_r) \right\}, \end{aligned} \quad (10.64)$$

where $P_{SW,v}$ is the cross correlation power spectrum defined by

$$\left\langle \left(\frac{1}{4}\delta_\gamma + \Phi \right) (\mathbf{k}, \eta) \cdot k' v_B^*(\mathbf{k}', \eta) \right\rangle = \frac{1}{(2\pi)^3} P_{SW,v}(k, \eta) \delta(\mathbf{k} - \mathbf{k}').$$

Let us discuss in general terms the consequences of (10.64) for the adiabatic scalar perturbations at $l \lesssim 1000$. This discussion is quite similar to that given in Section 9.2.3.

The range $l \ll l^{(r)} \simeq 100$ corresponds to the superhorizon perturbations at recombination. The velocity is small for these perturbations, so the polarization is small too.

In the range $l = 100 - 1000$, we use the results (9.65) and (9.77). The product $j_l[(\eta_0 - \eta_r)k] \cdot j'_l[(\eta_0 - \eta_r)k]$ rapidly oscillates as a function of k , so the second term

in braces in (10.64) can be neglected. It follows from (9.65) and (9.77) that the cross correlation power spectrum $P_{SW,v}(k, \eta_r)$ oscillates as a function of k around zero,

$$P_{SW,v} = -P_\Phi \frac{\sqrt{3}}{\sqrt{1 + R_B(\eta_r)}} \left(\frac{A^2}{2} \sin 2kr_s - AB \sin kr_s \right). \quad (10.65)$$

Therefore, C_l^{TE} also oscillates as a function of l around zero value. The distance between maxima and minima is determined by the first term in parenthesis in (10.65); with the standard relation $k \simeq l/\eta_0$, we find that this distance is

$$\Delta l \simeq \frac{\pi r_s}{2\eta_0} = \frac{\pi}{2} l^{(r)} \simeq 150.$$

The second term in (10.65) yields the modulation of the oscillations: odd peaks (corresponding to momenta $k \sim (\pi + 2\pi n)/r_s$) are enhanced as compared to even peaks. This modulation is the strongest for the first peaks, since the function $A(l) \equiv A(k = l/\eta_0)$ is independent of l at large l , while $B(l)$ decreases as $1/l^2$, see (9.68). All these properties are seen in Fig. 10.4. Note that Fig. 10.4 shows the absolute value of \mathcal{D}_l^{TE} .

Let us consider the cross correlation spectrum in some details at $100 \ll l \lesssim 1000$. To this end, we treat $l^{(r)}/l$ as a small parameter. Let us assume that the primordial power spectrum is flat; the generalization to slightly tilted primordial spectrum is given by (9.38), as usual.

We omit the second term in the integrand in (10.64), replace rapidly oscillating factor $j_l^2[k(\eta_0 - \eta_r)]$ by its average value and use the asymptotics (F.15). We change the integration variable to $u = k\eta_0/l$ and obtain

$$\begin{aligned} C_l^{TE} &= \frac{\pi}{5} \frac{\sqrt{3}\Delta\eta_r}{\sqrt{1 + R(\eta_r)}} \sqrt{\frac{(l+2)!}{2(l-2)!}} \cdot \frac{A_\Phi}{l^3} \\ &\times \int_1^\infty \frac{du}{u^3\sqrt{u^2-1}} \left[\frac{A^2}{2} \sin \left(2u \frac{l}{l^{(r)}} \right) - AB \sin \left(u \frac{l}{l^{(r)}} \right) \right]. \end{aligned} \quad (10.66)$$

The integral in (10.66) is saturated at $u \approx 1$ in a narrow region $\Delta u \sim l^{(r)}/l$. We then use the known formula for the integral

$$\int_1^\infty \frac{du}{u^3\sqrt{u^2-1}} \sin \left(u \frac{2l}{l^{(r)}} \right) \simeq \frac{1}{\sqrt{2}} \int_1^\infty \frac{du}{\sqrt{u-1}} \sin \left(u \frac{2l}{l^{(r)}} \right) = \frac{1}{2} \sqrt{\frac{\pi l^{(r)}}{l}} \sin \left(\frac{2l}{l^{(r)}} + \frac{\pi}{4} \right).$$

The second term in (10.66) is evaluated in a similar way. Hence, our approximate expression is

$$\begin{aligned} C_l^{TE} &= \frac{\pi}{5} \frac{\sqrt{3}}{\sqrt{1 + R(\eta_r)}} \sqrt{\frac{(l+2)!}{2(l-2)!}} \cdot \frac{A_\Phi}{l^3} \\ &\times \sqrt{\frac{\pi l^{(r)}}{2l}} \left[\frac{A^2}{2\sqrt{2}} \sin \left(2 \frac{l}{l^{(r)}} + \frac{\pi}{4} \right) - AB \sin \left(\frac{l}{l^{(r)}} + \frac{\pi}{4} \right) \right]. \end{aligned} \quad (10.67)$$

where $A = A(l)$, $B = B(l)$. In the high multipole region, A is independent of l , while $B \propto 1/l^2$, so the first term dominates. The multipoles C_l^{TE} oscillate around zero value, and the amplitude of these oscillations decreases as $C_l^{TE} \propto l^{-3/2}$. This is consistent with the numerical result shown in Fig. 10.4. The second term in (10.67) gives rise to the modulation of these oscillations at relatively small l . The first and second terms in (10.67) have maxima and minima at

$$l_{(max)} = \left(\frac{\pi}{8} + \pi n \right) l^{(r)}, \quad \text{and} \quad l_{(min)} = \left(\frac{5\pi}{8} + \pi n \right) l^{(r)},$$

and

$$l_{(max)} = \left(\frac{5\pi}{4} + 2\pi n' \right) l^{(r)}, \quad \text{and} \quad l_{(min)} = \left(\frac{\pi}{4} + 2\pi n' \right) l^{(r)},$$

respectively, where $n, n' = 0, 1, 2, \dots$. We see that the maxima of the two terms in (10.67) are close to each other at odd n , while the maximum of the first term is close to the minimum of the second term for even n . So, the odd and even maxima are enhanced and suppressed, respectively. This effect is pronounced for the first maxima only, when $B(l)$ is not very small. The first maximum at

$$l \simeq \left(\frac{\pi}{8} + \pi \right) l^{(r)} \simeq 340$$

is enhanced, and the second one at

$$l \simeq \left(\frac{\pi}{8} + 2\pi \right) l^{(r)} \simeq 645$$

is suppressed. The first minima of the two terms are not close to each other, so the depths of the minima are unmodulated. All this agrees with the result shown in Fig 10.5. We note that numerical evaluation of the integral (10.66) shows that all extrema are somewhat shifted towards smaller l as compared to our analytical estimate. This shift gets smaller at larger l where our approximation is more adequate.

The result (10.67) is analogous to the formula (9.75) for the oscillations in the temperature multipoles C_l^{TT} in the same range of harmonics, $100 \ll l \lesssim 1000$. The phases of the oscillatory terms in (10.67) are shifted by $\pi/2$ with respect to the phases in (9.75), so that C_l^{TT} has extrema at approximately the same l as zeroes of C_l^{TE} . This is in good agreement with both numerical analysis and experimental data at sufficiently large l , see Fig. 10.5. The modulation of oscillations at smaller l makes this correspondence less precise. This is especially true for the first peak in the temperature anisotropy spectrum.

Problem 10.14. Find approximately the spectrum C_l^{TE} generated by CMB isocurvature mode.

Let us consider low multipoles which are dominated by the reionization effect. Proceeding as before, we obtain instead of (10.61),

$$C_l^{TE} = \tau_{rei} \frac{27\pi}{100} \sqrt{\frac{(l+2)!}{2(l-2)!}} \cdot \int \frac{dk}{k} \mathcal{P}_\Phi j_2(k\eta_{rei}) \frac{j_l^2[(\eta_0 - \eta_{rei})k]}{(\eta_0 - \eta_{rei})^2 k^2}. \quad (10.68)$$

This shows that the reionization contribution is sizeable again at low multipoles only.

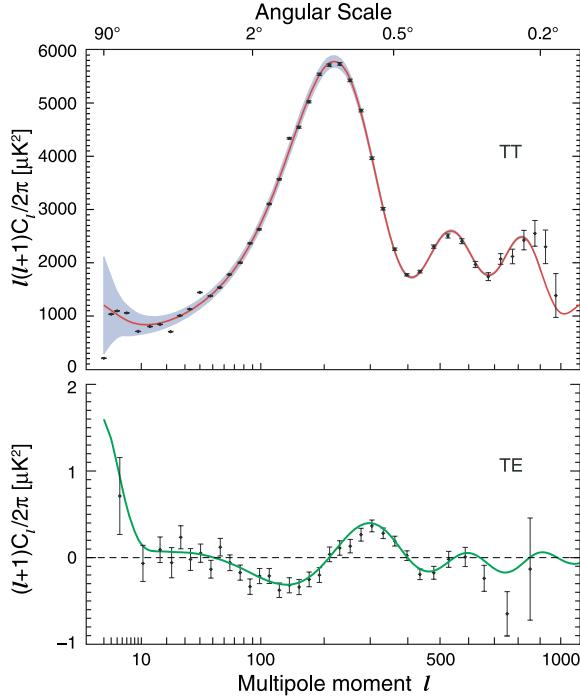


Fig. 10.5 The temperature anisotropy spectrum C_l^{TT} and cross correlation spectrum C_l^{TE} obtained by WMAP experiment [68]. The positions of extrema in C_l^{TT} are near the positions of zeroes in C_l^{TE} .

The integral in (10.68) can again be analytically estimated for $l \gg 1$. We use the approximation (F.15) for the spherical Bessel function j_l , average over its oscillations and obtain

$$C_l^{TE} = \tau_{rei} \frac{27\pi}{200} \frac{1}{(l + \frac{1}{2})^4} \sqrt{\frac{(l+2)!}{2(l-2)!}} \int_1^\infty \frac{du}{u^4} \frac{j_2 \left[u^{\frac{\eta_{rei}}{\eta_0 - \eta_{rei}}} (l + \frac{1}{2}) \right]}{\sqrt{u^2 - 1}} \mathcal{P}_\Phi, \quad (10.69)$$

where $u = k(\eta_0 - \eta_{rei}) / (l + \frac{1}{2})$. Clearly, the integral in (10.69) is saturated at $u \sim 1$ in the interval $\Delta u \sim (\eta_0 - \eta_{rei}) / ((l + \frac{1}{2}) \eta_{rei}) \ll 1$. The form of the integrand implies that C_l^{TE} are positive at small enough l . This is true at least when the spherical Bessel function j_2 in the integrand is positive at $u \sim 1$, i.e., for $l \lesssim 2\pi(\eta_0 - \eta_{rei})/\eta_{rei} \sim 12$. This region is of particular interest, since the reionization effect is irrelevant for higher multipoles anyway. In this region, the contribution (10.69) decreases faster than l^{-2} , since the range of integration in (10.69) decreases as l grows. Hence, \mathcal{D}_l^{TE} decreases, which is consistent with Fig. 10.4.

Problem 10.15. Evaluate the integral (10.69) numerically and by comparing it with Fig. 10.4 show that the approximation (10.69) indeed gives correct estimate for the angular spectrum C_l^{TE} .

Problem 10.16. Find the contribution of the reionization epoch to C_l^{TE} at $l_{rei} \ll l \ll l^{(r)}$.

10.3.3 Tensor perturbations

We now turn to the possible effect of tensor perturbations on CMB polarization. As the main example, we study the generation of polarization at the recombination epoch. Making use of (10.30) we find that in the approximation (10.39), the contribution of tensor perturbations with momentum \mathbf{k} to the polarization tensor is

$$\begin{aligned} \mathcal{P}_{ab}(\mathbf{k}, \mathbf{n}) &= \frac{3}{8\pi} \int d\mathbf{n}' \left[\frac{1}{2} \left(1 - (\mathbf{n}\mathbf{n}')^2 \right) g_{ab} - \mathbf{n}' \mathbf{e}_a \cdot \mathbf{n}' \mathbf{e}_b \right] \cdot e^{i\mathbf{k}\mathbf{n}(\eta_0 - \eta_r)} \cdot \Delta\eta_r \\ &\times \sum_{A=+, \times} n'_i \epsilon_{ij}^{(A)} n'_j \cdot h^{(A)'}(\mathbf{k}, \eta_r), \end{aligned} \quad (10.70)$$

while the polarization tensor itself is given by (10.45).

Let us again work in the spherical coordinate frame with azimuthal axis directed along \mathbf{k} . In this frame, the gravity waves of the two polarizations obey (9.52). The product $n'_i n'_j$ in (10.70) contains trigonometric functions of $2\theta'$, so the only non-vanishing integrals over the azimuthal angle in (10.70) come from the terms in (10.47) that depend on $2\theta'$. The latter are

$$\begin{aligned} \frac{1}{2} \left(1 - (\mathbf{n}\mathbf{n}')^2 \right) g_{\theta\theta} - \mathbf{n}' \mathbf{e}_\theta \cdot \mathbf{n}' \mathbf{e}_\theta &\longrightarrow -\frac{1}{4} (1 + \cos^2 \theta) \cdot \sin^2 \theta' \cdot \cos 2(\phi - \phi'), \\ \frac{1}{2} \left(1 - (\mathbf{n}\mathbf{n}')^2 \right) g_{\phi\phi} - \mathbf{n}' \mathbf{e}_\phi \cdot \mathbf{n}' \mathbf{e}_\phi &\longrightarrow \frac{1}{4} \sin^2 \theta \cdot (1 + \cos^2 \theta) \cdot \sin^2 \theta' \cdot \cos 2(\phi - \phi'), \\ -\mathbf{n}' \mathbf{e}_\theta \cdot \mathbf{n}' \mathbf{e}_\phi &\longrightarrow \frac{1}{2} \sin \theta \cdot \cos \theta \cdot \sin^2 \theta' \sin 2(\phi - \phi'). \end{aligned}$$

By comparing these expressions with (10.13) we find that the Stokes parameters generated by a tensor perturbation of polarization (+) are

$$\begin{aligned} Q_{(+)} &= -\frac{3}{32\pi} \Delta\eta_r \cdot h^{(+)' \cdot} (1 + \cos^2 \theta) e^{ik(\eta_0 - \eta_r) \cos \theta} \\ &\times \int_0^\pi d\theta' \cdot \sin^5 \theta' \int_0^{2\pi} d\phi' \cdot \cos 2\phi' \cdot \cos 2(\phi - \phi'), \end{aligned} \quad (10.71)$$

$$\begin{aligned} U_{(+)} &= -\frac{3}{16\pi} \cdot \Delta\eta_r \cdot h^{(+)' \cdot} \cos \theta e^{ik(\eta_0 - \eta_r) \cos \theta} \\ &\times \int_0^\pi d\theta' \cdot \sin^5 \theta' \int_0^{2\pi} d\phi' \cdot \cos 2\phi' \cdot \sin 2(\phi - \phi'). \end{aligned} \quad (10.72)$$

Similar formulas with $\sin 2\phi'$ substituted for $\cos 2\phi'$ are valid for a tensor perturbation of polarization (\times). We perform the integrations in (10.71), (10.72), and obtain

$$Q_{(+)}(\mathbf{k}, \mathbf{n}) = -\frac{1}{10} \Delta\eta_r \cdot h^{(+)' \cdot} e^{ik(\eta_0 - \eta_r) \cos \theta} \cdot (1 + \cos^2 \theta) \cdot \cos 2\phi, \quad (10.73a)$$

$$U_{(+)}(\mathbf{k}, \mathbf{n}) = -\frac{1}{5} \Delta\eta_r \cdot h^{(+)' \cdot} e^{ik(\eta_0 - \eta_r) \cos \theta} \cdot \cos \theta \cdot \sin 2\phi, \quad (10.73b)$$

and similarly for $Q_{(\times)}$ and $U_{(\times)}$ with the interchange $\cos 2\phi \longleftrightarrow \sin 2\phi$. We see that the tensor perturbations contribute to both of the Stokes parameters. Hence, both E - and B -modes are generated.

To calculate the contribution of tensor modes with momentum \mathbf{k} to the coefficients $a_{lm}^{E,B}$, we have to evaluate the integrals over the unit sphere, defined by (10.10), (10.11). Since the components of the polarization tensor depend on the azimuthal angle ϕ through the factors $\cos 2\phi$ or $\sin 2\phi$, the integrals (10.10), (10.11) do not vanish for spherical harmonics $Y_{l,\pm 2}$ only. Hence, in our coordinate frame the only non-vanishing coefficients are $a_{l,\pm 2}^{E,B}$. The integration over the polar angle θ is performed by making use of the definition (F.39) of spherical harmonics, expansion of the exponential factor in (10.73) in the Legendre polynomials and properties of the Legendre polynomials collected in Appendix F.2. The result for the contribution of the polarization (+) is (its derivation is given below)

$$\begin{aligned} a_{lm,(+)}^E(\mathbf{k}) &= \frac{1}{10} \Delta \eta_r h^{(+)}' \cdot i^l \cdot \sqrt{2\pi(2l+1)} \cdot (\delta_{m,+2} + \delta_{m,-2}) \\ &\quad \times \left\{ \frac{(l+2)(l+1)}{(2l-1)(2l+1)} j_{l-2}[(\eta_0 - \eta_r)k] - \frac{6(l+2)(l-1)}{(2l-1)(2l+3)} j_l[(\eta_0 - \eta_r)k] \right. \\ &\quad \left. + \frac{l(l-1)}{(2l+1)(2l+3)} j_{l+2}[(\eta_0 - \eta_r)k] \right\}, \end{aligned} \quad (10.74)$$

$$\begin{aligned} a_{lm,(+)}^B(\mathbf{k}) &= \frac{1}{5} \Delta \eta_r h^{(+)}' \cdot i^l \cdot \sqrt{2\pi(2l+1)} \cdot (\delta_{m,+2} - \delta_{m,-2}) \\ &\quad \times \left\{ \frac{l+2}{2l+1} j_{l-1}[(\eta_0 - \eta_r)k] - \frac{l-1}{2l+1} j_{l+1}[(\eta_0 - \eta_r)k] \right\}. \end{aligned} \quad (10.75)$$

We see that

$$a_{lm,(+)}^E(\mathbf{k}) \propto (\delta_{m,+2} + \delta_{m,-2}), \quad a_{lm,(+)}^B(\mathbf{k}) \propto (\delta_{m,+2} - \delta_{m,-2}).$$

This implies that E - and B -modes are uncorrelated, i.e., $C_l^{EB} \propto \sum_m a_{lm,(+)}^E a_{lm,(+)}^B = 0$. Furthermore, the contribution (9.53) of the tensor perturbations to the temperature anisotropy has the following structure,

$$a_{lm,(+)}^T \propto (\delta_{m,+2} + \delta_{m,-2}), \quad a_{lm,(\times)}^T \propto (\delta_{m,+2} - \delta_{m,-2}).$$

Therefore, $C_l^{TB} = 0$, while the cross correlation spectrum C_l^{TE} does not vanish. The correlation between scalar and tensor contributions to CMB polarization is also absent.

Problem 10.17. Show that irrespective of correlations between primordial scalar and tensor perturbations, the spectrum C_l^{TE} is the uncorrelated sum of scalar and tensor contributions. Same for C_l^{EE} .

For tensor perturbations of polarization (\times) one finds (the replacement $h^{(+)} \rightarrow h^{(\times)}$ is assumed)

$$a_{l,\pm 2,(\times)}^E = \mp i a_{l,\pm 2,(+)}^E, \quad a_{l,\pm 2,(\times)}^B = \pm i a_{l,\pm 2,(+)}^B, \quad (10.76)$$

so that

$$a_{lm,(\times)}^E \propto (\delta_{m,+2} - \delta_{m,-2}), \quad a_{lm,(\times)}^B \propto (\delta_{m,+2} + \delta_{m,-2}).$$

Hence, the contributions of modes (+) and (\times) to the polarization tensor are orthogonal to each other.

Let us derive the formulas (10.74) and (10.75). To this end, we make use of the second integral representation for $a_{lm}^{E,B}$, see (10.10), (10.11). It follows from (10.73a), (10.73b), (10.13) that the contribution of the mode of momentum \mathbf{k} to the polarization tensor with upper indices is (recall that we use the standard metric (10.12) on the unit sphere)

$$\mathcal{P}_{(+)}^{ab}(\mathbf{n}, \mathbf{k}) = -\frac{1}{20} \Delta \eta_r \cdot h^{(+)' } \begin{pmatrix} (1 + \cos^2 \theta) \cos 2\phi & -2 \frac{\cos \theta}{\sin \theta} \sin 2\phi \\ -2 \frac{\cos \theta}{\sin \theta} \sin 2\phi & -\frac{1 + \cos^2 \theta}{\sin^2 \theta} \cos 2\phi \end{pmatrix}.$$

Then we obtain from (10.45), (10.10) and (10.16) for E -mode and from (10.45), (10.11) and (10.17) for B -mode, respectively,

$$a_{lm,(+)}^E(\mathbf{k}) = -\frac{1}{10} \Delta \eta_r \cdot h^{(+)' } \sqrt{\frac{(l-2)!}{2 \cdot (l+2)!}} \int d\mathbf{n} e^{i(\eta_0 - \eta_r)k \cos \theta} [W_{lm}^* \cdot (1 + \cos^2 \theta) \times \cos 2\phi - 2X_{lm}^* \cos \theta \sin 2\phi], \quad (10.77)$$

$$a_{lm,(+)}^B(\mathbf{k}) = -\frac{1}{10} \Delta \eta_r \cdot h^{(+)' } \sqrt{\frac{(l-2)!}{2 \cdot (l+2)!}} \int d\mathbf{n} e^{i(\eta_0 - \eta_r)k \cos \theta} [X_{lm}^* \cdot (1 + \cos^2 \theta) \times \cos 2\phi + 2W_{lm}^* \cos \theta \sin 2\phi]. \quad (10.78)$$

We now recall Eq. (F.40) for the spherical harmonics and rewrite W_{lm}^* given in (10.14) as follows,

$$W_{lm}^* = \left[-2 \frac{\cos \theta}{\sin \theta} \frac{\partial}{\partial \theta} + 2 \frac{m^2}{\sin^2 \theta} - l(l+1) \right] Y_{lm}^*.$$

As we have noticed, the integrals over the azimuthal angle do not vanish for $m = \pm 2$ only. The relevant spherical harmonics (F.39) are conveniently expressed in terms of the Legendre polynomials,

$$\begin{aligned} Y_{l,\pm 2} &= \sqrt{\frac{(l-2)!}{(l+2)!}} \sqrt{\frac{2l+1}{4\pi}} (1-x^2) \frac{d^2 P_l}{dx^2} \cdot e^{\pm 2i\phi} \\ &= \sqrt{\frac{(l-2)!}{(l+2)!}} \sqrt{\frac{2l+1}{4\pi}} \left[2x \frac{dP_l}{dx} - l(l+1) P_l \right] \cdot e^{\pm 2i\phi}, \end{aligned} \quad (10.79)$$

where $x = \cos \theta$ and the second equality is obtained by making use of Eq. (F.23). Upon integration in (10.79) over the azimuthal angle we have

$$\begin{aligned} a_{l,\pm 2,(+)}^E &= -\frac{1}{10} \Delta \eta_r \cdot h^{(+)\prime} \frac{\pi}{\sqrt{2}} \sqrt{\frac{2l+1}{4\pi}} \frac{(l-2)!}{(l+2)!} \cdot I_l^E, \\ a_{l,\pm 2,(+)}^B &= \pm \frac{i}{10} \Delta \eta_r \cdot h^{(+)\prime} \frac{\pi}{\sqrt{2}} \sqrt{\frac{2l+1}{4\pi}} \frac{(l-2)!}{(l+2)!} \cdot I_l^B, \end{aligned}$$

where

$$\begin{aligned} I_l^E &= \int_{-1}^1 dx e^{i(\eta_0 - \eta_r)kx} \\ &\times \left[2x^3 \frac{d}{dx} - 6x \frac{d}{dx} + \frac{8}{1-x^2} - (1+x^2)(l+1)l \right] \left[2x \frac{dP_l}{dx} - l(l+1)P_l \right] \\ &= \frac{(l+2)!}{(l-2)!} \int_{-1}^1 dx e^{i(\eta_0 - \eta_r)kx} \left[\frac{(l+2)(l+1)}{(2l-1)(2l+1)} P_{l-2} \right. \\ &\quad \left. + \frac{6(l+2)(l-1)}{(2l-1)(2l+3)} P_l + \frac{l(l-1)}{(2l+3)(2l+1)} P_{l+2} \right], \end{aligned} \quad (10.80)$$

and

$$\begin{aligned} I_l^B &= 2 \int_{-1}^1 dx e^{i(\eta_0 - \eta_r)kx} \\ &\times \left[-2 \frac{d}{dx} + \frac{4x}{1-x^2} + 2x - xl(l+1) \right] \left[2x \frac{dP_l}{dx} - l(l+1)P_l \right], \\ &= 2 \frac{(l+2)!}{(l-2)!} \int_{-1}^1 dx e^{i(\eta_0 - \eta_r)kx} \left(\frac{l+2}{2l+1} P_{l-1} + \frac{l-1}{2l+1} P_{l+1} \right). \end{aligned} \quad (10.81)$$

We used here Eq. (F.23) and recurrence relations (F.25), (F.26) for the Legendre polynomials. Finally, we expand the exponential function in the Legendre polynomials, perform the integration over x using the orthogonality relation (F.34) and obtain the results (10.74), (10.75).

Problem 10.18. Prove the relations (10.76).

The important property of tensor perturbations is that they generate B -mode. So, let us study the angular spectrum of B -mode in some details. We find from (10.75) that the perturbations of polarization (+) give²

$$C_l^{BB} = \frac{1}{2l+1} \sum_m \int d^3k \langle a_{lm}^B(\mathbf{k}) a_{lm}^{B*}(\mathbf{k}) \rangle \quad (10.82a)$$

$$= \frac{4\pi}{25} \Delta \eta_r^2 \int \frac{d^3k}{(2\pi)^3} \hat{P}_T^{(+)}(k, \eta_r) \cdot \left[\frac{l+2}{2l+1} j_{l-1}(k\eta_0) - \frac{l-1}{2l+1} j_{l+1}(k\eta_0) \right]^2, \quad (10.82b)$$

²Formula (10.82a) is somewhat loose, as we have in mind that one integration over momenta is already performed with the use of the δ -function in the two-point correlation function of $h^{(+)\prime}(\mathbf{k}, \eta)$.

where $\hat{P}_T^{(+)}(k, \eta)$ is the power spectrum of $h^{(+)\prime}(\mathbf{k}, \eta)$ (notice that this is the power spectrum of the *derivative* of the metric perturbation with respect to conformal time), and we replaced $(\eta_0 - \eta_r)$ by η_0 . The contribution of polarization (\times) to C_l^{BB} coincides with that of polarization $(+)$ modulo the replacement of $\hat{P}_T^{(+)}$ by $\hat{P}_T^{(\times)}$. Realistically, the power spectra of perturbations with polarizations $(+)$ and (\times) are the same, so the total effect of tensor modes is twice the expression (10.82).

We now use the formula (9.47b) for tensor amplitudes and the explicit form (5.44) of the correlator. We obtain finally

$$\begin{aligned} C_l^{BB} &= \frac{4\pi}{25} \Delta\eta_r^2 \int_0^\infty \frac{dk}{k} \mathcal{P}_T(k) \mathfrak{h}'^2(k, \eta_r) \\ &\times \left[\frac{l+2}{2l+1} j_{l-1}(k\eta_0) - \frac{l-1}{2l+1} j_{l+1}(k\eta_0) \right]^2. \end{aligned} \quad (10.83)$$

Perturbations that are superhorizon at recombination give small contribution to the integral (10.83) even at low l , since the gravitational field evolves slowly at that time. The main contribution at $l \lesssim \eta_0/\eta_r \sim 50$ comes from the perturbations that just enter the horizon at recombination. Their momenta are $k\eta_r \sim 1$, and they obey $\mathfrak{h}' \sim k \sim \eta_r^{-1}$, see (9.48), while $j_l(k\eta_0) \propto (k\eta_0)^{-1}$. This gives for the flat primordial spectrum

$$C_l^{BB} = \text{const} \cdot A_T \left(\frac{\Delta\eta_r}{\eta_0} \right)^2 = \text{const} \cdot A_T \left(\frac{\Delta\eta_r}{\eta_r} \right)^2 \left(\frac{\eta_r}{\eta_0} \right)^2, \quad l \lesssim 50,$$

where the numerical constant is actually rather small. Since C_l^{BB} are almost independent of l at small l , the angular spectrum $\mathcal{D}_l^{BB} = T_0^2 \cdot \frac{l(l+1)}{2\pi} C_l^{BB}$ increases as l^2 in this range of multipoles, in agreement with Fig. 10.4.

For $50 < l < 1000$, the integral (10.83) is saturated, as usual, at $k \sim l/\eta_0$. These modes obey $\mathfrak{h}' \sim 1/(k\eta_r^2) \sim \eta_0/(l\eta_r^2)$, while $j_l(k\eta_0) \propto 1/l$. This gives the estimate

$$C_l^{BB} = \text{const} \cdot A_T \left(\frac{\Delta\eta_r}{\eta_r} \right)^2 \left(\frac{\eta_0}{\eta_r} \right)^2 \frac{1}{l^4}, \quad 50 < l < 1000.$$

It shows that \mathcal{D}_l^{BB} decreases as l^{-2} . This is also seen in Fig. 10.4. The oscillatory behavior of C_l^{BB} is, as usual, due to the oscillations of $\mathfrak{h}'(\eta_r)$ as a function of momentum.

Let us refine the above estimates. We make use of the recurrence relations (F.3), (F.4), and obtain from (9.48)

$$\mathfrak{h}'(k, \eta_r) = \frac{d}{d\eta_r} \left[\frac{3j_1(k\eta_r)}{k\eta_r} \right] = -\frac{3j_2(k\eta_r)}{\eta_r}. \quad (10.84)$$

Again using the recurrence relations (F.3) and (F.4), we rewrite the expression in braces in (10.83) as follows,

$$\frac{l+2}{2l+1} j_{l-1}(k\eta_0) - \frac{l-1}{2l+1} j_{l+1}(k\eta_0) = \frac{d}{d(k\eta_0)} j_l(k\eta_0) + \frac{2j_l(k\eta_0)}{k\eta_0}. \quad (10.85)$$

Let us first consider B -mode at $10 \lesssim l \lesssim \eta_0/\eta_r \simeq 50$. Since $l \gg 1$, the spherical Bessel functions j_l and their derivatives are approximated by the asymptotic formulas (F.15) and (F.22), respectively. Since the spherical Bessel functions oscillate, and the phases of the asymptotic expressions (F.15) and (F.22) differ by $\pi/2$, the interference term in (10.83), which is proportional to $j_l \cdot j'_l$, can be neglected. We average the oscillating factors in j_l^2 and $j_l'^2$ and obtain

$$\begin{aligned} C_l^{BB} &= \frac{36\pi}{25} \frac{\Delta\eta_r^2}{\eta_r^2} \int_{(l+1/2)/\eta_0}^{\infty} \frac{dk}{k} \mathcal{P}_T(k) j_2^2(k\eta_r) \\ &\times \left[\frac{\sqrt{(k\eta_0)^2 - (l + \frac{1}{2})^2}}{2(k\eta_0)^3} + \frac{2}{(k\eta_0)^3 \sqrt{(k\eta_0)^2 - (l + \frac{1}{2})^2}} \right]. \end{aligned} \quad (10.86)$$

The first term in square brackets in (10.86) dominates at $l \gg 1$.

Let us specialize to the flat primordial power spectrum of tensor perturbations. Since $j_2(x) \propto x^2$ at small x , see (F.8), the integral in (10.86) is saturated at $k\eta_r \sim 1$. We introduce the integration variable $y = k\eta_r$ and write

$$C_l^{BB} \simeq \frac{18\pi}{25} \left(\frac{\Delta\eta_r}{\eta_0} \right)^2 \cdot A_T \cdot \int_{\eta_r(l + \frac{1}{2})}^{\infty} \frac{dy}{y^3} j_2^2(y).$$

The lower limit of integration here can be extended to zero, and we arrive at the known integral

$$\int_0^{\infty} \frac{dy}{y^3} j_2^2(y) = \frac{1}{72}.$$

Hence, we obtain

$$\mathcal{D}_l^{BB} = T_0^2 \cdot \frac{l(l+1)}{2\pi} C_l^{BB} \simeq T_0^2 \cdot \frac{A_T}{200} \left(\frac{\Delta\eta_r}{\eta_r} \right)^2 \left(\frac{\eta_r}{\eta_0} \right)^2 \cdot l^2.$$

The angular spectrum \mathcal{D}_l^{BB} indeed increases as $\mathcal{D}_l^{BB} \propto l^2$. Our expression is in reasonable agreement with the complete numerical result shown in Fig. 10.4, though the numerical evaluation of the integral (10.86) shows somewhat slower increase than l^2 . For $l = 30$, $r = A_T/A_R = 0.38$ and $A_R \simeq 2.5 \cdot 10^{-9}$ our formula gives

$$\mathcal{D}_{30}^{BB} \approx 2 \cdot 10^{-2} \mu\text{K}^2,$$

in qualitative agreement with the exact result shown in Fig. 10.4.

We now obtain the estimate for smaller angular scales, $l \gg \eta_0/\eta_r \simeq 50$, where the main contribution comes from the perturbations which enter the horizon before recombination. We use Eq. (10.86), approximate j_2 at $l\eta_r/\eta_0 \gg 1$ by its large-argument asymptotic expression (F.10) and find for the flat primordial spectrum

$$C_l^{BB} = \frac{9\pi}{25} \left(\frac{\Delta\eta_r}{\eta_r} \right)^2 \left(\frac{\eta_0}{\eta_r} \right)^2 \frac{A_T}{(l + \frac{1}{2})^4} \int_1^{\infty} \frac{du}{u} \frac{\sqrt{u^2 - 1}}{u^5} \left\{ 1 - \cos \left[2u \frac{\eta_r}{\eta_0} \left(l + \frac{1}{2} \right) \right] \right\}, \quad (10.87)$$

where $u = k\eta_0/l$. This shows that the angular spectrum C_l^{BB} oscillates at $l > 50$ as function of l with the period $l \simeq \pi\eta_0/\eta_r \simeq 160$, in agreement with the numerical result shown in the right panel of Fig. 10.4. The average value around which C_l^{BB} oscillates, as well as the

amplitude of oscillations, rapidly decrease with l . The average value is found by making use of the known integral

$$\int_1^\infty \frac{du}{u^6} \sqrt{u^2 - 1} = \frac{2}{15}.$$

We find approximately

$$\mathcal{D}_l^{BB} \simeq \frac{3T_0^2}{125} \left(\frac{\Delta\eta_r}{\eta_r} \right)^2 \left(\frac{\eta_0}{\eta_r} \right)^2 \cdot \frac{A_T}{(l + \frac{1}{2})^2}.$$

As expected, for $l > 50$ the angular spectrum \mathcal{D}_l^{BB} decreases as $\mathcal{D}_l^{BB} \propto l^{-2}$, cf. Fig. 10.4. For $l = 500$, $r = A_T/A_R = 0.38$ and $A_R \simeq 2.5 \cdot 10^{-9}$ we obtain numerically

$$\mathcal{D}_{500}^{BB} \approx 2.5 \cdot 10^{-3} \mu\text{K}^2,$$

in agreement with the numerical result shown in Fig. 10.4.

At large angular scales, $l \lesssim 20$, the main effect is again due to the reionization. We find this contribution by making use of the approximate formula (10.44). The contribution of perturbations with momentum \mathbf{k} to the polarization tensor is (cf. (10.70))

$$\begin{aligned} \mathcal{P}_{ab}(\mathbf{k}, \mathbf{n}) &= \frac{3}{8\pi} \int d\mathbf{n}' \left[\frac{1}{2}(1 - (\mathbf{n}\mathbf{n}')^2)g_{ab} - \mathbf{n}'\mathbf{e}_a \cdot \mathbf{n}'\mathbf{e}_b \right] \cdot e^{i\mathbf{k}\mathbf{n}(\eta_0 - \eta_{rei})} \cdot \tau_{rei} \\ &\times \sum_{A=+, \times} n'_i \epsilon_{ij}^{(A)} n'_j \cdot \int_{\eta_r}^{\eta_{rei}} d\eta h^{(A)'}(\mathbf{k}, \eta) \cdot e^{-i\mathbf{k}\mathbf{n}'(\eta_{rei} - \eta)}. \end{aligned}$$

Further analysis is similar in many respects to the one we have performed for the recombination epoch, so we only highlight the peculiarities. We consider tensor perturbation $h^{(+)}$ for definiteness. The expressions for $Q_{(+)}$ and $U_{(+)}$ analogous to (10.71) and (10.72) involve new functions of the polar angle θ' , so the relevant integral is now

$$\int_0^\pi d\theta' \sin^5 \theta' \cdot e^{-ik(\eta_{rei} - \eta) \cos \theta'}.$$

We use the representation (F.35) for the exponential function, and evaluate integral over the polar angle θ' by making use of the equality

$$\int_{-1}^{+1} (1 - x^2)^2 P_l(x) dx = \frac{16}{15} \left(\delta_{l,0} - \frac{2}{7} \delta_{l,2} + \frac{1}{21} \delta_{l,4} \right). \quad (10.88)$$

Problem 10.19. Evaluate the integral (10.88) in two ways: (1) Use the Rodrigues formula (F.29) and then integrate by parts; (2) Expand $(1 - x^2)^2$ in the Legendre polynomials and then use the orthogonality relation (F.34).

We obtain in this way

$$\begin{aligned} & \int_0^\pi d\theta' \sin \theta' \sin^4 \theta' e^{-ik(\eta_{rei}-\eta) \cos \theta'} \\ &= 16 \left\{ \frac{j_0[(\eta_{rei}-\eta)k]}{3 \cdot 5} + \frac{2j_2[(\eta_{rei}-\eta)k]}{3 \cdot 7} + \frac{j_4[(\eta_{rei}-\eta)k]}{5 \cdot 7} \right\} \\ &= 16 \frac{j_2[(\eta_{rei}-\eta)k]}{k^2 (\eta_{rei}-\eta)^2}, \end{aligned}$$

where we used the equality between the expressions (9.53) and (9.54). As a result, the Stokes parameters are (cf. (10.73))

$$\begin{aligned} Q_{(+)}(\mathbf{k}, \mathbf{n}) &= -\frac{3}{2} \tau_{rei} \int_{\eta_r}^{\eta_{rei}} d\eta h^{(+)\prime}(\mathbf{k}, \eta) \frac{j_2[(\eta_{rei}-\eta)k]}{k^2 (\eta_{rei}-\eta)^2} \\ &\quad \times e^{ik(\eta_0-\eta_{rei}) \cos \theta} (1 + \cos^2 \theta) \cos 2\phi, \\ U_{(+)}(\mathbf{k}, \mathbf{n}) &= -3\tau_{rei} \int_{\eta_r}^{\eta_{rei}} d\eta h^{(+)\prime}(\mathbf{k}, \eta) \frac{j_2[(\eta_{rei}-\eta)k]}{k^2 (\eta_{rei}-\eta)^2} \\ &\quad \times e^{ik(\eta_0-\eta_{rei}) \cos \theta} \cos \theta \cdot \sin 2\phi, \end{aligned}$$

The analogs of the formulas (10.74), (10.75) have the forms

$$\begin{aligned} a_{lm,(+)}^E(\mathbf{k}) &= \frac{3}{2} \tau_{rei} (\delta_{m,+2} + \delta_{m,-2}) \cdot i^l \left\{ \frac{(l+2)(l+1)}{(2l-1)(2l+1)} j_{l-2}[(\eta_0 - \eta_{rei})k] \right. \\ &\quad - \frac{6(l+2)(l-1)}{(2l-1)(2l+3)} j_l[(\eta_0 - \eta_{rei})k] \\ &\quad \left. + \frac{l(l-1)}{(2l+1)(2l+3)} j_{l+2}[(\eta_0 - \eta_{rei})k] \right\} \\ &\quad \times \sqrt{2\pi(2l+1)} \cdot \int_{\eta_r}^{\eta_{rei}} d\eta h^{(+)\prime}(\mathbf{k}, \eta) \frac{j_2[(\eta_{rei}-\eta)k]}{k^2 (\eta_{rei}-\eta)^2}, \end{aligned} \quad (10.89)$$

$$\begin{aligned} a_{lm,(+)}^B(\mathbf{k}) &= 3\tau_{rei} \cdot i^l \left\{ \frac{l+2}{2l+1} j_{l-1}[(\eta_0 - \eta_{rei})k] - \frac{l-1}{2l+1} j_{l+1}[(\eta_0 - \eta_{rei})k] \right\} \\ &\quad \times \sqrt{2\pi(2l+1)} \cdot (\delta_{m,+2} - \delta_{m,-2}) \cdot \int_{\eta_r}^{\eta_{rei}} d\eta h^{(+)\prime}(\mathbf{k}, \eta) \frac{j_2[(\eta_{rei}-\eta)k]}{k^2 (\eta_{rei}-\eta)^2}. \end{aligned} \quad (10.90)$$

The coefficients for the two polarizations (+) and (\times) still obey (10.76). This gives for B -mode

$$\begin{aligned} C_l^{BB} &= 36\pi \cdot \tau_{rei}^2 \cdot \int_0^\infty \frac{dk}{k} \mathcal{P}_T(k) \cdot \left\{ \int_{\eta_r}^{\eta_{rei}} d\eta \mathfrak{h}'(k, \eta) \frac{j_2[(\eta_{rei}-\eta)k]}{(\eta_{rei}-\eta)^2 k^2} \right\}^2 \\ &\quad \times \left\{ \frac{l+2}{2l+1} j_{l-1}[(\eta_0 - \eta_{rei})k] - \frac{l-1}{2l+1} j_{l+1}[(\eta_0 - \eta_{rei})k] \right\}^2. \end{aligned} \quad (10.91)$$

In the lowest multipole region, $l \sim (\eta_0 - \eta_{rei})/\eta_{rei} \sim 2$, the integrals in (10.91) are saturated at $\eta \sim \eta_{rei}$, $k \sim \eta_{rei}^{-1}$. The functions $j_{l\pm 1}[(\eta_0 - \eta_{rei})k]$ are not suppressed. The dependence of C_l^{BB} on l is weak in this region, and \mathcal{D}_l^{BB} increases approximately as l^2 for the flat primordial spectrum,

$$\mathcal{D}_l^{BB} = T_0^2 \frac{l(l+1)}{2\pi} C_l^{BB} \sim \text{const} \cdot T_0^2 A_T \tau_{rei}^2 l^2. \quad (10.92)$$

This is consistent with the numerical result presented in Fig. 10.4.

The numerical coefficient in (10.92) is small. To illustrate this property, let us evaluate the inner integral in (10.91) at $k\eta_{rei} \ll 1$:

$$\int_{\eta_r}^{\eta_{rei}} d\eta \mathfrak{h}'(k, \eta) \frac{j_2[(\eta_{rei} - \eta)k]}{(\eta_{rei} - \eta)^2 k^2} \sim 3k^2 \int_{\eta_r}^{\eta_{rei}} \frac{\eta d\eta}{15^2} \simeq \frac{k^2 \eta_{rei}^2}{150}.$$

This indeed shows the numerical suppression in (10.91).

At higher multipoles, $l \gg (\eta_0 - \eta_{rei})/\eta_{rei}$, the situation is different. The relevant momenta are $k \gtrsim l/(\eta_0 - \eta_{rei}) \gg \eta_{rei}^{-1}$. In this case, the factor $j_2[(\eta_{rei} - \eta)k]/[(\eta_{rei} - \eta)^2 k^2]$ in the inner integral in (10.91) is strongly suppressed unless η is close to η_{rei} . Hence, the inner integral is saturated at $\eta = \eta_{rei} - O(k^{-1})$. At this time, $\mathfrak{h}'(k, \eta) \sim 1/(k\eta_{rei}^2)$, and the inner integral is of order $(k\eta_{rei})^{-2} \propto l^{-2}$. Since $j_l \propto l^{-1}$, we obtain that the multipoles C_l^{BB} decrease as l^{-6} in the range of l we consider, and \mathcal{D}_l^{BB} decrease as l^{-4} . This fall off is clearly seen in Fig. 10.4. The total spectrum C_l^{BB} becomes dominated by the recombination contribution already at fairly large angles, $l \sim 20$.

To estimate the spectrum at $l \gg (\eta_0 - \eta_{rei})/\eta_{rei}$ quantitatively, we make use of the formula (10.84) and the asymptotic behavior (F.10) of the spherical Bessel function j_2 at large values of its argument. Then we find the dominant contribution to the inner integral in (10.91), which comes from the time $\eta = \eta_{rei} - O(k^{-1})$, in the following form

$$\begin{aligned} \int_{\eta_r}^{\eta_{rei}} d\eta \mathfrak{h}'(k, \eta) \frac{j_2[(\eta_{rei} - \eta)k]}{(\eta_{rei} - \eta)^2 k^2} &\simeq 3 \int_0^{\eta_{rei} - \eta_r} d\tilde{\eta} \frac{j_2[(\eta_{rei} - \tilde{\eta})k] j_2(\tilde{\eta}k)}{\eta_{rei} - \tilde{\eta}} \frac{d\tilde{\eta}}{\tilde{\eta}^2 k^2} \\ &\simeq -3 \int_0^\infty dx \frac{\sin(\eta_{rei}k - x) j_2(x)}{k^2 \eta_{rei}^2} \frac{j_2(x)}{x^2}, \end{aligned}$$

where $\tilde{\eta} = \eta_{rei} - \eta$, $x = k\tilde{\eta}$, and we extended the integration interval, keeping in mind that the integral rapidly converges. This gives

$$\begin{aligned} &\int_{\eta_r}^{\eta_{rei}} d\eta \mathfrak{h}'(k, \eta) \frac{j_2[(\eta_{rei} - \eta)k]}{(\eta_{rei} - \eta)^2 k^2} \\ &\simeq -3 \frac{\sin(k\eta_{rei})}{k^2 \eta_{rei}^2} \int_0^\infty dx \cos x \frac{j_2(x)}{x^2} + 3 \frac{\cos(k\eta_{rei})}{k^2 \eta_{rei}^2} \int_0^\infty dx \sin x \frac{j_2(x)}{x^2} \\ &= \frac{1}{4} \frac{\cos(k\eta_{rei})}{k^2 \eta_{rei}^2}, \end{aligned}$$

where we used known integrals involving the spherical Bessel function j_2 . The remaining integral over momenta is evaluated by introducing the integration variable

$u \equiv (\eta_0 - \eta_{rei}) k / (l + \frac{1}{2})$, using Eq. (10.85), recalling the large order asymptotics (F.22) for j'_l and averaging the oscillating functions. The result is

$$C_l^{BB} \simeq \frac{9\pi}{16} \left(\frac{\eta_0 - \eta_{rei}}{\eta_{rei}} \right)^4 \frac{\tau_{rei}^2}{(l + \frac{1}{2})^6} \int_1^\infty \frac{du}{u} \cdot \mathcal{P}_T \left[\frac{(l + \frac{1}{2}) u}{\eta_0 - \eta_{rei}} \right] \cdot \frac{\sqrt{u^2 - 1}}{u^7}.$$

This gives for the flat primordial spectrum

$$\mathcal{D}_l^{BB} \simeq T_0^2 \cdot \frac{3\tau_{rei}^2}{140} \left(\frac{\eta_0 - \eta_{rei}}{\eta_{rei}} \right)^4 \cdot \frac{A_T}{(l + \frac{1}{2})^4}.$$

The spectrum indeed decreases, $\mathcal{D}_l^{BB} \propto l^{-4}$. To compare with Fig. 10.4, let us set $r = A_T/A_R = 0.38$ and $A_R \simeq 2.5 \cdot 10^{-9}$. We find for $l = 10$ and $\tau_{rei} = 0.08$

$$\mathcal{D}_{10}^{BB} \simeq 1.2 \cdot 10^{-3} \mu\text{K}^2.$$

This is rather close to the numerical result shown in Fig. 10.4, especially in view of the fact that Fig. 10.4 is obtained for earlier reionization corresponding to twice larger optical depth.

To end this Section, we give for completeness the expressions for the angular spectrum of E -mode generated by tensor perturbations. The contribution due to recombination epoch is

$$\begin{aligned} C_l^{EE} &= \frac{\pi}{25} \Delta \eta_r^2 \int_0^\infty \frac{dk}{k} \mathcal{P}_T(k) \mathfrak{h}'^2(k, \eta) \left\{ \frac{(l+2)(l+1)}{(2l-1)(2l+1)} j_{l-2}[(\eta_0 - \eta_r)k] \right. \\ &\quad \left. - \frac{6(l+2)(l-1)}{(2l-1)(2l+3)} j_l[(\eta_0 - \eta_r)k] \right\}^2 \\ &\quad + \frac{l(l-1)}{(2l+1)(2l+3)} j_{l+2}[(\eta_0 - \eta_r)k] \Bigg\}^2. \end{aligned} \quad (10.93)$$

$$\begin{aligned} C_l^{TE} &= \frac{\pi \Delta \eta_r}{10\sqrt{2}} \sqrt{\frac{(l+2)!}{(l-2)!}} \int_0^\infty \frac{dk}{k} \mathcal{P}_T(k) \mathfrak{h}'(k, \eta_r) \int_{\eta_r}^{\eta_0} d\eta \mathfrak{h}'(k, \eta) \frac{j_l[(\eta_0 - \eta)k]}{(\eta_0 - \eta)^2 k^2} \\ &\quad \times \left\{ \frac{(l+2)(l+1)}{(2l-1)(2l+1)} j_{l-2}[(\eta_0 - \eta_r)k] - \frac{6(l+2)(l-1)}{(2l-1)(2l+3)} j_l[(\eta_0 - \eta_r)k] \right. \\ &\quad \left. + \frac{l(l-1)}{(2l+1)(2l+3)} j_{l+2}[(\eta_0 - \eta_r)k] \right\}. \end{aligned} \quad (10.94)$$

The effect of reionization epoch has the following form

$$\begin{aligned} C_l^{EE} &= 9\pi \cdot \tau_{rei}^2 \cdot \int_0^\infty \frac{dk}{k} \mathcal{P}_T(k) \cdot \left\{ \int_{\eta_r}^{\eta_{rei}} d\eta \mathfrak{h}'(k, \eta) \frac{j_2[(\eta_{rei} - \eta)k]}{(\eta_{rei} - \eta)^2 k^2} \right\}^2 \\ &\quad \times \left\{ \frac{(l+2)(l+1)}{(2l-1)(2l+1)} j_{l-2}[(\eta_0 - \eta_{rei})k] - \frac{6(l+2)(l-1)}{(2l-1)(2l+3)} j_l[(\eta_0 - \eta_{rei})k] \right. \\ &\quad \left. + \frac{l(l-1)}{(2l+1)(2l+3)} j_{l+2}[(\eta_0 - \eta_{rei})k] \right\}^2. \end{aligned} \quad (10.95)$$

$$\begin{aligned}
C_l^{TE} = & \frac{3\pi \cdot \tau_{rei}}{2\sqrt{2}} \sqrt{\frac{(l+2)!}{(l-2)!}} \cdot \left\{ \int_0^\infty \frac{dk}{k} \mathcal{P}_T(k) \int_{\eta_r}^{\eta_0} d\eta \mathfrak{h}'(k, \eta) \frac{j_l[(\eta_0 - \eta)k]}{(\eta_0 - \eta)^2 k^2} \right\} \\
& \times \left\{ \int_{\eta_r}^{\eta_{rei}} d\eta \mathfrak{h}'(k, \eta) \frac{j_2[(\eta_{rei} - \eta)k]}{(\eta_{rei} - \eta)^2 k^2} \right\} \cdot \left\{ \frac{(l+2)(l+1)}{(2l-1)(2l+1)} j_{l-2}[(\eta_0 - \eta_{rei})k] \right. \\
& \left. - \frac{6(l+2)(l-1)}{(2l-1)(2l+3)} j_l[(\eta_0 - \eta_{rei})k] + \frac{l(l-1)}{(2l+1)(2l+3)} j_{l+2}[(\eta_0 - \eta_{rei})k] \right\}. \tag{10.96}
\end{aligned}$$

The analysis similar to that performed for *B*-mode shows, in particular, that the angular spectra C_l^{BB} and C_l^{EE} are very similar, cf. Fig. 10.4.

10.4 Discussion

The first evidence for CMB polarization has been obtained by the analysis of the spectra C_l^{EE} and C_l^{TE} obtained by DASI experiment (Degree Angular Scalar Interferometer) [95] at South Pole. Later on, other experiments also reported the detection of the *E*-mode, see Figs. 10.5 and 10.6. The existing data on *E*-mode are consistent with the generation mechanism (the Compton scattering of locally anisotropic radiation) studied in this Chapter: the spectra C_l^{EE} and C_l^{TE} are in accordance with the prediction obtained from the known temperature angular spectrum C_l^{TT} , see Fig. 10.5.

Measurements of C_l^{EE} and C_l^{TE} give additional cosmological information and are instrumental for removing the degeneracy in the cosmological parameters. In particular, the spectrum C_l^{EE} at $l \sim 100 - 1000$ is mostly determined by the velocity of the baryon-photon medium at the recombination epoch, while the temperature spectrum C_l^{TT} is most sensitive to the density perturbation and gravitational potential. At low multipoles, the spectra C_l^{EE} and C_l^{TE} strongly depend on the optical depth τ_{rei} at reionization, $C_l^{EE} \propto \tau_{rei}^2$, $C_l^{TE} \propto \tau_{rei}$. At the same time, the temperature spectrum has rather weak dependence on τ_{rei} , namely, $C_l^{TT} \propto e^{-\tau_{rei}}$. The current value [91] $\tau_{rei} = 0.084 \pm 0.016$ is determined mostly from the measurements of C_l^{EE} and C_l^{TE} . We note that the level of precision at which τ_{rei} is determined from the measurements of C_l^{EE} and C_l^{TE} is already close to the cosmic variance limit.

CMB polarization is particularly important from the viewpoint of the search for tensor perturbations. The analysis of temperature anisotropy alone has limited sensitivity to tensor perturbations, see, e.g., Ref. [97]. As an example, for the flat primordial tensor spectrum, the best projected sensitivity of measurements of C_l^{TT} is $r = A_T/A_R \simeq 0.07$. This limitation is due to both cosmic variance and degeneracy in parameters. The measurement of the polarization in *E*-mode helps improving the sensitivity to the projected value $r \simeq 0.02$. The analysis of *B*-mode enables one,

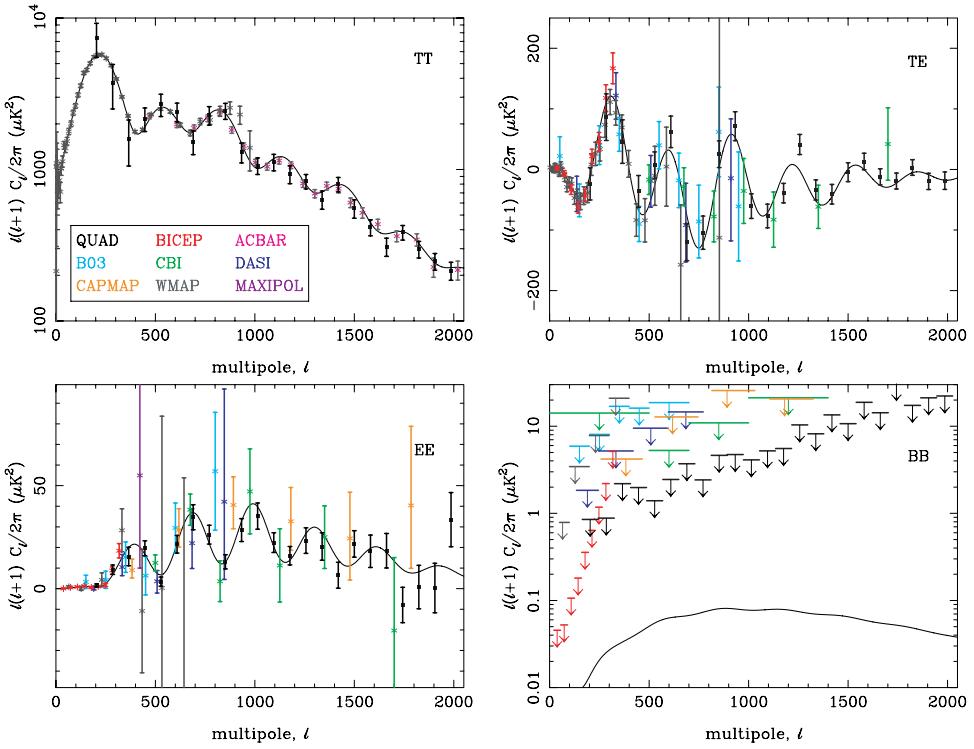


Fig. 10.6 Data on CMB temperature anisotropy and polarization [96], see Fig. 17.11 for color version. a) Temperature anisotropy spectrum \mathcal{D}_l^{TT} ; b) Cross correlation spectrum \mathcal{D}_l^{TE} ; c) Spectrum of E -mode \mathcal{D}_l^{EE} ; d) Bounds on the spectrum of B -mode \mathcal{D}_l^{BB} at 95% C.L.

in principle, to go much further. Particularly important is the study of B -mode at intermediate angular scales, where the cosmic variance is not very significant, while the secondary effect of gravitational lensing is not yet dominant.

No evidence for B -mode has been found so far, see Fig. 10.6. The current limit for low multipoles $l = 2 - 6$ is [91]

$$T_0^2 \cdot \frac{l(l+1)}{2\pi} C_l^{BB} < 0.15 \mu\text{K}^2$$

at 95% C.L. This yields the bound $r \lesssim 20$, which is much weaker than the bound quoted in this book. Presently, the strongest bound on r follows from the analysis of the spectra C_l^{TT} and C_l^{TE} . Future experiments designed specifically for the study of CMB polarization will dramatically change the situation. They will be capable of detecting the tensor modes at the level $r \sim 0.1$ and even somewhat smaller [97]. In the long run, the sensitivity of the CMB polarization measurements will reach 1% and even better for the ratio r .

It is worth noting (see Section 13.4 for details), that tensor perturbations of sizeable amplitude and almost flat primordial spectrum are predicted by various

inflationary models, including popular ones. Hence, the discovery of tensor perturbations with these properties would be a very strong argument in favor of inflation. This issue is quite hot, since the existing bounds on r are close to predictions of several inflationary models and even rule out some of them.

As we have noticed, B -mode is generated at small angular scales from E -mode due to gravitational lensing. Hence, the measurement of B -mode at large l would shed extra light on the abundance and evolution of structure in the recent Universe. This would be useful for the study of dark matter properties. Remarkably, the information on B -mode at small angular scales would be interesting from particle physics viewpoint. As an example, conventional neutrinos suppress the power spectrum of matter perturbations at small distance scales, see Section 8.4.3. This effect depends on neutrino masses, and if the latter are of the order of m_{atm} (see Appendix I.C), the measurement of CMB temperature anisotropy alone is in principle insufficient to detect the effect. Measuring B -mode at small angular scales is more promising, as it has potential sensitivity to the neutrino masses down to $\sum m_\nu \sim 3 \cdot 10^{-2}$ eV [98].

Finally, B -mode is sensitive to some extensions of the Standard Model of particle physics, in particular, to models with cosmic strings. Strings are sources of tensor (and also vector) perturbations. These may give considerable contribution to B -mode at scales below 1 angular degree, i.e., $l \gtrsim 1000$ [99]. The coefficients a_{lm}^B are expected to be strongly non-Gaussian in that case. Also, the study of B -mode would test some models of the generation of extragalactic magnetic fields, since magnetic fields in ionized gas yield the mixing between E - and B -modes due to the Faraday rotation.

To conclude, we emphasize that the existing data on CMB polarization do not contradict Λ CDM model with nearly flat primordial spectrum of adiabatic scalar perturbations in the absence of tensor perturbations and isocurvature scalar modes. The situation may change in near future.

Chapter 11

Drawbacks of the Hot Big Bang Theory. Inflation as Possible Way Out

The Universe we observe is homogeneous and isotropic to good approximation. It is large and warm. As discussed in Section 11.1, these properties, obvious at first sight, do not find satisfactory explanation within the Hot Big Bang theory. This is a serious drawback of this theory, or, rather, several drawbacks related to each other. A possible way out is the inflationary theory. Most notably, the inflationary theory also suggests a mechanism of generation of the cosmological perturbations. Such a mechanism is absent in the Hot Big Bang theory, which is its another serious drawback. This is why the inflationary theory is fairly compelling.

According to the inflationary theory, the Hot Big Bang epoch is preceded by the inflationary epoch of the fast (nearly exponential) cosmological expansion. At the latter epoch, matter equation of state is almost vacuum, $p \approx -\rho$, the energy density varies in time slowly, the Universe rapidly expands and becomes homogeneous, isotropic and spatially flat to excellent accuracy. Inflation ends up with the reheating epoch at which the vacuum-like energy turns into heat. All this occurred in a very short time interval; the corresponding estimates are given below.

Although the inflationary theory has not been unequivocally confirmed by observations, it is in overall agreement with the existing data. Some of its versions give specific, observationally testable predictions, among which tensor perturbations play a profound role. Remarkably, the study of the Universe at present cosmological scales may give us information on the properties of the Universe at the earliest epoch of its evolution that preceded the hot stage and was characterized by huge energy density and extremely small spatial and temporal scales. This is one of the most astonishing connections between micro- and macrophysics in Nature.

11.1 Drawbacks of the Hot Big Bang Theory

We have discussed in the accompanying book various unsolved problems of the Hot Big Bang theory, such as the nature of dark energy, origin of dark matter and

baryon asymmetry, etc. Here we concentrate on problems of a quite different sort. These have to do with the initial conditions which have to be imposed in the Hot Big Bang theory. Recall, in the first place, that this theory assumes that the early Universe was filled with hot plasma of high energy density ρ and positive pressure p , and the space-time metric had the form

$$ds^2 = dt^2 - a^2(t)dl^2, \quad (11.1)$$

where dl^2 is the metric of unit 3-sphere, 3-dimensional Euclidean space or unit 3-hyperboloid. Such a theory predicts the existence of singularity at the initial moment of time, when the scale factor $a(t)$ vanishes, while the energy density and pressure are infinite.

One would be tempted to argue that the existence of singularity is the consequence of the choice of homogeneous and isotropic metric (11.1), and that deviations from homogeneity and isotropy may enable one to get around the initial singularity. However, “theorems of singularity” proven in General Relativity, show that this is not the case, and that the initial singularity is a general property of expanding cosmological solutions.

The existence of singularity in a solution to the classical field equations is not specific to the FLRW metric (11.1) and theory of gravity in general. As an example, the Coulomb field of point-like charge in classical electrodynamics has a singularity at the position of the charge. One cannot use classical laws of physics near the singularity; physics is essentially quantum there. In the cosmological context this means that classical field equations are not applicable at the very beginning of the evolution. Classical equations are valid at sufficiently late times, $t \gtrsim t_{Pl}$, when the energy density, temperature and curvature are much smaller than the Planckian values.¹ The initial, quantum epoch of evolution would then set the initial data for subsequent classical expansion.

Quantum gravity is not developed to the extent that one can predict the most likely initial data for the Hot Big Bang evolution. So, some (and even all) of the problems we discuss below may be due to gross misunderstanding of putative quantum gravity regime. Nevertheless, one can envision with some degree of confidence the initial data that emerge after the quantum epoch.

A key is the observation that the only dimensionful parameter in General Relativity is the Planck mass M_{Pl} . Hence, one thinks on dimensional grounds that at time $t \sim t_{Pl}$ the size of a causally connected region in the Universe is of order l_{Pl} . So, one guesses that two regions of space separated by spatial distance $L > l_{Pl}$ at time $t \sim t_{Pl}$ have very different properties. One also guesses that the spatial curvature is comparable to the Planck value M_{Pl}^{-2} .

¹Instead of the Planck scale, one can think of some other scale at which gravity becomes quantum. Say, this can be the superstring scale. Although numerical estimates depend on this scale, the qualitative conclusions remain the same. We continue to talk about the Planck scale for definiteness. Likewise, we continue to count time from the would-be singularity.

Now, making use of the Einstein equations and assuming that the Hot Big Bang theory is valid starting from $t \sim t_{Pl}$ we can see whether the initial conditions needed for describing our Universe have any similarity with these guesses. Stated in this way, problems of the Hot Big Bang theory we discuss below emerge as gross discrepancy between the *a priori guessed* initial conditions and the initial data that *have to be introduced* into this theory in order that the outcome have any resemblance to our Universe. It is sufficient for our purposes to make crude, but simple estimates that highlight the heart of the matter.

11.1.1 Horizon problem

As the first exercise, let us calculate the spatial size of the visible Universe at the Planck epoch. We recall that the present size is about

$$l_{H,0}(t_0) \approx 1.4 \cdot 10^4 \text{ Mpc}.$$

In the Hot Big Bang theory, the size of this region was

$$l_{H,0}(t_{Pl}) = \frac{a(t_{Pl})}{a(t_0)} l_{H,0}(t_0) \sim \frac{T_0}{T_{Pl}} l_{H,0}(t_0), \quad (11.2)$$

where we neglected the factor involving the effective number of relativistic degrees of freedom. Numerically,

$$l_{H,0}(t_{Pl}) \sim 10^{30} M_{Pl}^{-1}. \quad (11.3)$$

Hence, the Hot Big Bang theory extrapolated back to nearly Planckian epoch says that the size of the visible Universe at that time was 30 orders of magnitude larger than the natural size of a causally connected region. In other words, the visible Universe consists of regions that were causally disconnected at the beginning of the classical evolution, whose number is

$$[l_{H,0}(t_{Pl}) M_{Pl}]^3 \sim 10^{90}.$$

It is natural to think that the relative density fluctuations from one such region to another were of order 1. As we discussed in previous Chapters, the primordial fluctuations *grow* at radiation and matter domination. Why then the Universe is homogeneous to excellent extent, say, at recombination? This question does not find its answer in the Hot Big Bang theory; it is the essence of the horizon problem.

This problem can be stated differently, without reference to the hypothetical Planck epoch. We defined the cosmological horizon $l_H(t)$ in Section I.3.2 as the maximum length a signal emitted at the Big Bang travels by the time t . This is the size of the causally connected region at that time. The horizon problem is due to the fact that the Universe expands slowly, so an observer sees more and more regions that have never been in causal contact with each other; nevertheless, the

Universe is homogeneous with high precision at very large scales. Quantitatively, the size of a causally connected region at time t_1 is

$$l_{H,1} = a(t_1) \int_{t_{Pl}}^{t_1} \frac{dt}{a(t)}. \quad (11.4)$$

Because of the cosmological expansion this size is stretched by now to

$$l_{H,1}(t_0) = a(t_0) \int_{t_{Pl}}^{t_1} \frac{dt}{a(t)}. \quad (11.5)$$

Within the Hot Big Bang theory, the expansion law is (until very recently)

$$a(t) \propto t^\alpha,$$

where $\alpha < 1$ ($\alpha = 1/2$ and $\alpha = 2/3$ at radiation and matter domination, respectively). Hence, the integral in (11.5) is saturated at the upper limit. We conclude that the size $l_{H,1}(t_0)$ is considerably smaller than the present horizon size $l_{H,0}$. To make an order-of-magnitude estimate, let us ignore the recent accelerated expansion and write the ratio of the two sizes as follows,

$$\frac{l_{H,0}}{l_{H,1}(t_0)} = \frac{\int_0^{t_0} dt/a(t)}{\int_0^{t_1} dt/a(t)} = \frac{\int_0^{a_0} da/[a^2 H(a)]}{\int_0^{a_1} da/[a^2 H(a)]} \simeq \frac{a(t_1)H(t_1)}{a(t_0)H(t_0)}. \quad (11.6)$$

At matter domination we have $a \propto t^{2/3}$, $H \propto t^{-1}$, so that $aH \propto t^{-1/3} \propto a^{-1/2}$, and

$$\frac{l_{H,0}}{l_{H,1}(t_0)} \simeq \sqrt{1 + z(t_1)}.$$

As an example, let t_1 be the time of recombination t_r . We recall that $z_r = 1100$ and find that the angle $\Delta\theta_r$ at which the horizon at recombination is seen today equals roughly²

$$\frac{1}{\Delta\theta_r} = \frac{l_{H,0}}{l_{H,r}(t_0)} \sim 35.$$

Hence, the sphere of last scattering observed via CMB consists of

$$\left(\frac{l_{H,0}}{l_{H,r}(t_0)} \right)^2 \sim 1000$$

parts which have never been in causal contact before CMB decoupling. This is illustrated in Fig. 11.1. Within the Hot Big Bang theory, there is no reason to think that photons coming from these causally disconnected regions have comparable, let alone the same, temperature. Still, CMB is isotropic to accuracy better than 10^{-4} !

This formulation of the horizon problem is practically insensitive to physics at the Planck scale. It is sufficient that the cosmological expansion is assumed to be dominated by relativistic and/or non-relativistic matter from the very beginning of the classical evolution.

²The accurate estimate is $(\Delta\theta_r)^{-1} \simeq 50$, see Section 2.1.2.

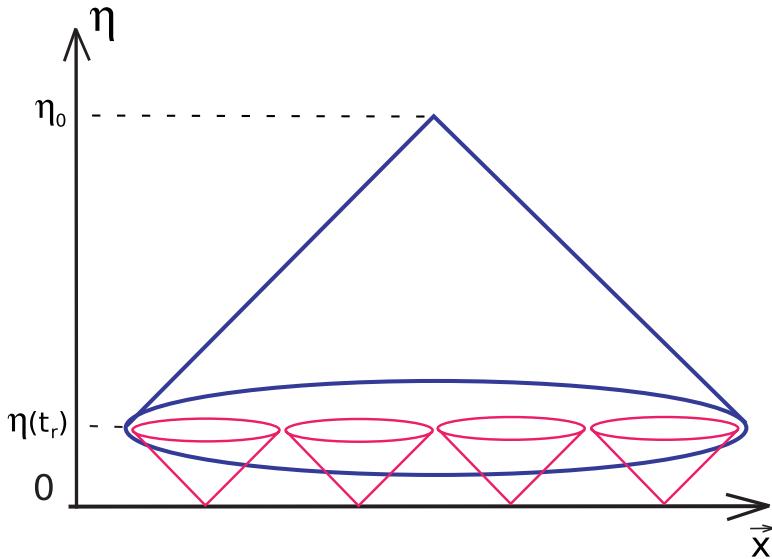


Fig. 11.1 In conformal coordinates, light cones look the same as in Minkowski space-time. In the Hot Big Bang theory, the coordinate size of a region, causally connected by the time t_r , equals $x_H(t_r) = \eta(t_r)$, where $\eta(t_r) \equiv \eta_r = \int_{t_{Pl}}^{t_r} \frac{dt}{a(t)}$. The coordinate size of the sphere of last scattering seen today is $(\eta_0 - \eta_r)$. The number of regions which are causally disconnected at time t_r and visible today is estimated as $(\eta_0/\eta_r)^3$. It is large, since $\eta_0 \gg \eta_r$. This estimate, of course, coincides with the one made in the text.

11.1.2 Flatness problem

Let us see how close to the reality is our second guess that the spatial curvature of the Universe was roughly of the order of the Planck scale at the Planck epoch, i.e., that the spatial curvature term in the right hand side of the Friedmann equation was comparable to other terms. Recall that the curvature term is equal to a^{-2} in absolute value, while all terms together give H^2 . Hence, our guess is that³

$$|\Omega_{curv}(t_{Pl})| \equiv \frac{|\rho_{curv}(t_{Pl})|}{\rho_c(t_{Pl})} = \frac{1}{a^2(t_{Pl})H^2(t_{Pl})} \sim 1. \quad (11.7)$$

As we discussed in Section I.4, the contribution of the spatial curvature to the Friedmann equation at the present epoch is small,

$$|\Omega_{curv,0}| < 0.02.$$

It changes in time as follows,

$$\Omega_{curv}(t) \propto \frac{1}{a^2(t)H^2(t)}.$$

³In this Section, and only in this Section we deviate from our convention that the parameter Ω_{curv} refers to the present epoch only.

We have already noticed when discussing the horizon problem that the ratio (11.6) decreases in time in the Hot Big Bang theory. Hence, the ratio

$$\frac{\Omega_{curv}(t)}{\Omega_{curv,0}} = \frac{(a_0 H_0)^2}{[a(t)H(t)]^2} \quad (11.8)$$

is very small at early times. At the Planck epoch

$$\frac{\Omega_{curv}(t)}{\Omega_{curv,0}} = \left(\frac{a_0}{a(t_{Pl})} \right)^2 \left(\frac{H_0}{H(t_{Pl})} \right)^2 \sim \left(\frac{M_{Pl}}{T_0} \right)^2 \left(\frac{H_0}{M_{Pl}} \right)^2 = \frac{H_0^2}{T_0^2} \sim 10^{-58}. \quad (11.9)$$

We see that the Hot Big Bang theory indeed reproduces the observed Universe only if the following initial condition is imposed,

$$|\Omega_{curv}(t_{Pl})| \lesssim 10^{-60}. \quad (11.10)$$

The spatial curvature had to be 60 orders of magnitude (!) smaller than the *a priori* estimate. This huge discrepancy is called the flatness problem.

Similar to the horizon problem, the flatness problem does not necessarily appeal to the Planck epoch. Say, for the Big Bang Nucleosynthesis epoch that begins at $T_{BBN} \sim 1 \text{ MeV}$, $t_{BBN} \sim 1 \text{ s}$, the estimate is

$$\begin{aligned} \Omega_{curv}(t_{BBN}) &\sim \left[\frac{a_0 H_0}{a(t_{BBN})H(t_{BBN})} \right]^2 \Omega_{curv,0} \\ &\sim \left(\frac{T_{BBN}}{T_0} \right)^2 \left(\frac{t_{BBN}}{t_0} \right)^2 \Omega_{curv,0} \lesssim 10^{-18}. \end{aligned}$$

Hence, the spatial curvature was definitely very small in the early Universe; in the context of Hot Big Bang theory this initial condition must be put in “by hand”.

It is worth noting that the parameter Ω_{curv} can be viewed as the ratio of the 3-dimensional to 4-dimensional curvature of the Universe (for $\Omega_{curv} \ll 1$, the 4-dimensional curvature is of order H^2). From this viewpoint, the result (11.10) implies huge mismatch between the two geometrical quantities, which is unexplained by the Hot Big Bang theory.

11.1.3 Entropy problem

The entropy problem also has to do with unnatural initial data in the Hot Big Bang theory. We have calculated the total entropy in the observable part of the Universe in Section I.5.2 and obtained the huge value,

$$S_0 \sim 10^{88}.$$

The cosmological expansion is reasonably adiabatic in the Hot Big Bang theory, so this gigantic dimensionless parameter has to be put in as the initial condition. This again contradicts *a priori* expectation based on the dimensional argument. As an example, for a closed Universe that can be characterized by the total entropy, the initial entropy must be greater than 10^{88} , while one would think that the natural

initial value is of the order of the number of particle species (i.e., of order 100 in the Standard Model of particle physics).

Clearly, the entropy problem is related to the horizon problem. We have seen that the size of the visible Universe at the Planck epoch was 30 orders of magnitude larger than the Planck length. Accordingly, its entropy exceeds the natural value by $3 \cdot 30$ orders of magnitude.

Hence, the Hot Big Bang theory does not explain why our Universe is so homogeneous, isotropic, “large” (spatially flat) and warm. Instead, these fundamental properties appear as consequences of very contrived initial conditions.

11.1.4 Primordial perturbation problem

Our discussion shows that there must exist a mechanism responsible for the homogeneity and flatness of the early Universe in the very beginning of the Hot Big Bang epoch at huge length scales. Still, the Universe is not absolutely homogeneous: there are structures like galaxies, clusters of galaxies, voids. As we discussed in previous Chapters, their presence, as well as the observed CMB anisotropy, requires the existence of primordial density perturbations of the relative amplitude at the level $\delta\rho/\rho \sim 5 \cdot 10^{-5}$, whose power spectrum is close to the flat, Harrison–Zeldovich spectrum. The Hot Big Bang theory does not provide the mechanism of the generation of the primordial perturbations; they must be put in “by hand” as yet another initial condition. This is the primordial perturbation problem, which is nicely solved by the inflationary theory.

11.2 Inflation: The Basic Idea

The important property behind the horizon and flatness problems is the fact that the combination $a(t)H(t)$ is the *decreasing function of time* in the hot expanding Universe. The general idea of inflation as a solution to the problems of the Hot Big Bang theory is that the hot stage was preceded by the inflationary epoch when $a(t)H(t)$ *increased with time*, i.e., the growth rate of the scale factor exceeded that of the cosmological horizon size. The combination $a(t)H(t)$ grows in time if the scale factor $a(t)$ increases faster than t , i.e., if the Universe undergoes the accelerated expansion. This is the origin of the term “inflation”. Technically, the effect of the accelerated expansion is that the integral in (11.4) is saturated at the lower, rather than upper limit of integration, and the horizon size has large value by the time t_{reh} when the Universe enters the hot stage.

Historically, the first well-developed inflationary model was proposed by Starobinsky [100, 101]. It makes use of the gravity theory whose Lagrangian contains both first and higher order terms in the Riemann curvature scalar. The Friedmann equation in this theory is modified at large values of the Hubble parameter, and the scale factor exponentially increases during sufficiently long period of time. This

model also has the mechanism of reheating, i.e., the transition from the inflationary to the hot stage. Hence, the Starobinsky model is successful. In this book, however, we consider models in which gravity is described by General Relativity, while the inflationary regime occurs due to fairly unconventional matter. In fact, the Starobinsky model also belongs to this category, since we know from Section I.A.5 that gravity theories with Lagrangians non-linear in curvature scalar are equivalent to General Relativity with a scalar field. We briefly consider the Starobinsky model at the end of Section 12.2.

We know from Section I.3.2 that within General Relativity, the accelerated cosmological expansion occurs when matter has negative pressure,

$$p < -\frac{1}{3}\rho. \quad (11.11)$$

Problem 11.1. *Strictly speaking, this result has been obtained in Section I.3.2 only for the barotropic and linear equation of state $p = w\rho$. Show that the inequality (11.11) must hold for the Universe that undergoes the accelerated expansion irrespective of the form of the equation of state.*

Concrete models in which the Universe is filled with matter with this fairly exotic equation of state are discussed in Section 12.2. Here we simply assume that inflation does occur in a time interval from $t \sim t_{Pl}$ to a certain time t_e (the notation here refers to the inflation end), after which the Universe reheats and the hot stage begins at temperature

$$T_{reh} \sim \sqrt{M_{Pl}^* H(t_e)}. \quad (11.12)$$

We assume here that the reheating is instantaneous, so the hot stage starts right after the end of inflation. This is a simplified picture, which is, however, sufficient for crude estimates. In this picture, the Hubble parameter at temperature T_{reh} is equal to the Hubble parameter at the inflation end, and Eq. (11.12) is merely the Friedmann equation right after the reheating. In order that the relation (11.7) be satisfied, i.e., the flatness problem be solved, we need (cf. (11.8))

$$1 \gtrsim \frac{[a(t_{Pl})H(t_{Pl})]^2}{(a_0 H_0)^2} = \frac{[a(t_{Pl})H(t_{Pl})]^2}{[a(t_e)H(t_e)]^2} \frac{[a(t_e)H(t_e)]^2}{(a_0 H_0)^2}. \quad (11.13)$$

In other words, the condition for successful inflation is

$$\frac{a(t_e)H(t_e)}{a_{Pl}H_{Pl}} \gtrsim \frac{a(t_e)H(t_e)}{a_0 H_0} \sim \frac{T_0}{T_{reh}} \frac{H(t_e)}{H_0}. \quad (11.14)$$

Let us see that the same inequality is sufficient for solving the horizon problem. To this end, we calculate the present size of a region which is causally connected by the time t_e . We use (11.5) with $t_1 = t_e$ and obtain

$$l_{H,e}(t_0) = a_0 \int_{t_{Pl}}^{t_e} \frac{dt}{a(t)} = a_0 \int_{t_{Pl}}^{t_e} da \frac{1}{Ha^2} \simeq \frac{a_0}{a(t_{Pl})H(t_{Pl})}.$$

We accounted here for the fact that in the inflationary regime, the integral is saturated at the lower limit of integration. The ratio of this size to the present Hubble length is of order

$$\frac{l_{H,e}(t_0)}{l_{H,0}} \sim \frac{a_0}{a(t_{Pl})} \frac{H_0}{H(t_{Pl})}.$$

Once the condition (11.13) is satisfied, we have

$$\frac{l_{H,e}(t_0)}{l_{H,0}} \gtrsim 1.$$

Thus, the observed Universe was causally connected by the end of inflation, and the horizon problem (in its second version) is solved, see Fig. 11.2.

Turning to the first version of the horizon problem, we note that the inflationary epoch is not at all thermal, so the last estimate in Eq. (11.2) is not valid. Instead, we have

$$l_{H,0}(t_{Pl}) = \frac{a(t_{Pl})H(t_{Pl})}{a_0 H_0} \frac{H_0}{H(t_{Pl})} l_{H,0} \lesssim \frac{H_0}{H(t_{Pl})} l_{H,0} \lesssim M_{Pl}^{-1},$$

where we used the inequality (11.13) and set $H(t_{Pl}) \sim M_{Pl}$. We see that inflation solves the horizon problem in the first version as well: the entire visible Universe started up as part of a Planck-size region.

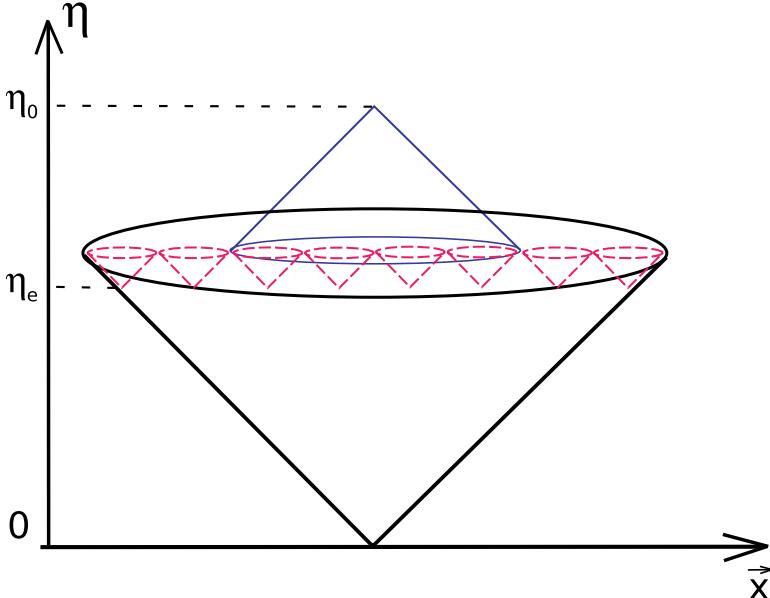


Fig. 11.2 The duration of the history of the Universe in conformal time, and hence the coordinate horizon size, is much greater in the inflationary theory than in the Hot Big Bang theory. Inflation takes the origin of conformal time back into the past, and in this way enlarges the size of a causally connected region.

To estimate roughly the minimal duration of successful inflation, we write the condition (11.14) as follows,

$$N_e^{(tot)} > \log \left(\frac{T_0 H(t_{Pl})}{T_{reh} H_0} \right) \simeq \log \left(\frac{T_0}{H_0} \right) + \log \frac{M_{Pl}}{T_{reh}}, \quad (11.15)$$

where

$$N_e^{(tot)} = \log \frac{a(t_e)}{a(t_{Pl})}$$

is the total *number of e-foldings* during inflation. The terminology here reflects the fact that the Universe expands by a factor of $\exp[N_e^{(tot)}]$. $N_e^{(tot)}$ is estimated as

$$N_e^{(tot)} = \int_{t_{Pl}}^{t_e} H(t) dt \sim H_{infl} \cdot \Delta t_{infl}, \quad (11.16)$$

where H_{infl} is the typical Hubble parameter at inflation and Δt_{infl} is the total duration of the inflationary epoch. We have already estimated the argument of the first logarithm in (11.15), see (11.9), so the condition (11.15) reads,

$$N_e^{(tot)} > N_e^{(min)} \simeq 67 + \log \frac{M_{Pl}}{T_{reh}}. \quad (11.17)$$

Hence, for reheat temperatures T_{reh} in the range from M_{Pl} to 1 TeV we have

$$N_e^{(min)} \simeq 70 - 100. \quad (11.18)$$

We see that the horizon and flatness problems are solved if inflation lasts for only 70–100 Hubble times of that epoch. This is fairly short time: from the rough estimate

$$H_{infl} \sim H(t_e) \sim \frac{T_{reh}^2}{M_{Pl}}$$

we find the minimum duration

$$\Delta t_{infl}^{(min)} \sim 10^{-42} - 10^{-9} \text{ s}$$

for $T_{reh} = M_{Pl} - 1$ TeV. In such a short time, the Universe expands to the extent that initially Planck-size regions become cosmologically large.

A few comments are in order. First, when obtaining the estimate (11.18) we assumed that the reheating is instantaneous. In realistic models, reheating may take many Hubble times. Hence, our estimate of $N_e^{(min)}$ may be somewhat too high. The traditionally quoted value is

$$N_e^{(min)} \simeq 60.$$

So, one traditionally talks about the minimum of 60 e-foldings, though this number is actually model-dependent.

Second, the right hand side of (11.15) is the estimate for the *minimum* number of e-foldings needed for solving the horizon and flatness problems. Generally speaking, there is no reason to think that the actual number of e-foldings coincides with this

minimum number; on the contrary, such a coincidence would require fine tuning of parameters. Without fine tuning, one has

$$N_e^{(tot)} \gg N_e^{(min)}.$$

This gives the important prediction that the spatial curvature of the present Universe is very small,⁴

$$\Omega_{curv} \ll 0.02.$$

A discovery of non-vanishing spatial curvature would be a serious argument *against* the inflationary theory, at least in its simple versions.

Third, we arbitrarily assumed in our discussion that inflation begins at the Planck time t_{Pl} . This is not at all necessary; the above estimates do not change much if inflation begins at some moment of time $t_i \gg t_{Pl}$ when the Hubble parameter is considerably smaller than M_{Pl} .

Fourth, we have not touched upon the entropy problem yet. It is clear, nevertheless, that inflation can solve this problem too. Realistic models are equipped with a mechanism of the post-inflationary reheating which ensures that the Universe does enter the hot stage. Thus, there is automatically enough entropy after the reheating. At the same time, it is not at all necessary that the Universe *before* inflation has large entropy. Entropy is generated at the post-inflationary reheating epoch, when the energy of vacuum-like matter transforms into heat. In other words, post-inflationary reheating is a strongly inequilibrium process from the thermodynamical viewpoint, and it efficiently produces entropy in the Universe.

Finally, we will see in Section 13 that inflation solves the primordial perturbation problem as well. Here we also mention that it helps solving other potential problems emerging in particle physics models. As an example, the hypothesis of Grand Unification yields the prediction of the existence of magnetic monopoles whose mass is roughly of order 10^{17} GeV. These particles would be produced at the Grand Unified phase transition at temperature of order $M_{GUT} \sim 10^{16}$ GeV. We discussed magnetic monopoles in Section I.12.2 where we have found that their present mass density would exceed the critical density by many orders of magnitude. This problem is solved in the inflationary scenario if the reheating temperature T_{reh} is smaller than M_{GUT} (which is quite realistic): monopoles are not produced at the hot stage in that case. Furthermore, if there were monopoles before inflation, the expansion of the Universe at inflation dilutes their number density to an unobservable value. Historically, the monopole problem was one of the strong motivations for the inflationary scenario.

⁴We restore the convention that the parameter Ω_{curv} refers to the *present epoch*.

This page is intentionally left blank

Chapter 12

Inflation in Slow Roll Regime

12.1 Slow Roll Conditions

We now turn to the discussion of how the physical conditions necessary for successful inflation can be implemented. In the first place, we would like to understand what sort of matter may have negative pressure. We have studied one example in Section I.3.2.3. Namely, we have seen that negative pressure is characteristic of *vacuum*, whose energy-momentum tensor has the following form,

$$T_{\mu\nu} = \rho_{vac} g_{\mu\nu},$$

where ρ_{vac} is the vacuum energy density, constant in space and time. Hence, the vacuum equation of state is

$$p = -\rho.$$

With this equation of state, and positive ρ_{vac} , the scale factor evolves exponentially,

$$a(t) = \text{const} \cdot e^{H_{vac} t}, \quad (12.1)$$

where the Hubble parameter H_{vac} is independent of time and is given by

$$H_{vac} = \sqrt{\frac{8\pi}{3} \frac{\rho_{vac}}{M_{Pl}^2}}.$$

If there is normal matter in the Universe, besides vacuum of positive energy density, the late-time cosmological expansion still obeys the exponential law (12.1), since the vacuum energy density stays constant, while the energy density of matter decreases as the Universe expands. Thus, positive vacuum energy density leads naturally to the inflationary expansion. However, we need a mechanism that terminates the inflation and generates hot matter in the Universe. So instead of vacuum, one thinks of a homogeneous field whose energy density is almost constant during inflation. At some point this field decays producing conventional particles, and inflation ends. The simplest possibility is that this field is a scalar.

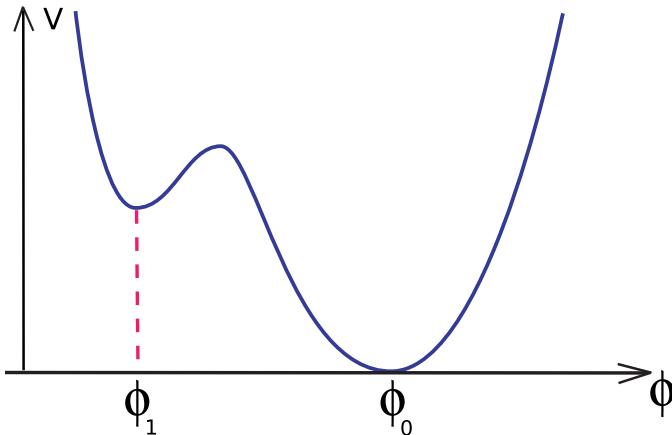


Fig. 12.1 The scalar potential in the “old inflation” model.

Historically, the first inflationary model that explicitly used a scalar field was proposed by Guth [102] and is called “old inflation”. It is based on the scalar field with the scalar potential shown in Fig. 12.1. This potential has two minima. One of them ($\phi = \phi_0$) has zero energy and corresponds to the true vacuum where we stay now.¹ Another minimum corresponds to the metastable vacuum with positive energy density

$$\rho_{vac} = V(\phi_1).$$

Old inflation is based on the assumption² that the Universe is initially in the wrong vacuum with $\phi = \phi_1$. The cosmological expansion dilutes the energy density of particles, if any, and the Universe enters the inflationary regime characterized by the energy density ρ_{vac} . The false vacuum is unstable; it decays via the quantum process of nucleation of bubbles of the true vacuum that expand at the speed of light — the Universe boils. The energy of the false vacuum transforms into the kinetic energy of bubble walls. It was assumed that the reheating occurs due to bubble wall collisions which create particles, and in the end the kinetic energy of bubble walls turns into heat.

However, there is a caveat in this scenario. Sufficiently long inflation requires that the bubble nucleation probability per unit time per unit volume Γ is much smaller than the parameter H_{vac}^4 determining the cosmological expansion rate.

¹Hereafter we neglect the present dark energy as it is negligibly small by the inflationary standards.

²This assumption holds, e.g., if inflation was preceded by yet another hot stage, and the field ϕ is of the Higgs type. Then the effective potential for ϕ has unique minimum at high temperature, $\phi = \phi_1 = 0$, see Section I.10.2, and the expectation value of the field ϕ vanishes initially. As the Universe cools down, the second minimum of the effective potential develops, but the phase transition may be delayed, and the inflationary regime may set in.

In this situation, the motion of bubbles from each other caused by the exponential expansion prevents the bubble walls to collide: the distance between the walls of neighboring bubbles actually increases. In other words, since the bubble nucleation rate per Hubble time per Hubble volume is smaller than 1, bubbles nucleate outside the event horizon of each other, and they never get in causal contact (while the wall collision *is* the causal contact). Bubble walls do not collide, and the Universe does not get hot. The Guth model with the reheating due to bubble wall collisions does not work.

Problem 12.1. *Let old inflation begin at time t_i . Let there be comoving observers in the Universe, homogeneously distributed over space. Show that the relative number of observers staying in the false vacuum at time $t > t_i$ is³*

$$P(t) = \exp \left[-\frac{4\Gamma\pi}{3} \int_{t_i}^t dt' a^3(t') \left(\int_{t'}^t \frac{dt''}{a(t'')} \right)^3 \right]. \quad (12.2)$$

When deriving this formula, assume that the initial bubble size is negligibly small. Using (12.2), show that successful inflation, such that the scale factor increases at least e^{60} times for most of observers, requires

$$\Gamma \lesssim 0.004H_{vac}^4.$$

There are, however, even simpler models yielding both the inflationary expansion and post-inflationary reheating. These are based on the slow roll mechanism. To introduce this mechanism, let us consider classical equations describing scalar field in the expanding Universe. The action for the scalar field minimally coupled to gravity is

$$S = \int d^4x \sqrt{-g} \mathcal{L} = \int d^4x \sqrt{-g} \left[\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) \right], \quad (12.3)$$

and the energy-momentum tensor is given by

$$T_{\mu\nu} = \frac{2}{\sqrt{-g}} \frac{\delta S}{\delta g^{\mu\nu}} = \partial_\mu \phi \partial_\nu \phi - g_{\mu\nu} \mathcal{L}. \quad (12.4)$$

Consider *homogeneous* scalar field in the spatially flat Universe with the standard FLRW metric (11.1). The action (12.3) yields the following equation of motion,

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0. \quad (12.5)$$

The energy-momentum tensor coincides formally with that of ideal fluid, whose energy density and pressure are

$$\rho = \frac{1}{2} \dot{\phi}^2 + V(\phi), \quad (12.6)$$

$$p = \frac{1}{2} \dot{\phi}^2 - V(\phi). \quad (12.7)$$

³We prefer not to talk in terms of probability, since its definition is a delicate matter in this context.

In the absence of other matter, the second equation determining the evolution of the entire system is the Friedmann equation

$$H^2 = \frac{8\pi}{3M_{Pl}^2} \left(\frac{1}{2}\dot{\phi}^2 + V(\phi) \right). \quad (12.8)$$

Equation (12.5) has simple analogy: it is equivalent to the equation governing mechanical roll down in a potential well $V(\phi)$ with time-dependent friction coefficient $3H$. The slow roll inflationary regime occurs when the Hubble friction term $3H\dot{\phi}$ dominates over the acceleration term in Eq. (12.5), i.e.,

$$\left| \frac{\ddot{\phi}}{3H\dot{\phi}} \right| \ll 1. \quad (12.9)$$

Under this condition, the field ϕ slowly rolls down the potential. Another slow roll condition ensures that the kinetic energy is small compared to potential energy in Eq. (12.8),

$$\frac{\dot{\phi}^2}{2V(\phi)} \ll 1. \quad (12.10)$$

This condition yields the approximate relation

$$p \approx -\rho,$$

which shows that the scalar field indeed behaves as vacuum-like matter and gives rise to almost exponential expansion of the Universe.

Once the conditions (12.9) and (12.10) are satisfied, the system of equations (12.5), (12.8) takes the following simple form,

$$\dot{\phi} = -\frac{1}{3H}V'(\phi), \quad (12.11)$$

$$H = \frac{1}{M_{Pl}} \left(\frac{8\pi V}{3} \right)^{1/2}. \quad (12.12)$$

Equation (12.12) implies that the dependence of the scale factor $a(t)$ on time in the slow roll regime is given by

$$a(t) = a_i \exp \left(\int_{t_i}^t H(t') dt' \right) = a_i \exp \left\{ \left(\frac{8\pi}{3M_{Pl}^2} \right)^{1/2} \int_{t_i}^t \sqrt{V[\phi(t)]} dt \right\}, \quad (12.13)$$

where the subscript i refers to the beginning of inflation. To see that the expansion is nearly exponential, let us calculate the change of the Hubble parameter in the Hubble time, \dot{H}/H^2 . By differentiating Eq. (12.12) we find

$$\dot{H} = \frac{1}{2M_{Pl}} \left(\frac{8\pi}{3V} \right)^{1/2} V'(\phi)\dot{\phi}. \quad (12.14)$$

We then take the ratio of Eqs. (12.14) and (12.12), express $V'(\phi)$ using (12.11) and obtain

$$\frac{\dot{H}}{H} = \frac{1}{2} \frac{V'}{V} \dot{\phi} = -\frac{3}{2} \frac{\dot{\phi}^2}{V} H.$$

Hence, the inequality (12.10) gives

$$\left| \frac{\dot{H}}{H^2} \right| = 3 \frac{\dot{\phi}^2}{2V} \ll 1. \quad (12.15)$$

This shows that the expansion is indeed almost exponential: the Hubble parameter does not change considerably in the Hubble time.

The same result can be stated in a somewhat different way. According to the basic definition, inflation means the growth in time of the combination (aH) . This is indeed the case, since

$$\frac{1}{a} \frac{d}{dt} (aH) = H^2 + \dot{H},$$

and the right hand side here is positive in view of (12.15).

Let us now find the conditions on the scalar potential $V(\phi)$ which guarantee that the inequalities (12.9) and (12.10) are satisfied. We insert the Hubble parameter found from Eq. (12.12) into Eq. (12.11) and obtain the following expression for the speed of rolling,

$$\dot{\phi} = -\frac{M_{Pl}}{(24\pi)^{1/2}} \frac{V'}{V^{1/2}}. \quad (12.16)$$

The condition (12.10) is now written as

$$\frac{M_{Pl}^2}{48\pi} \left(\frac{V'}{V} \right)^2 \ll 1. \quad (12.17)$$

Turning to the condition (12.9), we evaluate the time derivative of Eq. (12.16) and obtain

$$\ddot{\phi} = -\frac{M_{Pl}}{(24\pi)^{1/2}} \left(\frac{V''}{V^{1/2}} - \frac{1}{2} \frac{V'^2}{V^{3/2}} \right) \dot{\phi} = -\frac{M_{Pl}^2}{8\pi} \left(\frac{V''}{V} - \frac{1}{2} \left(\frac{V'}{V} \right)^2 \right) H \dot{\phi}, \quad (12.18)$$

where we used Eq. (12.12) to arrive at the second equality. We recall (12.17) and find that the condition (12.9) is satisfied provided that

$$\left| \frac{V''}{V} \right| \ll \frac{24\pi}{M_{Pl}^2}. \quad (12.19)$$

Hence, the slow roll regime occurs if and only if the scalar potential obeys (12.17) and (12.19). The inequalities (12.17) and (12.19) are dubbed the slow roll conditions.

Traditionally, one introduces the slow roll parameters

$$\epsilon = \frac{M_{Pl}^2}{16\pi} \left(\frac{V'}{V} \right)^2 \quad (12.20)$$

$$\eta = \frac{M_{Pl}^2}{8\pi} \frac{V''}{V}. \quad (12.21)$$

The values of these parameters in concrete inflationary models determine the predicted properties of scalar and tensor perturbations, see Chapter 13. For this reason, we discuss in Section 12.2 the dependence of ϵ and η on the shape of the scalar potential. In terms of these parameters, the slow roll conditions are simply

$$\epsilon \ll 1, \quad \eta \ll 1. \quad (12.22)$$

In most inflationary models inflation ends when at least one of these parameters becomes close to 1. The time derivatives considered above are expressed in terms of the slow roll parameters as follows,

$$\frac{\dot{\phi}^2}{2V} = \frac{\epsilon}{3}, \quad \frac{\dot{H}}{H^2} = -\epsilon, \quad \frac{\ddot{\phi}}{H\dot{\phi}} = \epsilon - \eta.$$

We note that the parameters ϵ and η depend on time at inflation, since the field ϕ varies in time, albeit slowly.

Hence, once the conditions (12.22) are satisfied, the scalar field slowly rolls down the potential $V(\phi)$, its value ϕ and velocity $\dot{\phi}$, as well as the Hubble parameter H change only slightly in the Hubble time, the equation of state is nearly vacuum and the cosmological expansion is nearly exponential. This is the *slow roll inflation regime*, and the field ϕ is called *inflaton*.

Problem 12.2. *Find the equation of state parameter w and its time derivative to the leading non-trivial order in ϵ and η .*

12.2 Inflationary Models

We consider in this Section several examples of scalar potentials in which the inflationary slow roll regime occurs for long enough time, then inflation ends and the Universe enters the hot stage. We emphasize that these examples by no means exhaust all possibilities. Furthermore, there are alternative mechanisms of inflation, such as k -inflation [103] that exploits the unconventional kinetic term in the inflaton Lagrangian. The examples we study here are, nevertheless, particularly instructive, since they illustrate possible variants of the inflaton and scale factor evolution in the slow roll regime. We use the traditional nomenclature that sometimes reflects physics and sometimes is of historical origin (like “old inflation” briefly discussed in the beginning of Section 12.1).

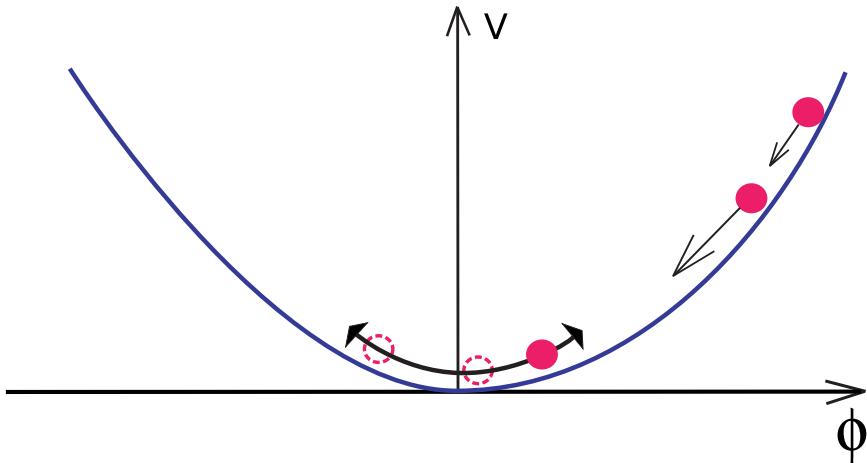


Fig. 12.2 The inflaton potential in the large field model. Circles show the inflaton field at various stages of its evolution.

12.2.1 Large field inflation (“chaotic”)

The slow roll conditions (12.17), (12.19) require that the inflaton potential $V(\phi)$ slowly varies as function of ϕ . At first sight, they strongly restrict the shape of the inflaton potential. To see that this is not the case, consider the simple power-law potential shown in Fig. 12.2, namely

$$V = g\phi^n, \quad n > 0. \quad (12.23)$$

The coupling g has dimension

$$[g] = (\text{mass})^{4-n}.$$

The slow roll conditions are valid in the theory with the potential (12.23) at large field values,

$$\phi \gg \frac{nM_{Pl}}{4\sqrt{3\pi}}. \quad (12.24)$$

Hence, the condition for slow roll is that the field takes super-Planckian value. This does not mean, however, that the dynamics cannot be described within the classical theory of the scalar field and classical General Relativity. Indeed, quantum gravity effects are small provided that the energy density is sub-Planckian,

$$V(\phi) \ll M_{Pl}^4. \quad (12.25)$$

This condition is satisfied for the field obeying

$$\phi \ll g^{-1/n} M_{Pl}^{4/n}. \quad (12.26)$$

We compare the conditions (12.24) and (12.26) and see that they can be simultaneously satisfied if the coupling g is small in the Planck units. The range of the field

values in which the evolution is classical and at the same time it proceeds in the slow roll regime, is⁴

$$\frac{nM_{Pl}}{4\sqrt{3}\pi} \ll \phi \ll \left(\frac{M_{Pl}^4}{g}\right)^{1/n}. \quad (12.27)$$

We show in Chapter 13 that the coupling g must indeed be small (in the Planck units) to ensure the agreement of the theory with observations, so the whole picture is self-consistent.

The simplest and most familiar power-law potentials are quadratic and quartic,

$$V_2(\phi) = \frac{m^2}{2}\phi^2, \quad (12.28)$$

$$V_4(\phi) = \frac{\lambda}{4}\phi^4; \quad (12.29)$$

the constant m has dimension of mass and the coupling λ is dimensionless. According to the above discussion, they have to be small,

$$m \ll M_{Pl}, \quad \lambda \ll 1.$$

The slow roll conditions are satisfied for

$$\phi \gg \frac{M_{Pl}}{\sqrt{3}\pi},$$

where we have omitted a factor of order 1 which is not very relevant for what follows. We note here that when writing the expressions (12.28), (12.29), and, generally, (12.23), we assume, in fact, that they give the dominant terms in the potential $V(\phi)$ at $\phi \gg M_{Pl}$; if the potential has other terms, they must be negligibly small at super-Planckian fields.

Let us now describe a scenario for the beginning of this large field inflation. Let us assume that the Universe “in the beginning” was rather wild, as we discussed in Section 11.1. Namely, it was strongly inhomogeneous at scales exceeding l_{Pl} , strongly curved (the radius of spatial curvature was of order l_{Pl} everywhere), and the energy density was high, $\rho \sim M_{Pl}^4$. Let the theory contain the inflaton field whose potential obeys the slow roll conditions at large ϕ . Our guess is that the typical value of the inflaton field at the Planck epoch is near the upper boundary of the interval (12.27), i.e., its potential energy density is also Planckian. Unlike in the homogeneous situation discussed above, the gradient terms $\partial_i\phi\partial_j\phi$ in the energy-momentum tensor are, generally speaking, of the order of the potential term. Also, the spatial curvature term in the Friedmann equation is important. We note, however, that both of these effects decrease like $1/a^2$ as the Universe expands, while the scalar potential remains almost constant. It is natural to think that there may exist (or may emerge as a result of fluctuation) a patch of size somewhat

⁴The discussion here is somewhat oversimplified, see Section 14.1, but the qualitative conclusion that the classical slow roll inflation occurs in a wide range of field values is correct.

larger than Planckian (but not necessarily very much larger) where the gradient and spatial curvature terms in the Friedmann equation are somewhat smaller than the potential term (but, again, not very much smaller). This patch starts to expand, and it soon enters the inflationary regime. This is one possible way of making the classical inflation consistent with our guess on the Planckian epoch. This scenario, and large field inflation in general, was proposed by Linde [104]; since initial conditions ensuring inflation may emerge accidentally in this scenario, it is also called “chaotic inflation”.

Inflation lasts until the inflaton field reaches the lower boundary of the interval (12.27). Then the Hubble friction term $3H\dot{\phi}$ becomes small compared to the term $\dot{\phi}$ in Eq. (12.23), and the inflaton starts to oscillate around the minimum of the potential $V(\phi)$. Fast oscillations of the field ϕ give rise to particle creation, provided, of course, the inflaton interacts with conventional particles. The particles thermalize in the end, and the Universe enters the hot stage of its evolution. The dynamics of the post-inflationary reheating is fairly complex and interesting; some of its aspects are studied in Chapter 15. For the time being, we do not need anything other than intuitive understanding that non-stationary processes that begin with the inflaton oscillations, in the end establish thermal equilibrium throughout the Universe.

The upper bound on the initial temperature of the hot stage T_{reh} is obtained by noticing that the thermal energy just after inflation cannot exceed the inflaton potential energy at the end of inflation,

$$T_{reh}^4 \lesssim T_{max}^4 \sim \frac{1}{g_*} V(\phi \approx M_{Pl}) , \quad (12.30)$$

where $g_* = g_*(T_{reh})$ is, as usual, the effective number of degrees of freedom. It follows from (12.25) that this temperature is much smaller than M_{Pl} . In fact, the post-inflationary reheating often takes many Hubble times, so the Universe expands considerably during that epoch and the reheat temperature is several orders of magnitude lower than T_{max} determined by the right hand side of (12.30).

We now come back to the question of how inflation solves the entropy problem of the Hot Big Bang theory. Matter at inflation does not carry entropy, but has high energy density. This energy transforms into heat at the reheating epoch. This occurs in huge volume created in the course of the inflationary expansion, so huge overall entropy is generated in the inflated patch of the Universe.

Let us estimate the total duration of inflation and total number of e -foldings in the outlined scenario. It is convenient for what follows to study slightly more general quantity. Namely, let us calculate the number of e -foldings $N_e(\phi)$ from the time t_ϕ at the inflationary epoch when the inflaton field takes a given value ϕ , to the end of inflation. We will see in Section 13 that the range $N_e \simeq 50\text{--}60$ is of particular interest. In the spirit of definition given in Section 11.2, we write

$$N_e(\phi) = \log \left[\frac{a_e}{a(\phi)} \right] = \int_{t_\phi}^{t_e} H(t) dt ,$$

where a_e and $a(\phi)$ are the scale factors at the time t_e of inflation end and time t_ϕ , respectively. We now use sequentially Eqs. (12.11) and (12.12), valid for the slow roll inflation regime, and obtain

$$N_e(\phi) = \int_{\phi}^{\phi_e} H(\phi) \frac{d\phi}{\dot{\phi}} = \int_{\phi_e}^{\phi} 3H^2 \frac{d\phi}{V'} = \frac{8\pi}{M_{Pl}^2} \int_{\phi_e}^{\phi} \frac{V}{V'} d\phi, \quad (12.31)$$

where ϕ_e is the value of the inflaton field at the end of inflation. We are interested in rather large $N_e(\phi)$ which corresponds to the epoch long before the inflation end. For power-law potentials (12.23) this means

$$\phi \gg \phi_e, \quad (12.32)$$

and our final result is

$$N_e(\phi) = \frac{4\pi}{n} \frac{\phi^2}{M_{Pl}^2}. \quad (12.33)$$

Among other things, the latter formula shows that the interesting number of e -foldings, $N_e = 50-60$, is indeed obtained for super-Planckian fields,

$$\phi(N_e) = \sqrt{\frac{n N_e}{4\pi}} \cdot M_{Pl} \quad (12.34)$$

$$= (2.8 - 3.1) \cdot M_{Pl}, \quad n = 2, \quad N_e = 50 - 60 \quad (12.35)$$

$$= (4.0 - 4.4) \cdot M_{Pl}, \quad n = 4, \quad N_e = 50 - 60. \quad (12.36)$$

We also note the simple relations between the slow roll parameters and number of e -foldings before the inflation end,

$$\epsilon = \frac{n}{4N_e}, \quad (12.37)$$

$$\eta = \frac{n-1}{2N_e}. \quad (12.38)$$

These relations follow for power-law potentials from the definitions (12.20), (12.21) and the relation (12.33); they are valid well before the end of inflation when the inequality (12.32) holds.

Coming back to the estimate of the total number of e -foldings, we assume, according to the above scenario, that inflation begins at the time when the inflaton energy density is comparable to the Planck value, $V(\phi_i) \sim M_{Pl}^4$. Then we have for the power-law potentials (12.23)

$$\phi_i \sim g^{-1/n} M_{Pl}^{4/n} \quad (12.39)$$

and

$$N_e^{(tot)} = N_e(\phi_i) = \frac{4\pi}{n} \left(\frac{M_{Pl}^{4-n}}{g} \right)^{2/n}. \quad (12.40)$$

We see that this number is large for small coupling g in the Planck units, which implies successful inflation. In realistic models with correct amplitude of primordial

perturbations generated at inflation, the right hand side of (12.40) is very large, and the increase of the scale factor is huge.

As two examples, consider the quadratic and quartic potentials (12.28) and (12.29). According to (12.40), the total numbers of e -foldings in these models are estimated by

$$n = 2: \quad N_e^{(tot)} \simeq 4\pi \frac{M_{Pl}^2}{m^2}, \quad (12.41)$$

$$n = 4: \quad N_e^{(tot)} \simeq \frac{2\pi}{\sqrt{\lambda}}. \quad (12.42)$$

We will see in Section 13 that the correct amplitude of density perturbations, $\delta\rho/\rho \sim 5 \cdot 10^{-5}$, is obtained for

$$n = 2: \quad m \sim 10^{-6} M_{Pl}, \quad (12.43)$$

$$n = 4: \quad \lambda \sim 10^{-13}. \quad (12.44)$$

Hence, the total numbers of e -foldings are $N_e^{(tot)} \sim 10^{13}$ and $N_e^{(tot)} \sim 10^7$, respectively. The microscopic, Planck-size patch of the Universe stretches out to a region of size

$$R \sim 10^{10^{13}} \text{ and } 10^{10^7}, \quad (12.45)$$

no matter in which units, the Planck lengths or Gigaparsecs. Inflation does its job very well.

Let us finally estimate the total duration of inflation, again using quadratic and quartic potentials for definiteness. We write by definition

$$\Delta t^{(tot)} = \int_{t_i}^{t_e} dt = \int_{\phi_i}^{\phi_e} \frac{d\phi}{\dot{\phi}}.$$

We now use the slow roll equations (12.11), (12.12) and obtain

$$\Delta t^{(tot)} = \int_{\phi_e}^{\phi_i} \frac{3Hd\phi}{V'(\phi)} = 3\sqrt{\frac{8\pi}{3}} \frac{1}{M_{Pl}} \int_{\phi_e}^{\phi_i} \frac{\sqrt{V}d\phi}{V'}.$$

Omitting numerical factors of order 1 and using (12.39), (12.43), (12.44) we find

$$n = 2: \quad \Delta t^{(tot)} \sim \frac{1}{M_{Pl}} \frac{\phi_i}{m} \sim \frac{M_{Pl}}{m^2} \sim 10^{-31} \text{ s}, \quad (12.46a)$$

$$n = 4: \quad \Delta t^{(tot)} \sim \frac{1}{M_{Pl}} \frac{1}{\sqrt{\lambda}} \log \frac{\phi_i}{\phi_e} \sim \frac{1}{M_{Pl}} \frac{1}{\sqrt{\lambda}} \log \frac{1}{\lambda} \sim 10^{-35} \text{ s}. \quad (12.46b)$$

Hence, a patch of the Universe inflates from microscopic to huge size in a very short period of time. This is of course due to the exponential character of the cosmological expansion at the inflationary epoch.

Problem 12.3. Find the dependence on time of the inflaton field and Hubble parameter at inflation for quadratic and quartic scalar potentials by solving the slow roll equations explicitly. Confirm the estimates (12.46) in this way.

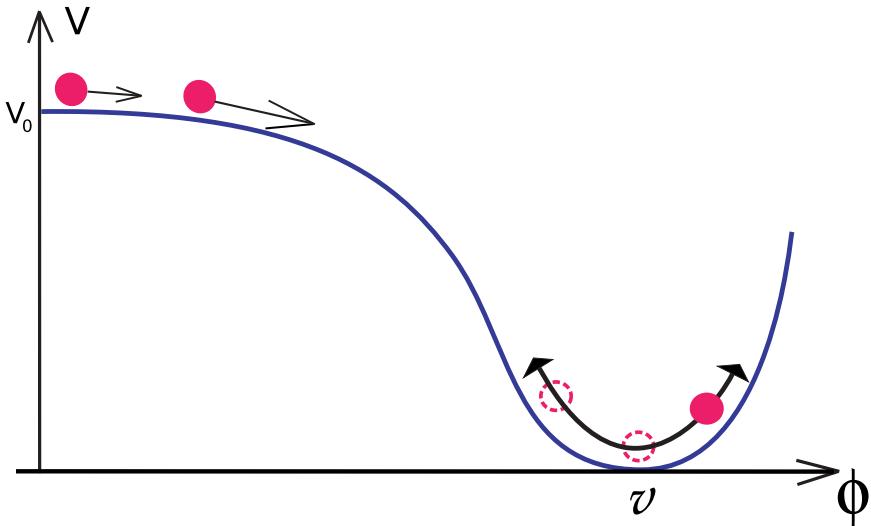


Fig. 12.3 The inflaton potential in the small field inflationary model.

12.2.2 Small field inflation (“new”)

Unlike the large field inflation, the model we study here does not invoke super-Planckian fields. The view on the possible beginning of inflation is also quite different from the idea of the chaotic initial conditions discussed in Section 12.2.1. As we show in Chapter 13, the two models also differ in their predictions for scalar and tensor perturbations. At the same time, the flatness requirement for the inflaton potential is common to both models.

Consider the inflaton potential shown in Fig. 12.3. Assume that the potential has the following form near the origin,

$$V(\phi) = V_0 - g\phi^n, \quad (12.47)$$

where

$$n \geq 3.$$

The most popular is the quartic potential

$$V(\phi) = V_0 - \frac{\lambda}{4}\phi^4. \quad (12.48)$$

In what follows we consider this particular case, and comment on arbitrary n in appropriate places.

According to their definitions (12.20) and (12.21), the slow roll parameters at small ϕ are

$$\epsilon = \frac{M_{Pl}^2}{16\pi} \frac{\lambda^2 \phi^6}{V_0^2}, \quad (12.49)$$

$$\eta = -\frac{M_{Pl}^2}{8\pi} \frac{3\lambda\phi^2}{V_0}. \quad (12.50)$$

We see that the slow roll regime $\epsilon, \eta \ll 1$ indeed holds at small fields. Once one assumes that the initial inflaton value is close to $\phi = 0$, one obtains the model of inflation.

Small initial value of the inflaton field, necessary in this scenario, may come out in the following way. Assume that the Universe was filled with hot matter close to thermal equilibrium before inflation. In this sense the pre-inflationary stage was similar to the Hot Big Bang epoch, but with no strong requirements imposed on homogeneity, isotropy and spatial flatness. We have seen in Section I.10.2 that the effective potential of a scalar field is different from the zero-temperature potential $V(\phi)$ and that the minimum of the former is typically at $\phi = 0$. If the inflaton also has this property, its initial value is naturally equal to zero. As the Universe expands and cools down, the effective potential gradually develops its zero-temperature form, and the inflationary epoch begins. The point $\phi = 0$ is unstable in the theory with the potential of Fig. 12.3, and quantum effects slightly shift the field from zero. We give the estimate of these effects and hence the total number of inflationary e -foldings later on.

The case of interest is when the slow roll ends at small inflaton field, such that

$$\lambda\phi_e^4 \ll V_0. \quad (12.51)$$

The opposite situation is basically a version of the large field inflation studied in Section 12.2.1. Hence, we concentrate on the case (12.51). Then there is hierarchy between the slow roll parameters,

$$\epsilon \ll |\eta|, \quad (12.52)$$

and we see from (12.50) that

$$\eta < 0. \quad (12.53)$$

The property (12.52) follows directly from (12.49), (12.50) and (12.51),

$$\frac{\epsilon}{|\eta|} = \frac{\lambda\phi^4}{6V_0} \leq \frac{\lambda\phi_e^4}{6V_0} \ll 1.$$

In view of (12.52), we find that inflation ends at

$$|\eta_e| = \frac{M_{Pl}^2}{8\pi} \frac{3\lambda\phi_e^2}{V_0} \sim 1,$$

so the condition (12.51) restricts the inflaton potential,

$$\frac{V_0}{M_{Pl}^4} \ll \left(\frac{3}{8\pi}\right)^2 \lambda. \quad (12.54)$$

This inflationary model [108, 109] is called small field inflation, or “new” inflation for historical reasons. It was the second, after Starobinsky’s, successful inflationary

model. Once inflaton reaches the value ϕ_e , the inflationary regime terminates, ϕ rapidly rolls down to the minimum of $V(\phi)$, and the further cosmological evolution proceeds in the same way as in the large field model.

Problem 12.4. Consider the inflaton potential which has the shape shown in Fig. 12.3, but behaves at small ϕ as follows,

$$V(\phi) = V_0 - \frac{m^2}{2}\phi^2 - \frac{\lambda}{4}\phi^4,$$

with no fine tuning of parameters. Let the initial inflaton value be close to $\phi = 0$. Show that either inflation does not occur at all, or it proceeds in the same regime as for the scalar potential (12.48), or the inflationary regime is the same as in the large field model. Hence, introducing quadratic term into the scalar potential does not give anything new, unless parameters are fine tuned.

We will see in Chapter 13 that the correct amplitude of the primordial density perturbations is obtained in this scenario at very small value of the coupling,

$$\lambda \sim 10^{-13}. \quad (12.55)$$

It follows from (12.54) that, like in the large field scenario, the reheat temperature is rather small,

$$T_{reh} \lesssim \left(\frac{1}{g_*} V_0 \right)^{1/4} \ll 10^{-3} M_{Pl}.$$

We note that the only restriction on V_0 is the upper bound (12.54), so V_0 can be much smaller than λM_{Pl}^4 . Accordingly, the reheat temperature can be quite small. We see that the small field inflation is an example of a cosmological model with low maximum temperature obtained without fine tuning of parameters.

Let us find the relation between the inflaton field at the inflationary stage and the number of e -foldings to the inflation end. We use (12.31) and obtain for $N_e \gg 1$ and hence $\phi \ll \phi_e$,

$$N_e(\phi) = \frac{3H^2}{2\lambda\phi^2} = \frac{4\pi V_0}{\lambda M_{Pl}^2 \phi^2}. \quad (12.56)$$

It now follows from (12.50) that the slow roll parameter is, cf. (12.38),

$$\eta = -\frac{3}{2N_e}.$$

In the general power-law case (12.47) the relation has the following form,

$$\eta = -\frac{n-1}{n-2} \frac{1}{N_e}. \quad (12.57)$$

Problem 12.5. Derive the last relation.

We find from (12.56) that the total number of e -foldings is determined by the initial inflaton value ϕ_i , namely,

$$N_e^{(tot)} = \frac{3H^2}{2\lambda\phi_i^2} = \frac{4\pi V_0}{\lambda M_{Pl}^2 \phi_i^2}. \quad (12.58)$$

If the average inflaton field is zero at the beginning of inflation, as we discussed above, the initial field ϕ_i is determined by quantum fluctuations. We will see in Chapter 13 that their amplitude is estimated as

$$\phi_i \sim \delta\phi \sim H.$$

It then follows from (12.58) that

$$N_e^{(tot)} \sim \frac{1}{\lambda}.$$

Recalling (12.55) we come to the same conclusion as in the large field case (see (12.45)): inflation stretches out a microscopic patch of the Universe to huge volume.

Problem 12.6. Obtain the estimate for the duration of inflation analogous to (12.46).

12.2.3 Hybrid inflation

Our third example [110] is similar to the large field model from the viewpoint of the beginning of inflation. However, it does not rely upon super-Planckian fields. In its simple version, the model contains two scalar fields, the inflaton field ϕ and an additional field χ (sometimes called waterfall field). In this regard, the model is called hybrid inflation, or multi-field inflation. The scalar potential is chosen as shown in Fig. 12.4. The idea is that the field ϕ is sufficiently large at inflation, and it slowly rolls down the valley along $\chi = 0$. Once the valley becomes a saddle, the system of fields rapidly rolls down in the orthogonal direction, inflation ends, and the oscillations near the minimum $\phi = 0$, $\chi = v$ reheat the Universe.

For definiteness, we choose the scalar potential as follows,

$$V(\phi, \chi) = \frac{1}{2}(g^2\phi^2 - \mu^2)\chi^2 + \frac{h}{4}\chi^4 + U(\phi) + V_0,$$

where g and h are positive dimensionless couplings, μ is a parameter of dimension of mass and $U(\phi)$ is monotonically increasing inflaton potential. By choosing V_0 we set

$$U(\phi = 0) = 0.$$

We assume, also for definiteness, that $U(\phi)$ has power-law behavior in the relevant range of ϕ . As an example,

$$U(\phi) = \frac{m^2}{2}\phi^2. \quad (12.59)$$

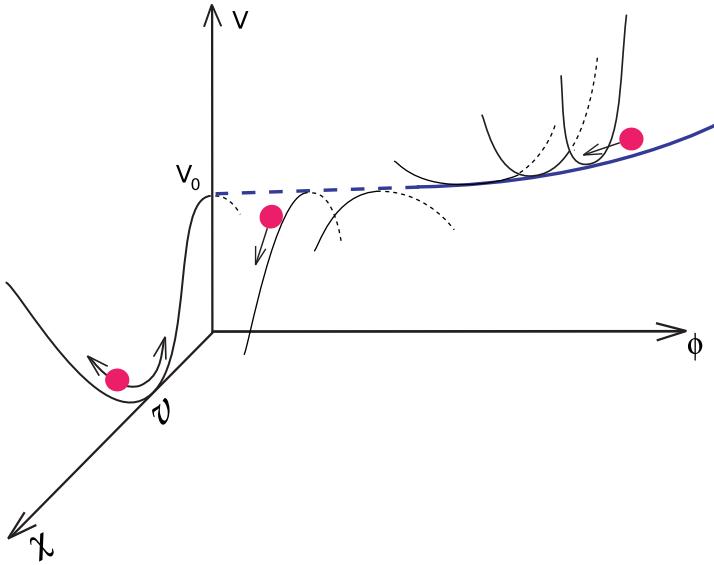


Fig. 12.4 The scalar potential in the hybrid inflation model.

The potential $V(\phi, \chi)$ has the minimum at

$$\phi = 0, \quad \chi = v = \frac{\mu}{\sqrt{h}}.$$

The constant V_0 is determined from the requirement that the potential vanishes at the minimum, which gives

$$V_0 = \frac{\mu^4}{4h}.$$

The line $\chi = 0$ is a valley of the potential $V(\phi, \chi)$ at $\phi > \phi_c$, where

$$\phi_c = \frac{\mu}{g}.$$

We are going to consider the case in which the energy of inflaton is small compared to the total energy at the critical point ϕ_c ,

$$U_c \equiv U(\phi_c) \ll V_0. \quad (12.60)$$

We also require that inflation continues up to $\phi = \phi_c$. The latter two conditions ensure that the dynamics in this model is substantially different from that in the large field model, and the relevant values of the inflaton field $\phi \sim \phi_c$ are small compared to M_{Pl} (see problem 12.9). The slow roll condition for the inflaton at $\phi \geq \phi_c$ is

$$\eta(\phi_c) = \frac{M_{Pl}^2}{8\pi} \frac{U''(\phi_c)}{V_0} \ll 1.$$

This gives for power-law potentials $U(\phi)$,

$$\eta(\phi_c) \sim \frac{M_{Pl}^2}{8\pi\phi_c^2} \frac{U_c}{V_0} \ll 1. \quad (12.61)$$

Note that

$$\eta(\phi_c) > 0.$$

The second slow roll parameter is small in this model, in analogy to the small field model,

$$\epsilon(\phi_c) = \frac{M_{Pl}^2}{16\pi} \left(\frac{U'(\phi_c)}{V_0} \right)^2 \sim \frac{M_{Pl}^2}{16\pi\phi_c^2} \left(\frac{U_c}{V_0} \right)^2 \sim \eta \cdot \frac{U_c}{V_0} \ll \eta.$$

It follows from (12.61) that inflation can indeed occur at sub-Planckian fields; for $\phi_c \ll M_{Pl}$, the small value of η can be due to the small ratio U_c/V_0 .

Let us now find the range of parameters for which the system, in a few Hubble times after the inflaton reaches the critical value ϕ_c , rapidly rolls down to the minimum $\phi = 0, \chi = v$. In Hubble time, the inflaton shifts from the critical value ϕ_c to $\phi_c - \Delta\phi$, where

$$\Delta\phi \sim -\dot{\phi}\Delta t \sim \frac{U'(\phi_c)}{3H^2},$$

where the last estimate is obtained by making use of the slow roll equation (12.11). At this value of the inflaton field, the potential for the field χ near $\chi = 0$ is

$$V_{eff}(\chi) = V_0 - \frac{\mu_{eff}^2}{2} \chi^2, \quad (12.62)$$

where

$$\mu_{eff}^2 = 2g^2\phi_c\Delta\phi \sim g^2 \frac{2U'\phi_c}{3H^2}.$$

We make use of the Friedmann equation, recall that $U(\phi)$ has power-law behavior and obtain

$$\mu_{eff}^2 \sim g^2 \frac{M_{Pl}^2}{4\pi} \frac{U_c}{V_0}. \quad (12.63)$$

The waterfall field χ rolls down the potential $V_{eff}(\chi)$. This occurs in the fast roll regime, provided that

$$\eta_\chi = \frac{M_{Pl}^2}{8\pi} \frac{V''_{eff}}{V_{eff}} \gtrsim 1$$

(the parameter ϵ_χ is small for small χ). We recall (12.62), (12.63) and write the latter inequality as follows,

$$\frac{g^2}{(4\pi)^2} \frac{U_c}{V_0} \frac{M_{Pl}^4}{V_0} \gtrsim 1. \quad (12.64)$$

Hence, the hybrid inflation indeed occurs, provided the inequalities (12.60), (12.61) and (12.64) are satisfied. Assuming also that the critical inflaton field is sub-Planckian, we write the final chain of inequalities,

$$\frac{(4\pi)^2}{g^2} \frac{V_0}{M_{Pl}^4} \lesssim \frac{U_c}{V_0} \ll \frac{\phi_c^2}{M_{Pl}^2} \ll 1. \quad (12.65)$$

These inequalities are satisfied in a wide range of the parameters entering the potential $V(\phi, \chi)$, i.e., hybrid inflation does not require fine tuning.

Problem 12.7. Write the chain of inequalities (12.65) in terms of the original parameters g , h , μ and m for quadratic inflaton potential (12.59). Check that the parameters may vary in a wide range without violating (12.65). Hint: Assume that the theory is weakly coupled, $g, h \lesssim 1$, and the dimensionful parameters m , μ and v are small compared to M_{Pl} .

Let us again find the dependence of the number of e -foldings before inflation end on the inflaton field. We use Eq. (12.31) and write

$$N_e(\phi) = \frac{8\pi V_0}{M_{Pl}^2} \int_{\phi_c}^{\phi} \frac{d\phi}{U'(\phi)}.$$

Here we accounted for the fact that inflation ends when the inflaton field reaches the critical value, i.e., $\phi_e = \phi_c$. If N_e is not very large (as we already pointed out, the interesting range is $N_e \simeq 50 - 60$), we have

$$N_e(\phi) \simeq \frac{8\pi V_0}{M_{Pl}^2 U'(\phi_c)} (\phi - \phi_c) \simeq \frac{1}{\eta(\phi_c)} \frac{\phi - \phi_c}{\phi_c},$$

where we used (12.61). We see that ϕ is not very distant from ϕ_c for interesting N_e . Hence,

$$\eta(\phi) \simeq \eta(\phi_c).$$

There is no relation between η and N_e similar to (12.38), and η is an independent parameter in the hybrid inflation model in the sense that it is unrelated to other phenomenologically important parameters. The only property of η is that it is small.

In fact, one can obtain a model-dependent bound on η in the hybrid inflation model by making use of the following argument. Consider for definiteness the quadratic inflaton potential (12.59). Then η is independent of ϕ ,

$$\eta = \frac{M_{Pl}^2 m^2}{8\pi V_0}.$$

Let us find the total number of e -foldings for the entire inflationary epoch. We use Eq. (12.31) and obtain

$$N_e^{(tot)} = \frac{8\pi V_0}{M_{Pl}^2 m^2} \log \frac{\phi_i}{\phi_c},$$

which gives

$$\eta = \frac{1}{N_e^{(tot)}} \log \frac{\phi_i}{\phi_c}.$$

If inflation begins at $\phi_i \sim M_{Pl}$ (this is the case, in particular, if the inflaton potential contains non-renormalizable terms (12.67)), then the latter equality becomes

$$\eta = \frac{1}{N_e^{(tot)}} \log \frac{M_{Pl}}{\phi_c}.$$

Since $N_e^{(tot)} \gtrsim 60$ for successful inflation, the latter equality implies an upper bound on η which, however, logarithmically depends on the parameters of the scalar potential through ϕ_c . Similar bound is stronger for the power-law inflaton potential $U \propto \phi^n$ with $n > 2$.

Problem 12.8. Assuming that there is no fine tuning, and, in particular, that $N_e^{(tot)} \gg 60$, find the bound on η at 60 e-foldings before the end of inflation in the hybrid inflation model with the inflaton potential $U \propto \phi^n$, $n > 2$.

Problem 12.9. Extend the analysis of the hybrid inflation model to the case $U_c \sim V_0$ in which $\phi_c \gtrsim M_{Pl}$. What region in the plane (ϵ, η) corresponds to models of this sort?

12.2.4 Concluding remarks

We end this Chapter with several comments.

First, the slow roll regime of the inflaton evolution, as well as nearly exponential cosmological expansion are not, generally speaking, pre-requisites of the inflationary solutions to the horizon, flatness and entropy problems. This remark is not very relevant, though, in view of the cosmological data showing that the primordial spectrum of density perturbations is nearly flat, see Chapter 13: these data rule out the simplest fast roll models.

Problem 12.10. Power-law inflation

Show that the Universe inflates (i.e., the horizon and flatness problems get solved) if it expands according to the law

$$a = a_0 t^\alpha, \quad \alpha > 1. \quad (12.66)$$

In a model where this expansion is driven by the inflaton field ϕ , find the inflaton potential without assuming slow roll. At what values of α and/or ϕ inflation occurs in the slow roll regime?

Second, the required flatness of the inflaton potential is a very non-trivial property from the particle physics viewpoint. The reason is that the dynamics of spatially homogeneous scalar field is determined not by the tree-level potential entering the classical Lagrangian but by the *effective potential* which includes quantum corrections. If the inflaton interactions with other fields are not particularly weak, these corrections are, generally speaking, large, and the effective potential is not flat

enough. This problem is easier in supersymmetric extensions of the Standard Model, but its solution is not simple even there [105]. In the large field models, there is an additional difficulty. These models invoke super-Planckian fields, but one expects, on general grounds, that the inflaton effective potential contains higher order terms (“non-renormalizable interactions”) suppressed by the gravitational mass scale M_{Pl} . They have the following form,

$$\Delta V(\phi) = c_6 \frac{\phi^6}{M_{Pl}^2} + c_8 \frac{\phi^8}{M_{Pl}^4} + \dots \quad (12.67)$$

where constants $c_{6,8,\dots}$ are offhand of order 1. These terms are negligible at $\phi \ll M_{Pl}$, but they dominate at super-Planckian fields and ruin the slow roll regime. This situation indeed takes place, generally speaking, in supergravity models [105]. Hence, the mechanism responsible for the flatness of the inflaton effective potential at $\phi \gg M_{Pl}$ is of considerable interest for theories pretending to describe physics at extreme energy scales, notably, superstring theory. Recently, there has been some progress in understanding this issue [106, 107].

Problem 12.11. Estimate, for quadratic potential (12.28), the range of parameters c_6 and c_8 where the new terms (12.67) are irrelevant at 60 e-foldings before the end of inflation. Do the same for quartic potential (12.29). The numerical values of m and λ are given in (12.43) and (12.44), respectively.

Third, as we mentioned in the beginning of Section 12.2, inflation is not necessarily driven by a scalar field rolling down its potential. In any case, an important aspect of all models is the mechanism by which inflation ends and the Universe gets reheated. This occurs automatically in models we have discussed, but some models have severe “graceful exit” problem, to the extent that it makes them unrealistic. An example is “old inflation” briefly discussed in the beginning of this Chapter.

Finally, we paid some attention in this Chapter to possible scenarios of the beginning of inflation. If there is no fine tuning of parameters, and the total number of e-foldings exceeds 100, this aspect is purely academic, at least at the current understanding of this point. Because of long inflationary epoch, the properties of the Universe before and at the beginning of inflation are completely forgotten, and the mechanism that started inflation is beyond experimental study. It may well be that the question about the beginning of the cosmological evolution does not belong to the category of questions which can be answered by comparing theory with experiment.

To end this Chapter, let us briefly consider the Starobinsky model which we mentioned in the beginning of Section 11.2. The original model does not involve scalar fields, but modifies the Einstein gravity. The gravitational action contains both linear and quadratic terms in the curvature scalar,

$$S = -\frac{M_{Pl}^2}{16\pi} \int \sqrt{-g} d^4x f(R), \quad f(R) = R - \frac{R^2}{6M^2}, \quad (12.68)$$

where the parameter M has dimension of mass. As shown in Section I.A.5, this theory is equivalent to General Relativity with a scalar field. Upon introducing a new metric, $\tilde{g}_{\mu\nu} = g_{\mu\nu}/\psi$, one obtains that the action of this equivalent theory has the form (I.A.77). We restore the dimensionful parameters and write that action as follows,

$$S = -\frac{M_{Pl}^2}{16\pi} \int \sqrt{-g} d^4x \tilde{R} + \frac{M_{Pl}^2}{16\pi} \int \sqrt{-g} d^4x \left(\frac{3}{2} \frac{\tilde{g}^{\mu\nu} \tilde{\nabla}_\mu \psi \tilde{\nabla}_\nu \psi}{\psi^2} + \frac{R_0(\psi)}{\psi} - \frac{f[R_0(\psi)]}{\psi^2} \right),$$

where $R_0(\psi) = 3(1-\psi)M^2$. The change of variables

$$\psi = \exp\left(\frac{4\sqrt{\pi}}{\sqrt{3}} \frac{\phi}{M_{Pl}}\right)$$

reduces the kinetic term to the canonical form, and the scalar potential becomes

$$V = \frac{3M^2 M_{Pl}^2}{32\pi^2} \left(1 - \frac{1}{\psi}\right)^2 = \frac{3M^2 M_{Pl}^2}{32\pi^2} \left(1 - e^{-\frac{4\sqrt{\pi}}{\sqrt{3}} \frac{\phi}{M_{Pl}}}\right)^2. \quad (12.69)$$

The potential flattens out at large fields $\phi \gtrsim M_{Pl}$, while it tends to zero as $\phi \rightarrow 0$. The inflationary behavior is obtained for large fields, and inflation occurs in the slow roll regime. We note that the interaction of the field ϕ with other fields, including the Standard Model ones, is induced by the redefinition of the metric. Thus, it is universal and of the gravitational strength.

Hence, the Starobinsky model is equivalent to chaotic inflation with one inflaton field and potential (12.69).

Problem 12.12. Find the range of the parameter M and field ϕ where the slow roll approximation is valid.

This page is intentionally left blank

Chapter 13

Generation of Cosmological Perturbations at Inflation

One of the main advantages of the inflation hypothesis is that it explains the existence of the primordial cosmological perturbations. In the inflationary scenario, the primordial density perturbations are generated as the result of the enhancement of the inflaton *vacuum quantum fluctuations*.¹ This enhancement is precisely due to the accelerated cosmological expansion at inflation [111–115]. Similar mechanism generates tensor perturbations from vacuum fluctuations of the gravitational field [100, 116–119].

We note that the enhancement of zero-point fluctuations in expanding Universe was first suggested as a mechanism of the generation of scalar perturbations without reference to inflation [120] (see also Ref. [121] where the possibility of the efficient generation of scalar perturbations at inflation was mentioned as well). The generation of gravity waves in the expanding Universe was also considered before the inflation idea has been put forward [122].

The predictions of the inflationary mechanism are in overall agreement with the existing observational data that show the high degree of Gaussianity of the scalar perturbations, the flat or nearly flat primordial scalar spectrum and the domination of the adiabatic mode (i.e., no or almost no admixture of isocurvature modes). We emphasize, however, that the data do not yet prove the validity of inflationary theory. We have pointed out in Section 5.1 that the adiabatic mode is physically very natural; conversely, isocurvature modes would require rather exotic generation mechanism of dark matter and/or baryon asymmetry. The flat, Harrison–Zeldovich spectrum (proposed, in fact, before the idea of inflation has been suggested) is also singled out on general grounds due to its scale invariance and hence it may not be specific to inflation. So, the confirmation of the inflationary theory needs the observation of more delicate features of the primordial cosmological perturbations. Inflation predicts that the scalar spectrum is slightly tilted, so the observational

¹We consider in Section 14.2 the possibility that the density perturbations result from the enhanced vacuum fluctuations of “curvaton”, a field other than inflaton. This is quite similar to the inflaton mechanism.

detection of the tilt would be an argument (not particularly strong) in favor of inflation. What is more important, some popular inflationary models predict sizeable tensor perturbations with almost flat primordial power spectrum. The observation of the latter would be a triumph of the inflationary theory.

We emphasize that the concrete results on the shape of the scalar spectrum and the amplitude of the tensor spectrum depend on the model of inflation (on the inflaton potential in the slow roll models), so inflation does not actually give definite predictions for them. There exists a relation common to all slow roll models, but it involves the tensor tilt and hence it will be tested in distant future, if ever.

We consider these fascinating issues in this Chapter. We begin with a simplified discussion of the generation of scalar perturbations which grasps the major properties of the inflationary mechanism and yields the correct result. The accurate analysis is left for Section 13.2. We then study the generation of tensor perturbations and discuss the predictions of the inflationary theory for the observational cosmology.

13.1 Simplified Analysis: Inflaton Fluctuations

The inflaton field, like any other field, has vacuum fluctuations. These fluctuations get strongly enhanced at inflation, so that macroscopic perturbations of inflaton, and hence energy density, are generated. These perturbations induce perturbations of metric, see Chapter 2. In this Section we use the approximation in which the latter effect is neglected. Hence, we study inflaton fluctuations in the homogeneous background metric of the inflating Universe. This approximation is not adequate for large wavelength perturbations (superhorizon modes, see Section 2.4), but it illustrates the physics well enough and quickly gives the result.

Before studying the inflaton fluctuations at inflation, it is instructive to consider vacuum fluctuations of a scalar field in the Minkowski space-time.

13.1.1 Vacuum fluctuations in flat space-time

Consider free real massless scalar field in the Minkowski space-time. Its action is

$$S_\varphi = \frac{1}{2} \int d^4x \eta^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi = \frac{1}{2} \int d^4x [(\partial_t \varphi)^2 - (\partial_i \varphi)^2],$$

and the energy functional is given by

$$E = \frac{1}{2} \int d^3x [(\partial_t \varphi)^2 + (\partial_i \varphi)^2]. \quad (13.1)$$

Let us estimate, at qualitative level first, the amplitude of vacuum fluctuations of typical momentum q and frequency $\omega_q = q$. To this end, we note that the energy of these fluctuations in a typical spatial volume $\lambda^3 \sim q^{-3}$ should be of the order of the zero-point fluctuation energy of an oscillator of frequency ω_q , $E = \hbar\omega_q/2$. On the

other hand, if $\delta\varphi_q$ is the amplitude of the fluctuations we discuss, then the energy density, estimated using (13.1), is of order

$$\rho \sim \omega_q^2 (\delta\varphi_q)^2.$$

Setting $\hbar = 1$, we have

$$\frac{\omega_q}{2} \sim \rho \lambda^3 \sim \omega_q^2 (\delta\varphi_q)^2 \lambda^3,$$

and with $\omega_q = q$, $\lambda \sim q^{-1}$ we obtain

$$\delta\varphi_q \sim q. \quad (13.2)$$

This gives the desired estimate of the amplitude of the vacuum fluctuations of typical momentum q ; this estimate is in fact evident on dimensional grounds. We see that the amplitudes of vacuum fluctuation are large at high momenta q and hence at small wavelengths, and, conversely, fluctuations are small at large wavelengths. These fluctuations do not show up at macroscopic distance scales.

To refine our estimate, we write the standard expression for the field operator,

$$\varphi(\mathbf{x}, t) = \int \frac{d^3 q}{(2\pi)^{3/2} \sqrt{2\omega_q}} (e^{i\omega_q t - i\mathbf{q}\mathbf{x}} A_{\mathbf{q}}^\dagger + e^{-i\omega_q t + i\mathbf{q}\mathbf{x}} A_{\mathbf{q}}), \quad (13.3)$$

where the creation and annihilation operators obey the standard commutational relations,

$$[A_{\mathbf{q}}, A_{\mathbf{q}'}^\dagger] = \delta(\mathbf{q} - \mathbf{q}'). \quad (13.4)$$

The equal-time vacuum correlation functions of the operators $\varphi(\mathbf{x}, t)$ are expressed in a standard way through the two-point function,

$$\langle 0 | \varphi(\mathbf{x}, t) \varphi(\mathbf{y}, t) | 0 \rangle,$$

where $|0\rangle$ denotes the vacuum state, $A_{\mathbf{q}}|0\rangle = 0$. Namely, the higher order correlators obey the Wick theorem. As we discuss in Section C.2, the latter property is characteristic of Gaussian random fields; in this sense the vacuum fluctuations of free quantum fields (the massless scalar field in our case) are Gaussian random fields. Let us define, in analogy to the theory of Gaussian random fields, the power spectrum $\mathcal{P}_\varphi(q)$ of the vacuum fluctuations by (we always deal with the *vacuum* fluctuations, and often omit the corresponding notation)

$$\langle \varphi^2(x) \rangle = \int_0^\infty \mathcal{P}_\varphi(q) \frac{dq}{q}, \quad (13.5)$$

where the integration is performed over the absolute value of momentum.² As usual, the power spectrum determines the contribution of a decimal interval of momenta

²The integral (13.5) has the ultraviolet divergency. It is due to the fact that the two operators φ in the left hand side of (13.5) are taken at coincident points. This does not prevent us to calculate the power spectrum by using (13.5). Various physical quantities, such as correlators at different points, are finite in free field theory.

into the fluctuation, $\langle \varphi^2(x) \rangle$ in our case. The left hand side of the last formula is straightforwardly calculated by using (13.3),

$$\langle \varphi^2(x) \rangle = \int \frac{d^3q}{(2\pi)^3 \cdot 2\omega_q} = \int_0^\infty \frac{4\pi q^2 dq}{2(2\pi)^3 q} = \int_0^\infty \frac{q^2}{(2\pi)^2} \frac{dq}{q}. \quad (13.6)$$

Hence,

$$\mathcal{P}_\varphi(q) = \frac{q^2}{(2\pi)^2}. \quad (13.7)$$

We *define* the amplitude of quantum fluctuations $\delta\varphi_q$ as the variance of the field with typical momentum q , namely

$$\delta\varphi_q \equiv \Delta_\varphi = \sqrt{\mathcal{P}_\varphi(q)}. \quad (13.8)$$

We see that for free massless scalar field it is equal to

$$\delta\varphi_q = \frac{q}{2\pi}. \quad (13.9)$$

This is the precise meaning of the estimate (13.2); we have also found the numerical coefficient.

In the expanding Universe, physical momenta q decrease in time (get redshifted). If the expansion is slow, the amplitude of vacuum fluctuations of a given conformal momentum decreases too. This is not the case in the inflating Universe.

13.1.2 Generation of inflaton perturbations

We now study the properties of the inflaton fluctuations at inflation. We write the total inflaton field as the sum of the homogeneous background field $\phi_c(t)$ and perturbation $\varphi(\mathbf{x}, t)$,

$$\phi(\mathbf{x}, t) = \phi_c(t) + \varphi(\mathbf{x}, t). \quad (13.10)$$

Here $\phi_c(t)$ is the homogeneous *classical* solution to the field equations; e.g., it is one of the solutions studied in Chapter 12 (we denoted it simply by $\phi(t)$; here we added the superscript c to avoid confusion); $\varphi(\mathbf{x}, t)$ is quantum field. We work within the linearized theory, assuming that the perturbation $\varphi(\mathbf{x}, t)$ is small (in the sense which will become clear later). The quadratic action for $\varphi(\mathbf{x}, t)$ in the background of the spatially flat metric is obtained by inserting the expansion (13.10) into the general expression for the action (12.3). We find

$$\begin{aligned} S_\varphi &= \frac{1}{2} \int d^4x \sqrt{-g} [g^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi - V''(\phi_c) \varphi^2] \\ &= \frac{1}{2} \int dt d^3x a^3 [\dot{\varphi}^2 - a^{-2} (\partial_i \varphi)^2 - V''(\phi_c) \varphi^2]. \end{aligned} \quad (13.11)$$

The linearized field equation reads,

$$\ddot{\varphi} + 3H\dot{\varphi} - \frac{1}{a^2}\partial_i\partial_i\varphi + V''(\phi_c)\varphi = 0.$$

Since the background is homogeneous and isotropic, the solution is naturally the sum of plane waves,

$$\varphi(\mathbf{x}, t) \propto e^{\pm i\mathbf{kx}}\varphi(\mathbf{k}, t),$$

where \mathbf{k} is the time-independent conformal momentum. The field equation in the momentum representation is

$$\ddot{\varphi} + 3H\dot{\varphi} + \frac{k^2}{a^2}\varphi + V''(\phi_c)\varphi = 0. \quad (13.12)$$

Hereafter we omit \mathbf{k} in the notation of the field wherever possible, and denote $k = |\mathbf{k}|$.

The idea of the calculation is as follows. The Hubble parameter H evolves slowly at inflation, while the scale factor a rapidly grows. Hence, the physical momentum $q(t) = k/a(t)$ is large at early times as compared to the Hubble parameter, and also $q^2 \gg V''(\phi_c)$. In this regime, the first and third terms dominate in Eq. (13.12), the field $\varphi(t)$ rapidly oscillates and behaves in very much the same way as in the Minkowski space-time. In other words, the mode $\varphi(\mathbf{k})$ is subhorizon, see Section 2.4; space-time curvature is effectively small and largely irrelevant. The quantum field theory vacuum is well-defined in such a situation; particle creation due to the non-stationary metric is strongly suppressed and vanishes in the formal limit $k/a \rightarrow \infty$. The amplitude of the vacuum fluctuations decreases as momentum gets redshifted. On the other hand, $q(t) \ll H$ at late times, and the second term with the Hubble friction dominates in Eq. (13.12). The mode is superhorizon, and its amplitude practically does not change in time; vacuum fluctuations are “frozen in”, even though their wavelength grows and momentum decreases. The transition from sub- to superhorizon regime occurs when the physical momentum of a given mode is of the order of the Hubble parameter,

$$q(t) \sim H.$$

According to (13.9), the vacuum fluctuation amplitude at that time is

$$\delta\varphi_k \equiv \Delta_\varphi \sim \frac{H}{2\pi}, \quad (13.13)$$

and it stays constant in time afterwards. This means that the large wavelength (superhorizon) inflaton field fluctuations are enhanced with respect to the fluctuations of the same wavelength in the Minkowski space-time by the factor

$$\frac{\delta\varphi}{\delta\varphi_{Mink}} \sim \frac{H}{q}.$$

Cosmologically interesting wavelengths are, roughly speaking, of the order of the present cosmological horizon; they exit the horizon at inflation when their wavelength is not very much longer than the Planck length (assuming that the Hubble parameter at inflation is not much less than M_{Pl} in logarithmic scale). It then follows from the results of Section 11.1.1 that the horizon exit of these waves occurred at the time corresponding to about $N_e \sim 60$ e -foldings before the end of inflation. The amplitude of perturbations is frozen in since the horizon exit, while the redshift factor for momentum is roughly e^{N_e} . This gives an idea of the enhancement factor for the vacuum fluctuations: by the end of inflation, the amplitude is enhanced, as compared to the Minkowski case, by a huge factor

$$\frac{H}{q(t_e)} \sim e^{N_e} \sim e^{60}.$$

This is the physical mechanism behind the inflationary generation of the inflaton field perturbations, and the density perturbations in the end.

To end the discussion at the qualitative level, we emphasize that the sequence of events for a given mode at inflation is opposite to that at radiation or matter domination. Namely, the mode is first subhorizon, $q(t) \equiv k/a(t) \gg H$, and then superhorizon, $q \ll H$. The situation is inverse at radiation or matter domination. It is worth noting that this property of inflation is directly related to the solution of the Hot Big Bang theory problems: inflation solves these problems precisely because the combination aH grows in time, see Section 11.1, hence $q(t) = k/a(t)$ necessarily decreases faster than $H(t)$. The dependence of the physical momentum and Hubble parameter on time at the inflationary epoch and later is illustrated in Fig. 13.1.

We now turn to the quantitative analysis. We specify to the slow roll inflation; in that case

$$V'' = \frac{8\pi}{M_{Pl}^2} V \cdot \eta = 3H^2 \cdot \eta \ll H^2,$$

where η is the slow roll parameter.³ Therefore, the last term in the left hand side of Eq. (13.12) can be neglected. We introduce the conformal time η , related to the cosmic time t in the standard way, $dt = ad\eta$. We recall for later reference that for the exponential expansion, $a(t) = a_0 \cdot \exp(Ht)$, the scale factor evolves in the conformal time as follows (see Section 2.1.1),

$$a(\eta) = -\frac{1}{H\eta}, \quad (13.14)$$

where η is negative and $|\eta|$ decreases in time. We note that this formula is *not exact* at inflation, since the Hubble parameter depends on time, albeit slowly. Nevertheless, it can be used for relatively short time intervals, when the evolution of H is negligible.

³Historically, the conformal time and one of the slow roll parameters are denoted by the same letter η . We also use this notation, and explicitly specify the meaning of η whenever necessary.

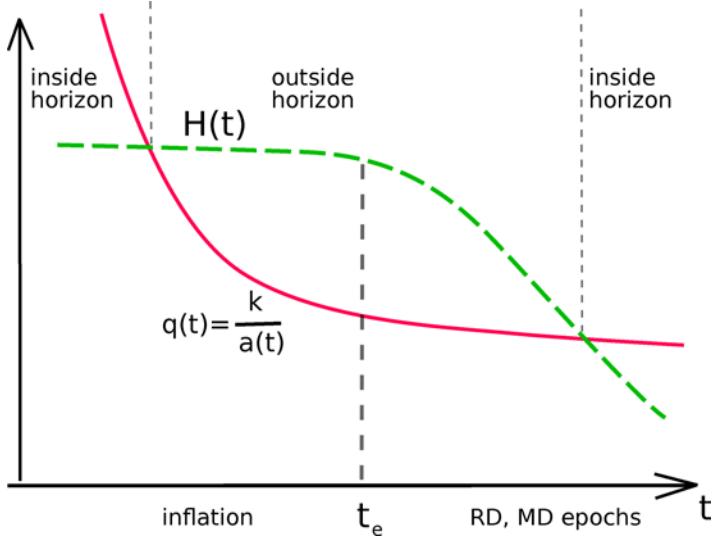


Fig. 13.1 The dependence of the physical momentum and Hubble parameter on time at inflation, radiation and matter domination for a mode of a given conformal momentum. The mode is subhorizon and superhorizon at $q \gg H$ and $q \ll H$, respectively.

We recall the form of metric in conformal time, (see Section 2.1.1), $ds^2 = a^2(\eta)\eta_{\mu\nu}dx^\mu dx^\nu$, and, neglecting the term with V'' , write the action (13.11) in the following form,

$$S_\varphi = \frac{1}{2} \int d^4x a^2(\eta) [(\partial_\eta \varphi)^2 - (\partial_i \varphi)^2]. \quad (13.15)$$

Thus, our simplified problem boils down to the study of a massless scalar field in the background metric of the inflating Universe. We keep identifying φ with the inflaton perturbation field, but our result for its spectrum is valid for any massless scalar field minimally coupled to gravity.

The equation for the inflaton perturbation is

$$\varphi'' + 2\frac{a'}{a}\varphi' - \Delta\varphi = 0, \quad (13.16)$$

where, as usual, $\Delta = \partial_i \partial_i$ and prime denotes the derivative with respect to η . It is convenient to introduce a new field χ by

$$\varphi = \frac{\chi}{a(\eta)}. \quad (13.17)$$

Inserting this into (13.15) we obtain the action for χ ,

$$\begin{aligned} S_\chi &= \frac{1}{2} \int d^3x d\eta \left[\left(\chi' - \frac{a'}{a}\chi \right)^2 - (\partial_i \chi)^2 \right] \\ &= \frac{1}{2} \int d^3x d\eta \left[\chi'^2 - (\partial_i \chi)^2 + \frac{a''}{a}\chi^2 \right], \end{aligned} \quad (13.18)$$

where the second form is obtained by integrating by parts. We write the equation for χ in the momentum representation,

$$\chi'' - \frac{a''}{a}\chi + k^2\chi = 0. \quad (13.19)$$

The combination a''/a is small compared to k^2 at early times, so the corresponding terms can be neglected both in Eq. (13.19) and in the action functional (13.18). Formally, these terms vanish in the limit $\eta \rightarrow -\infty$ (e.g., $a''/a = 2/\eta^2$ for exponential expansion); the mode is subhorizon. Therefore, the early time behavior of the field χ in the conformal coordinates is the same as that of a massless scalar field in the Minkowski space-time. We immediately write the expression for the field operator, completely analogous to (13.3):

$$\chi(\mathbf{x}, \eta) = \int \frac{d^3k}{(2\pi)^{3/2}\sqrt{2k}} (e^{ik\eta - i\mathbf{kx}} A_{\mathbf{k}}^\dagger + e^{-ik\eta + i\mathbf{kx}} A_{\mathbf{k}}), \quad \eta \rightarrow -\infty, \quad (13.20)$$

where the creation and annihilation operators $A_{\mathbf{k}}^\dagger$ and $A_{\mathbf{k}}$ obey the standard commutational relation (13.4) with \mathbf{q} replaced by \mathbf{k} .

Physically reasonable choice of the initial state of the quantum inflaton field φ for interesting momenta \mathbf{k} is vacuum, $A_{\mathbf{k}}|0\rangle = 0$. Indeed, modes of interest to us exit the horizon at $N_e(k) \sim 60$ e -foldings before the end of inflation, so in the beginning of inflation their physical momenta are

$$q(t_i) \sim H e^{N_e^{tot} - N_e(k)},$$

Generally, they are extremely high. Hence, the deviation from vacuum at the beginning of inflation would mean that there existed particles with huge physical momenta; such an assumption would be very contrived. Hence, we consider the field φ in the initial vacuum state with respect to the operators $A_{\mathbf{k}}^\dagger$ and $A_{\mathbf{k}}$. We emphasize that we work in the Heisenberg formulation, so this state does not evolve. In the context of quantum field theory in the de Sitter space, this vacuum state is known as the *Bunch-Davis vacuum*.

We note that according to (13.17), the amplitudes of the modes of the field φ decrease in time in the subhorizon regime as a^{-1} , i.e., in the same way as the physical momentum. This is in accordance with the above qualitative analysis.

On the other hand, the modes are superhorizon at late times at the inflationary epoch, and the term with k^2 is negligible in Eq. (13.19). The field χ has the growing mode in this regime,

$$\chi = \text{const} \cdot a(\eta), \quad (13.21)$$

so that the field φ is time-independent. There is also the decaying solution,

$$\chi_{dec}(\eta) = a(\eta) \int d\eta \frac{\text{const}}{a^2}. \quad (13.22)$$

It is not interesting in our context, since this solution rapidly decreases and soon after the horizon exit becomes negligibly small.

So, the operator χ grows as the scale factor, and φ is a constant operator at late times. We need the expression for this constant operator that determines the power spectrum of the inflaton perturbations at inflation. To obtain this expression, we match the solutions (13.20) and (13.21) by solving Eq. (13.19) in the time interval of a few Hubble times around the horizon exit. This time interval is short in the sense that we can use the formula (13.14) for the scale factor. Then Eq. (13.19) becomes

$$\chi'' - \frac{2}{\eta^2} \chi + k^2 \chi = 0.$$

Its solution which is positive-frequency in the formal limit $\eta \rightarrow -\infty$ has the form

$$\chi_k^{(+)} = e^{ik\eta} \left(1 + \frac{i}{k\eta} \right), \quad (13.23)$$

while the negative-frequency solution is

$$\chi_k^{(-)} = [\chi_k^{(+)}]^*.$$

As expected, these solutions oscillate with frequency k in the subhorizon regime (early times, $k|\eta| \gg 1$) and behave as $\eta^{-1} \propto a(\eta)$ in the superhorizon regime (late times, $k|\eta| \ll 1$). We require that the field is given by (13.20) before the horizon exit and write near the horizon exit

$$\chi(\mathbf{x}, \eta)_{\{k\}} = \int_{\{k\}} \frac{d^3 k}{(2\pi)^{3/2} \sqrt{2k}} (e^{-i\mathbf{k}\mathbf{x}} \chi_k^{(+)}(\eta) A_{\mathbf{k}}^\dagger + e^{i\mathbf{k}\mathbf{x}} \chi_k^{(-)}(\eta) A_{\mathbf{k}}), \quad (13.24)$$

where the subscript $\{k\}$ is used to emphasize that we are interested in modes of momenta around a given momentum k . After the horizon exit, i.e., at $|\eta| \ll k^{-1}$, the field (13.24) becomes

$$\chi(\mathbf{x}, \eta)_{\{k\}} = \int_{\{k\}} \frac{d^3 k}{(2\pi)^{3/2} \sqrt{2k}} \left(-\frac{1}{k\eta} \right) (e^{-i\mathbf{k}\mathbf{x}+i\alpha_k} A_{\mathbf{k}}^\dagger + e^{i\mathbf{k}\mathbf{x}-i\alpha_k} A_{\mathbf{k}}),$$

where α_k are irrelevant phases. Making use of (13.17) and (13.14), we obtain the inflaton perturbation φ after the horizon exit,

$$\varphi(\mathbf{x}, \eta)_{\{k\}} = \int_{\{k\}} \frac{d^3 k}{(2\pi)^{3/2} \sqrt{2k}} \frac{H}{k} (e^{-i\mathbf{k}\mathbf{x}+i\alpha_k} A_{\mathbf{k}}^\dagger + e^{i\mathbf{k}\mathbf{x}-i\alpha_k} A_{\mathbf{k}}). \quad (13.25)$$

We see that this field indeed does not depend on time. It is the Gaussian random field (the Wick theorem is obviously valid, see also Section C.2), and the contribution of momenta of order k to its fluctuation is

$$\langle \varphi(\mathbf{x})^2 \rangle_{\{k\}} = \int_{\{k\}} \frac{d^3 k}{(2\pi)^3} \frac{H^2}{2k^3} = \int_{\{k\}} \frac{dk}{k} \frac{H^2}{(2\pi)^2}. \quad (13.26)$$

This gives the power spectrum

$$\mathcal{P}_\varphi(k) = \frac{H_k^2}{(2\pi)^2} \quad (13.27)$$

and the amplitude of the inflaton perturbations⁴

$$\delta\phi_k \equiv \Delta_\varphi(k) = \frac{H_k}{2\pi}. \quad (13.28)$$

The notation in the right hand side emphasizes the following property. Modes of different momenta exit the horizon at different time, the larger k the later. The Hubble parameter at inflation evolves in time, albeit slowly. The above analysis, including the formula (13.23) and formulas that follow, refers to the time interval near the time η_k at which the mode of momentum k exits the horizon; the mode of the field φ stays constant afterwards. Therefore, the parameter H entering (13.25), and hence (13.27) and (13.28), is actually the Hubble parameter *at the horizon exit*,

$$H_k = H(\eta_k).$$

The time η_k is determined by the condition that the physical momentum is equal to the Hubble parameter,

$$H(\eta_k) = \frac{k}{a(\eta_k)}. \quad (13.29)$$

We see that the right hand sides of (13.27) and (13.28) weakly depend on k , i.e., on the perturbation wavelength. We note that Eq. (13.29) is generally an order-of-magnitude estimate, but since the Hubble parameter evolves slowly, it gives the desired result to the leading order in the slow roll parameters.

If one neglects the weak dependence of the inflationary Hubble parameter on time, i.e., sets $H_k = H = \text{const}$, one obtains flat power spectrum (13.27) of the superhorizon inflaton perturbations at inflation. As we argue in Section 13.1.3, this yields the flat spectrum of the primordial scalar perturbations. The slow variation of the Hubble parameter and the background inflaton field in time results in the weak dependence of the power spectrum on momentum, in particular, small scalar tilt. This is one of the predictions of the inflationary theory.

A comment is in order. Let us try to use the result (13.26) for calculating the total variance of the inflaton perturbation. We write

$$\langle \varphi^2(\mathbf{x}) \rangle = \int \frac{dk}{k} \frac{H^2}{(2\pi)^2}. \quad (13.30)$$

This integral is formally divergent at the lower limit of integration⁵ $k \rightarrow 0$ (the upper limit of integration is in fact $k_{max} \sim aH$, since modes of higher momenta are subhorizon and the result (13.26) is not valid for them). This is logarithmic

⁴The fact that the numerical coefficient in the estimate (13.13) turns out to be correct appears a mere coincidence.

⁵Note that this divergency is unrelated to the fact that the left hand side contains the product of the field operators at coincident points. The corresponding divergence exists as well, but, like in the Minkowski case, it is an ultraviolet property. We are interested in the infrared behavior, which is specific to inflation.

divergency, so the potentially large contribution to $\langle \varphi^2(\mathbf{x}) \rangle$ comes from *exponentially low* momenta. Modes of very low momenta are homogeneous not only over the horizon at inflation, but also over the present horizon. They exit the horizon at very early stage of the inflationary epoch. So, one can interpret their contribution, for all practical purposes, as the renormalization of the initial value of the classical inflaton field ϕ_c . At the same time, modes of super-large wavelengths can make the Universe strongly inhomogeneous at the length scales much exceeding the present horizon size. We will have to say more about this in Section 14.1.

13.1.3 Primordial scalar perturbations

We continue our simplified analysis and study primordial scalar perturbations generated by the inflaton fluctuations. The origin of the density perturbations can be understood in terms of the following qualitative picture (cf. Section 5.2). Different regions of the Universe of sizes exceeding the current Hubble size H^{-1} evolve independently, as different homogeneous isotropic universes. The existence of the inflaton perturbations of physical wavelengths greater than H^{-1} means that the total (perturbations included) inflaton field in these different local universes takes somewhat different values,⁶ while it is practically homogeneous in each of these local universes. Hence, the local universes are at somewhat different stages of their evolution: the region with the total field $\phi_c + \delta\phi$ greater than the average over the entire Universe field ϕ_c lags behind the average evolution, and vice versa,⁷ see Fig. 13.2.

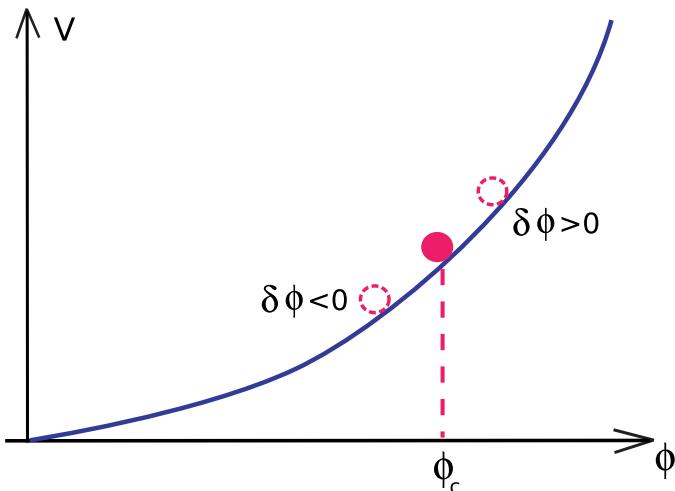


Fig. 13.2 The inflaton field, perturbations included, in different Hubble size regions.

⁶This picture is valid in an appropriate gauge only, but we ignore this subtlety in our qualitative discussion.

⁷We assume here for definiteness that the field ϕ decreases in time at inflation.

The time delay/advance δt is obtained by writing

$$\dot{\phi}_c \delta t = \delta \phi. \quad (13.31)$$

Because of this delay/advance, the local universes exit the inflationary regime at somewhat different times, and hence they have somewhat different energy densities: the energy density in the regions where inflation finished earlier is diluted by the time inflation ends on average. This yields the following estimate for the density perturbation,

$$\delta \rho \sim \dot{\rho} \delta t.$$

At the end of inflation one has

$$\dot{\rho} \sim -H\rho,$$

so (13.31) gives the estimate for the amplitude

$$\frac{\delta \rho}{\rho} \sim \frac{H}{\dot{\phi}_c} \delta \phi.$$

This argument gives not only the estimate for the amplitude of density perturbations, but also tells that the density perturbation field is proportional to the inflaton perturbation field,

$$\frac{\delta \rho(\mathbf{x})}{\rho} = \text{const} \cdot \frac{H}{\dot{\phi}_c} \varphi(\mathbf{x}). \quad (13.32)$$

Hence, the density perturbation field is Gaussian random, and its power spectrum is determined mostly by the power spectrum of inflaton perturbations.

Our argument is, of course, rather hand-waving. It does not give the numerical coefficient in (13.32). It is even more important that it does not tell at which moment of time one should evaluate the parameters H and $\dot{\phi}_c$ entering (13.32). Furthermore, the quantity $\delta \rho/\rho$ itself is inappropriate, since it is gauge-dependent, and in the conformal Newtonian gauge it depends on time even for superhorizon modes. We give the accurate analysis in Section 13.2 and here we simply quote the result. First, the appropriate characteristic of a superhorizon mode is \mathcal{R} , the spatial curvature in the comoving reference frame, see Section 5.2. Second, the coefficient in the result (13.32) written for \mathcal{R} is equal to 1, and the parameters H and $\dot{\phi}_c$ should be evaluated at the time η_k the mode exits the horizon at inflation. This time is determined by Eq. (13.29). Hence, the amplitude and power spectrum of the scalar perturbations are

$$\Delta_{\mathcal{R}}(k) = \left(\frac{H}{|\dot{\phi}_c|} \right)_{\eta_k} \Delta_{\varphi}(k) = \left(\frac{H^2}{2\pi|\dot{\phi}_c|} \right)_{\eta_k}, \quad (13.33)$$

$$\mathcal{P}_{\mathcal{R}}(k) = \left(\frac{H^2}{2\pi\dot{\phi}_c} \right)_{\eta_k}^2. \quad (13.34)$$

Because of slow roll, the time-dependence of H and $\dot{\phi}_c$, and hence the momentum-dependence in (13.33) are weak, and the spectrum is close to the flat, Harrison–Zeldovich spectrum.

Instead of the time η_k , it is convenient to consider the number of e -foldings $N_e(k)$ from that time to the end of inflation. The estimates for the latter are similar to the estimates given in Chapter 11. Namely, the present momentum of a wave is $q_0 = k/a_0$, so Eq. (13.29) is written as follows,

$$q_0(k) \frac{a_0}{a_e} e^{N_e(k)} = H(\eta_k). \quad (13.35)$$

To estimate the ratio of the scale factors here, we assume that *at the post-inflationary reheating epoch* the energy density behaves as $\rho(a) \propto a^{-\beta}$, where $\beta = \text{const}$. It follows from the results of Chapter 15 that β actually evolves from $\beta = 3$ (non-relativistic matter, coherent inflaton oscillations) to $\beta = 4$ (relativistic matter), but we take it time-independent for the estimate. Then

$$\frac{a_0}{a_e} = \frac{a_0}{a_{reh}} \frac{a_{reh}}{a_e} \simeq \frac{T_{reh}}{T_0} \left(\frac{\rho_e}{\rho_{reh}} \right)^{1/\beta},$$

where we neglected the change in the effective number of degrees of freedom at the hot stage, which is unimportant in what follows. We use the Friedmann equation, the relation $\rho_{reh} \simeq T_{reh}^4$, neglect factors of order 1 and write Eq. (13.35) as follows,

$$e^{N_e(k)} \simeq \frac{T_0}{q_0} \frac{T_{reh}^{4/\beta-1} H_e^{1-2/\beta}}{M_{Pl}^{2/\beta}} \frac{H(\eta_k)}{H_e}. \quad (13.36)$$

Since $3 \leq \beta < 4$, this formula shows that $N_e(k)$ decreases as the overall inflationary energy scale and reheat temperature decrease. It is conveniently written by introducing the parameter $T_e \simeq \sqrt{M_{Pl} H_e}$, which has the meaning of the maximum temperature at instantaneous reheating. Then (13.36) gives⁸

$$N_e(k) \simeq \log \frac{T_0}{q_0} - \left(\frac{4}{\beta} - 1 \right) \log \frac{T_e}{T_{reh}} - \frac{1}{2} \log \frac{M_{Pl}}{H_e} + \log \frac{H(\eta_k)}{H_e}. \quad (13.37)$$

The third term here determines the general dependence of $N_e(k)$ on the energy scale of inflation, whereas the second one encodes the ambiguity due to unknown reheating processes. Cosmologically interesting scales vary from hundreds kpc to the present Hubble scale of order 10 Gpc. The number of e -foldings depends on

⁸Note that the result (13.37), applied to the present Hubble wavelength, $k/a_0 \sim H_0$, is consistent with (11.15). Indeed, when obtaining the estimate (11.15) we assumed instantaneous reheating, took the Hubble parameter at the beginning of inflation equal to M_{Pl} , and required, in fact, that the present Universe be subhorizon at the beginning of inflation. In notations used here, all this means that we took $T_e = T_{reh}$, $H_e \sim T_{reh}^2/M_{Pl}$, $H(\eta_k) \sim M_{Pl}$ and $q_0 \sim H_0$. With this choice, the result (13.37) indeed reduces to (11.15).

momentum weakly (logarithmically), so we can take a certain value of momentum for an estimate, say

$$q_0 = 0.002 \text{ Mpc}^{-1}. \quad (13.38)$$

The largest term in the argument of logarithm in (13.37) is the ratio

$$\frac{T_0}{q_0} \simeq 10^{28}.$$

To estimate other terms in (13.37) consider, e.g., large field inflation with the potential $V(\phi) = m^2\phi^2/2$. In that case $H_e \sim m\phi_e/M_{Pl} \sim m \sim 10^{-6}M_{Pl}$, see (13.62), $T_e \sim 10^{-3}M_{Pl}$, while the reheat temperature may be anything in the interval, say, $T_{reh} = 10^8 - 10^{16} \text{ GeV}$. The last term in (13.37), is rather small, and we neglect it. Taking $3 \leq \beta \leq 4$ we obtain the estimate $N_e = 50 - 57$, depending on the reheat temperature. Similar values are found in other large field models. In the context of large field inflation, one traditionally uses the interval

$$N_e \simeq 50 - 60.$$

Hence, the formula (13.33) involves the parameters of the inflationary epoch at $50 - 60$ e -foldings before its end.

To end this Section, we make a few comments. First, to an excellent approximation there is no decaying mode of the inflaton perturbations by the end of inflation. According to (13.22), the amplitude of the decaying mode $\phi_{dec} = \chi_{dec}/a$ decreases by a factor $\exp(-N_e)$ in the last $N_e \sim 60$ e -foldings of inflation. Consequently, there is no decaying mode in the scalar perturbations. Second, the mechanism we outlined generates *adiabatic* scalar perturbations, since, by the end of inflation, all energy density is carried by the inflaton field, and the mechanisms that reprocess it into heat are the same everywhere in the Universe. The generation of isocurvature modes requires additional mechanisms, see Sections 14.2 and 14.4 in this regard. Finally, perturbations are Gaussian provided that the non-linearity in the equation for inflaton perturbations can be neglected, as we have done above. We argue in Section 13.5 that the non-linear terms are indeed very small, so the perturbations are indeed Gaussian to a very good approximation.

We discuss the consequences of the result (13.33) in Section 13.4. Before that, we study scalar perturbations rigorously, and then consider another important effect, the generation of tensor perturbations.

13.2 Scalar Perturbations in Full Linear Theory

The purpose of this Section is to give the accurate treatment of the generation of the scalar perturbations at inflation. We obtain the exact linearized equation that describes the inflaton and metric perturbations, and derive, in the slow roll approximation, the formula (13.33) for the quantity \mathcal{R} introduced in Section 5.2.

We recall that \mathcal{R} is the spatial curvature of constant time hypersurfaces in the comoving reference frame; in the course of our analysis we will have to clarify the notion of the comoving frame, i.e., specify the quantity that plays the role of velocity at inflation.

We use the conformal Newtonian gauge, as we did in Chapters 2, 4 and 6. The anisotropic stress is absent (see below), so the perturbed metric (2.69) reads

$$ds^2 = a^2(\eta)[(1 + 2\Phi)d\eta^2 - (1 - 2\Phi)d\mathbf{x}^2]. \quad (13.39)$$

We begin with the calculation of the energy-momentum tensor of the inflaton field. According to (12.4), the complete expression for the energy momentum tensor with one upper and one lower indices is

$$T_\nu^\mu = g^{\mu\lambda}\partial_\nu\phi\partial_\lambda\phi - \delta_\nu^\mu\left(\frac{1}{2}g^{\lambda\rho}\partial_\lambda\phi\partial_\rho\phi - V(\phi)\right). \quad (13.40)$$

We insert here the inflaton field in the form (13.10) (we still use the notation ϕ_c for background inflaton field) and the expression for the metric (13.39). In this way we obtain the linearized form of the components we are going to use,

$$\begin{aligned} \delta T_0^0 &\equiv \delta\rho = \frac{1}{a^2}[-\Phi\phi_c'^2 + \phi_c'\varphi'] + \frac{dV(\phi_c)}{d\phi_c}\varphi \\ &= \frac{1}{a^2}\left[-\Phi\phi_c'^2 + \phi_c'\varphi' - \left(\phi_c'' + 2\frac{a'}{a}\phi_c'\right)\varphi\right], \end{aligned} \quad (13.41)$$

$$\delta T_i^0 = \frac{1}{a^2}\phi_c'\partial_i\varphi, \quad (13.42)$$

where we have recalled the field equation for the background inflaton field, written in conformal time,

$$\phi_c'' + 2\frac{a'}{a}\phi_c' + a^2\frac{dV(\phi_c)}{d\phi_c} = 0.$$

It follows from the general form (13.40) and spatial homogeneity of ϕ_c that $\delta T_j^i \propto \delta_j^i$ to the linear order, i.e., the anisotropic stress indeed vanishes. We do not need the expression for δT_j^i in what follows.

Now, we recall that the velocity potential is defined by $\delta T_i^0 = -(\rho + p)\partial_i v$, see (2.39) and (2.68). It follows from (12.6) and (12.7) that the unperturbed energy density and pressure obey

$$\rho + p = \frac{1}{a^2}\phi_c'^2.$$

Therefore, we find from (13.42) the effective velocity potential in the theory of the inflaton field,

$$v = -\frac{\varphi}{\phi_c'}. \quad (13.43)$$

Hence, the comoving coordinate frame is the frame in which the inflaton field is constant in space, i.e., $\varphi = 0$; it is in this frame that the effective velocity vanishes.

We now write the relevant equations for the scalar perturbations, namely, 00- and $0i$ -components of the linearized Einstein equations, (2.74) and (2.75). These are

$$\Delta\Phi - 3\frac{a'}{a}\Phi' - 3\frac{a'^2}{a^2}\Phi = -4\pi G\phi_c'^2\Phi + 4\pi G\left[\phi_c'\varphi' - \left(\phi_c'' + 2\frac{a'}{a}\phi_c'\right)\varphi\right], \quad (13.44)$$

$$\Phi' + \frac{a'}{a}\Phi = 4\pi G\phi_c'\varphi. \quad (13.45)$$

These two equations are sufficient for studying the evolution of the two unknown functions, Φ and φ . This choice of the pair of equations is convenient due to the fact that the second time-derivatives do not enter here.

Problem 13.1. Show that vector perturbations are not generated in the model under discussion.

Problem 13.2. Show that ij -components of the linearized Einstein equations are identically satisfied provided that Eqs. (13.44) and (13.45) are solved. Write the linearized equation of motion that follows from the variation of the action with respect to ϕ . Show that it is also identically satisfied for solutions to Eqs. (13.44) and (13.45).

Hence, the equations not written in the text do not give anything new compared to the system (13.44), (13.45).

We have not yet used the fact that $a(\eta)$ and $\phi_c(\eta)$ obey the Einstein equations. Let us now incorporate this fact. Namely, we use Eqs. (2.4) and (2.5) and write

$$\frac{a''}{a} - 2\frac{a'^2}{a^2} = -4\pi G(\rho + p)a^2 = -4\pi G\phi_c'^2. \quad (13.46)$$

This enables us to rewrite Eq. (13.44) in the following way,

$$\Delta\Phi - 3\frac{a'}{a}\Phi' - \left(\frac{a''}{a} + \frac{a'^2}{a^2}\right)\Phi = 4\pi G\left[\phi_c'\varphi' - \left(\phi_c'' + 2\frac{a'}{a}\phi_c'\right)\varphi\right]. \quad (13.47)$$

The system (13.45), (13.47) can be simplified. We express Φ through Φ' and ϕ_c using (13.45) and insert it into Eq. (13.47). We again make use of (13.46) and cast the resulting equation into the form

$$\Delta\Phi = 4\pi G\frac{a}{a'}\phi_c'^2 \cdot \frac{d}{d\eta} \left(\Phi + \frac{a'}{a\phi_c'}\varphi\right). \quad (13.48)$$

Note that according to the definition (5.16) and expression (13.43) for the effective velocity potential, the quantity in parenthesis in (13.48) is nothing but $-\mathcal{R}$, since

$$\mathcal{R} = -\Phi + \frac{a'}{a}v = -\left(\Phi + \frac{a'}{a\phi_c'}\varphi\right).$$

As we said before, \mathcal{R} has the meaning of spatial curvature of constant inflaton field hypersurfaces.

It is clear from (13.48) that instead of φ , the natural variable is

$$\tilde{\varphi} = \varphi + \frac{a\phi'_c}{a'}\Phi. \quad (13.49)$$

By analogy to the field χ defined by (13.17), it is convenient to introduce the field u ,

$$\tilde{\varphi} = \frac{u}{a},$$

so that

$$u = -z\mathcal{R}, \quad (13.50)$$

where

$$z = \frac{a^2\phi'_c}{a'}.$$

The function u is called in literature the *Mukhanov–Sasaki variable*. Equation (13.48) is written in its terms as follows,

$$\Delta\Phi = 4\pi G\phi'_c \frac{z}{a} \cdot \frac{d}{d\eta} \left(\frac{u}{z} \right). \quad (13.51)$$

The second equation for Φ and u is Eq. (13.45),

$$\frac{a'}{a^2} \frac{d}{d\eta} \left(\frac{a^3}{a'} \Phi \right) = 4\pi G\phi'_c u. \quad (13.52)$$

The system (13.51), (13.52) can be reduced to the second order equation for u . We apply the operator Δ to Eq. (13.52), use Eq. (13.51) and obtain

$$\frac{d}{d\eta} \left[z^2 \frac{d}{d\eta} \left(\frac{u}{z} \right) \right] = z\Delta u,$$

or, finally,

$$u'' - \frac{z''}{z}u - \Delta u = 0. \quad (13.53)$$

We emphasize that this is the *exact* equation in a theory with one inflaton field and no other matter. The slow roll approximation has *not* been used in its derivation; neither have we used any assumptions about the shape of the inflaton potential.⁹ We note in this regard that in the superhorizon regime the last term in Eq. (13.53) can be neglected, and the growing solution is

$$u = \text{const} \cdot z.$$

⁹Equation of the type (13.53) is valid in the general case of scalar perturbations in the expanding Universe filled with single-component fluid [120, 121]. Its generalization to the more complicated kinetic term of the scalar field is given in the book [123].

In view of (13.50), the function \mathcal{R} is time-independent for this solution, which reiterates the general result obtained in Section 5.2.

Let us now turn to the slow roll case. One has to keep in mind that despite the slow evolution of ϕ_c in time t , the derivative $\dot{\phi}_c = a\dot{\phi}_c$ is not small and rapidly changes in time due to the rapid evolution of the scale factor. So, the function $z(\eta)$ is conveniently written as

$$z = a \frac{\dot{\phi}_c}{H}. \quad (13.54)$$

This expression involves the rapidly varying factor $a(t)$, while $\dot{\phi}_c$ and H are time-independent to the leading order in the slow roll parameters. Hence, to the leading order, Eq. (13.53) coincides with the equation we have already encountered, see Eq. (13.19),

$$u'' - \frac{a''}{a}u - \Delta u = 0, \quad (13.55)$$

and in the slow roll approximation

$$a = -\frac{1}{H\eta}. \quad (13.56)$$

At early times at inflationary epoch, a mode of a given momentum k is subhorizon, and the field u coincides with the field $\chi = a\varphi$. Indeed, it follows from (13.51) that $\Phi \propto k^{-1}u$ as $k \rightarrow \infty$, so the contribution of the gravitational potential Φ to the right hand side of (13.49) is negligible, and $\tilde{\varphi} = \varphi$. Hence, the quantum field u is given precisely by (13.20) at large negative η .

Problem 13.3. *Find the solution $\Phi(\eta)$ and $\varphi(\eta)$ to Eqs. (13.51), (13.52) (in the momentum representation) at large negative η (subhorizon regime).*

The observation we have made is, in fact, unnecessary for obtaining the quantum field u at large negative η . The field u can be quantized directly. To this end, one finds the expression for its action. The variation of the action must yield Eq. (13.53), hence the action is

$$S_u = \frac{c}{2} \int d\eta d^3x \left[u'^2 + \frac{z''}{z}u^2 - (\partial_i u)^2 \right], \quad (13.57)$$

where c is an unknown constant. This expression should of course follow from the original quadratic action of the linearized theory (the sum of the quadratic parts of the Einstein–Hilbert and scalar field actions) upon “integrating out” the gravitational potential Φ . This is one way of calculating the coefficient c . This lengthy calculation can be cut short, however. One can require instead that the term χ'^2 enters the action with the coefficient $1/2$, since it originates solely from the scalar field action. Since $u = \chi + z\Phi$, this gives $c = 1$. The quantization of the theory with the action (13.57) proceeds at early times exactly in the same way as in the scalar field theory in the Minkowski space-time.

Thus, in the slow roll approximation, the operator u is precisely the same as the operator χ in Section 13.1.2, and the operator $\tilde{\varphi}$ coincides with the operator χ/a ,

which in the simplified treatment of Section 13.1.2 coincides with the operator of the inflaton field perturbation. The field $\tilde{\varphi}$ stays constant in the superhorizon regime, and this constant is given by the right hand side of (13.25). The power spectrum of the corresponding Gaussian random field is given by (13.27). Finally, formulas (13.50) and (13.54) relate $\tilde{\varphi}$ and \mathcal{R} ,

$$\mathcal{R} = -\frac{H}{\dot{\phi}_c} \tilde{\varphi}.$$

In the slow roll approximation, this yields the formula (13.34) for the power spectrum of \mathcal{R} .

Problem 13.4. *Assuming that the slow roll parameters ϵ and η are small and time-independent, find the analog of Eq. (13.55) to the first non-trivial order in these parameters. [Hint: The formula (13.56) for the scale factor is not valid for $\epsilon \neq 0$.] By solving that equation, find the scalar power spectrum $\mathcal{P}_{\mathcal{R}}$ to the first non-trivial order in ϵ and η . Show that the dependence of ϵ and η on time is the higher order effect.*

13.3 Tensor Perturbations

Tensor perturbations are generated precisely in the same way as the perturbations of a massless scalar field studied in Section (13.1.2). Indeed, we have seen in Section 2.3.1 that tensor mode of a given polarization obeys Eq. (2.58), which coincides with Eq. (13.16) for massless scalar field. The only subtlety is that the action (2.59) for tensor modes differs from the scalar field action (13.15) by the overall factor $1/(32\pi G) \equiv M_{Pl}^2/(32\pi)$. Hence, we introduce the canonically normalized field of tensor perturbations of polarization A ,

$$\varphi^{(A)} = \sqrt{\frac{M_{Pl}^2}{32\pi}} h^{(A)}.$$

Its quantum dynamics is exactly the same as the dynamics of a massless scalar field in the background metric of the inflating Universe, so the superhorizon power spectrum coincides with (13.27), i.e.,

$$\mathcal{P}_{\varphi^{(A)}}(k) = \frac{H_k^2}{(2\pi)^2}.$$

This gives the power spectrum of tensor perturbations,

$$\mathcal{P}_T = \frac{32\pi}{M_{Pl}^2} \sum_{A=+,\times} \mathcal{P}_{\varphi^{(A)}}(k) = \frac{16}{\pi} \frac{H_k^2}{M_{Pl}^2}, \quad (13.58)$$

where we accounted for two polarizations. The dimensionless tensor amplitude is

$$\Delta_T = \sqrt{\mathcal{P}_T} = \frac{4}{\sqrt{\pi}} \frac{H_k}{M_{Pl}}. \quad (13.59)$$

As before, the right hand side involves the Hubble parameter at the time the mode of momentum k exits the horizon at inflation. For present wavelengths comparable to the present Hubble size, this time corresponds to $50 - 60$ e -foldings before the end of inflation. Like inflaton perturbations, primordial tensor perturbations are the Gaussian random field.¹⁰ Its power spectrum is nearly flat.

To end this Section, we make a comment similar to the comment made after the formula (13.30). It has to do with extremely large wavelength modes that exit the horizon early at the inflationary epoch and still remain superhorizon at present. Formally, these modes give logarithmically divergent infrared contribution $h_{ij}^{(IR)}$ to the total metric perturbation h_{ij} . This contribution is so slowly varying in space for all cosmologically interesting scales, that the resulting metric

$$g_{ij} = a^2 (\eta_{ij} + h_{ij}^{(IR)})$$

can be treated as independent of space-time point. One can perform time-independent transformation of spatial coordinates x^i and reduce this metric to the standard form $g_{ij} = a^2 \eta_{ij}$. The classical inflaton field remains homogeneous after this transformation, while the infrared tensor modes do not affect the evolution after they exit the horizon at inflation. Because of that, the analysis of smaller wavelength modes performed in this and previous Sections remains valid. We conclude that the infrared tensor perturbations, like the infrared scalar modes, are irrelevant for describing the visible Universe.

13.4 Amplitudes and Tilts of Power Spectra

Let us discuss in more details the predictions of inflationary models for the power spectra of the primordial scalar and tensor perturbations, and the values of the parameters of the inflaton potentials at which these predictions are consistent with observations. For illustrational purposes we consider three simple classes of models studied in Sections 12.2.1, 12.2.2 and 12.2.3.

13.4.1 *Scalar amplitude: flatness of inflaton potential*

We begin with the amplitude of scalar perturbations. Observational data are consistent with the general prediction of simple inflationary models that the primordial scalar perturbations are Gaussian and are in the adiabatic mode. The

¹⁰We briefly discuss in Section 13.5 to which extent this is correct.

observationally determined value of $A_{\mathcal{R}}$ is given in (5.37), i.e., the amplitude of scalar perturbations is

$$\Delta_{\mathcal{R}} = \sqrt{\mathcal{P}_{\mathcal{R}}} = 5.0 \cdot 10^{-5}. \quad (13.60)$$

In the inflationary theory, this amplitude is given by the expression (13.33). Making use of the slow roll equations (12.11) and (12.12), the result (13.33) can be written as

$$\Delta_{\mathcal{R}} = \frac{3H^3}{2\pi V'} = 4 \left(\frac{8\pi}{3} \right)^{1/2} \frac{V^{3/2}}{M_{Pl}^3 V'}. \quad (13.61)$$

As we have seen in Section 13.1.3, the right hand side of this formula should be evaluated at $N_e = 50 - 60$ e -foldings before the end of inflation.

Let us see that the amplitude (13.61) is determined by the couplings in the inflaton potential, and the small amplitude is obtained for small couplings, i.e., flat potential. As our first example, we consider the large field models with the power-law potentials (12.23). Hereafter we again omit subscript c labeling the classically rolling inflaton field. In the large field models

$$\frac{V^{3/2}}{V'} = \frac{\sqrt{g}}{n} \phi^{n/2+1}.$$

The field ϕ at relevant time is related to the number of e -foldings by (12.33). Making use of (12.33), we obtain the following result for the amplitude (13.61),

$$\Delta_{\mathcal{R}} = \frac{4}{n} \left(\frac{8\pi}{3} \right)^{1/2} \left(\frac{nN_e}{4\pi} \right)^{(n+2)/4} \left(\frac{g}{M_{Pl}^{4-n}} \right)^{1/2}.$$

We see that the coupling g must indeed be small (in the Planck units for $n \neq 4$),

$$\frac{g}{M_{Pl}^{4-n}} = \frac{3}{8\pi} \left(\frac{n}{4} \right)^2 \cdot \left(\frac{4\pi}{nN_e} \right)^{(n+2)/2} \cdot \Delta_{\mathcal{R}}^2.$$

In particular, the parameters of quadratic and quartic potentials (12.28) and (12.29) must be equal to

$$\frac{m}{M_{Pl}} = \sqrt{\frac{3}{4\pi}} \cdot \frac{\pi}{N_e} \cdot \Delta_{\mathcal{R}} \simeq 1.0 \cdot 10^{-6}, \quad (13.62)$$

$$\lambda = \frac{3}{2\pi} \cdot \left(\frac{\pi}{N_e} \right)^3 \cdot \Delta_{\mathcal{R}}^2 \simeq 1.0 \cdot 10^{-13}, \quad (13.63)$$

where we set $N_e = 60$ for numerical estimates. As we already discussed in Section 12.2.1, the inflaton potential must be pretty flat at super-Planckian fields.

Let us turn to the small field inflation of Section 12.2.2 with quartic potential, still assuming that the inequality (12.51) holds. We have in this model

$$\frac{V^{3/2}}{V'} = \frac{V_0^{3/2}}{\lambda \phi^3},$$

and the field ϕ is related to the number of e -foldings by (12.56). Making use of the latter relation, we again find the value (13.63) of the coupling λ .

Problem 13.5. *Find the couplings in the models of Section 12.2.2 with the inflaton potentials $V = V_0 - g\phi^n$ with $n > 2$, assuming that inflation occurs in the regime $V_0 \gg g\phi^n$.*

In the hybrid inflation model, the small scalar amplitude also requires the flat potential, but now in the vicinity of the critical inflaton field ϕ_c . Since one can choose $\phi_c \ll M_{Pl}$ in this model, flatness of the inflaton potential is not so problematic.

Problem 13.6. *Consider the hybrid inflation model of Section 12.2.3 with quadratic inflaton potential (12.59). Assuming that the field ϕ is not much different from its critical value at $N_e \sim 50 - 60$ e -foldings before the end of inflation, show that*

$$\Delta_R \simeq \frac{1}{2\pi\sqrt{3}} \cdot \frac{1}{\eta^{3/2}} \cdot \frac{m}{\phi_c},$$

where η is the slow roll parameter. For sub-Planckian ϕ_c and mild bound $\eta < 0.1$ (in fact, the bound is stronger) this gives the bound on the inflaton mass parameter $m \lesssim 10^{-6} M_{Pl}$.

These examples show that the mechanism of inflation based on General Relativity and the inflaton field minimally coupled to gravity requires the extension of the Standard Model of particle physics. As we noticed in Section 12.2.4, flatness of the inflaton potential is problematic in many realistic extensions. Among other possibilities we note the modification of General Relativity; an example is the Starobinsky model.

Problem 13.7. *Consider the possibility that inflation occurs in the Standard Model with the Higgs field H non-minimally coupled to gravity [124],*

$$\mathcal{L} = \mathcal{L}_{SM} - \frac{M^2}{16\pi} R - \xi H^\dagger H R.$$

Find the value of the dimensionless coupling ξ at which this model predicts the realistic scalar amplitude. Find the scalar and tensor tilts (see Section 13.4.3). Hint: Perform an appropriate conformal transformation of the metric that yields the theory of a minimally coupled scalar field (transformation from the Jordan to the Einstein frame).

13.4.2 Tensor amplitude

We now study the tensor-to-scalar ratio. The result (13.58) gives the tensor power spectrum in the slow roll models,

$$\mathcal{P}_T = \frac{128}{3} \frac{V}{M_{Pl}^4}. \quad (13.64)$$

Hence, the ratio of the tensor and scalar power spectra is

$$r \equiv \frac{\mathcal{P}_T}{\mathcal{P}_{\mathcal{R}}} = \frac{1}{\pi} \frac{M_{Pl}^2 V'^2}{V} = 16\epsilon, \quad (13.65)$$

where ϵ is the slow roll parameter (12.20). We see that the amplitude of the primordial tensor perturbations is fairly large, as compared to that of the scalar perturbations, for not too small ϵ . As an example, it follows from (12.37) that in the large field models,

$$r = \frac{4n}{N_e}.$$

Hence, for the quadratic and quartic inflaton potentials we have

$$\begin{aligned} r &= 0.13 - 0.16, & n = 2, \\ r &= 0.27 - 0.32, & n = 4. \end{aligned}$$

The primordial tensor amplitude in these models is at the level

$$\Delta_T = r^{1/2} \Delta_{\mathcal{R}} = (1.8 - 2.8) \cdot 10^{-5}. \quad (13.66)$$

On the other hand, we have seen in Sections 12.2.2 and 12.2.3 that the parameter ϵ is naturally small in the small field and hybrid models, $\epsilon \ll \eta$. Hence, the tensor amplitudes are small in these models unless their parameters are fine tuned.

Tensor perturbations generated at inflation can be detected by studying CMB temperature anisotropy [116–119]. This effect has not been observed, and the current limit on the tensor-to-scalar ratio is

$$r < 0.20 \quad (13.67)$$

at 95% C.L., see Section 5.4; we get back to this point later on. We see that observations become sensitive to tensor perturbations to the extent that there is already tension between the experiment and the predictions of the large field models with power-law potentials and $n \geq 4$. The sensitivity of CMB observations is expected to increase considerably due to the development of dedicated polarization detectors.

The bound (13.67), combined with the known scalar amplitude (13.60) implies the model-independent bound on the Hubble parameter, and hence inflationary energy scale towards the end of inflation (more precisely, 50 – 60 e -foldings before the end). We find from (13.67) and (13.60) the bound on the tensor power,

$$\mathcal{P}_T < 5.0 \cdot 10^{-10}, \quad \Delta_T < 2.2 \cdot 10^{-5},$$

and using (13.58) and (13.64) we get

$$H < 1 \cdot 10^{-5} M_{Pl}, \quad N_e \sim 50 - 60, \quad (13.68)$$

and

$$E_{inf} \equiv V^{1/4} < 2 \cdot 10^{-3} M_{Pl}, \quad N_e \sim 50 - 60,$$

respectively. Hence, the final stages of inflation definitely occurred at the sub-Planckian energy density. We note that the Hubble parameter and energy scale nearly saturate the above bounds in the large field models.

Before proceeding further, let us discuss the amplitudes of relic gravity waves of relatively small wavelengths. We have seen in Section 3.2.2 that the amplitude of tensor perturbation stays constant until it re-enters the horizon at radiation or matter domination. After the horizon re-entry, the amplitude decreases as a^{-1} . We are interested here in wavelengths smaller than 100 Mpc; these modes enter the horizon at radiation domination, see Section 2.4. The present amplitude of such a wave h_0 is related to the primordial amplitude h by the formula (3.11). Inserting there the numerical values of the parameters, we obtain the present amplitude of gravity waves with wavelength $\lambda_0 = 2\pi/q_0$,

$$\Delta_{T,0}(\lambda_0) \simeq 0.6 \cdot 10^{-18} \left(\frac{\lambda_0}{10^{13} \text{ cm}} \right) \cdot \Delta_T, \quad (13.69)$$

where we set $g_* \simeq 100$, having in mind the waves of wavelengths $\lambda_0 \lesssim 10^{16}$ cm which enter the horizon at temperature $T_x \gtrsim 100$ GeV, see Section 2.4 (the numerical coefficient in (13.69) is slightly larger for larger wavelengths because of the smaller value of g_* at their horizon entry). Making use of (13.66) for the estimate, we find that the large field inflationary models predict the amplitude of relic gravity waves at the level¹¹

$$\Delta_{T,0}(\lambda_0) \simeq (1 - 2) \cdot 10^{-23} \left(\frac{\lambda_0}{10^{13}} \right). \quad (13.70)$$

Recall that we are talking here about stochastic, Gaussian gravitational wave background, and its amplitude $\Delta_{T,0}$ is defined as the square root of its power spectrum $\mathcal{P}_{T,0}$.

We note that the larger value in the right hand side in (13.70) gives the absolute bound in inflationary models without exotica.¹² Indeed, the larger value in (13.70) corresponds to the bound (13.67), and the latter applies to wavelengths of several hundred Mpc and more. Inflationary models without exotica predict red tensor tilt, so the amplitudes of shorter waves cannot exceed the larger value in the right hand side in (13.70).

¹¹In fact, (13.70) overestimates the amplitude by a factor 1.5 – 2. The reason is that the estimates (13.66) are valid for long wavelengths, $\lambda_0 \sim 1$ Gpc, see (13.38). We consider here the perturbations of smaller wavelengths; they exit the horizon at the later stage of inflation, when the Hubble parameter is smaller. Accordingly, the amplitude (13.59) is smaller, see problem 13.9.

¹²Exotic models, such as theories with phantom field playing the role of inflaton or Lorentz-violating theories, sometimes predict larger amplitudes at smaller wavelengths.

The result (13.69) can be used to obtain the energy density of relic gravity waves in the present Universe. The modes of wavelengths well below the present Hubble size do not feel the cosmological expansion at the present epoch, and the action (2.59) reduces to the action for free gravity waves in the Minkowski space-time,

$$S_{TT} = \sum_A \frac{1}{64\pi G} \int d^4x [(\partial_t h^{(A)})^2 - \partial_k h^{(A)} \partial_k h^{(A)}],$$

where derivatives are evaluated over physical spatial coordinates and d^4x is the physical 4-volume element. Hence, the energy functional is

$$E_{TT} = \sum_A \frac{1}{64\pi G} \int d^3x [(\partial_t h^{(A)})^2 + \partial_k h^{(A)} \partial_k h^{(A)}].$$

This immediately gives the energy density of stochastic gravity wave background,

$$\rho_{GW} = \frac{M_{Pl}^2}{32\pi} \int \frac{dq_0}{q_0} (q_0 \Delta_{T,0})^2. \quad (13.71)$$

where the integration is performed over the absolute value of 3-momentum. Making use of (3.11) we find

$$\Omega_{GW} = \frac{1}{12} \Omega_{rad} \int \frac{dq_0}{q_0} \left(\frac{g_{*,0}}{g_*(\eta_\times)} \right)^{1/3} \Delta_T^2. \quad (13.72)$$

This result implies that the contribution of a decimal interval of momenta is

$$\Omega_{GW,log} \simeq 0.03 \Omega_{rad} \Delta_T^2 \simeq 0.9 \cdot 10^{-15} \cdot \left(\frac{\Delta_T}{2 \cdot 10^{-5}} \right)^2. \quad (13.73)$$

The integration interval in (13.72) is finite. The minimal momentum corresponds to the present Hubble wavelength, while the maximum momentum corresponds to modes that exit the horizon at the end of inflation. Hence, in the case of the flat primordial spectrum, the total contribution of all gravity waves to Ω_{GW} is larger than (13.73) by a factor equal to the number of e -foldings, $N_e \simeq 50 - 60$. This contribution is quite small: it is smaller than Ω_γ by at least 9 orders of magnitude. Relic gravity waves are irrelevant from the viewpoint of the cosmological expansion.

Problem 13.8. Obtain the relation (13.71) and the estimate (13.72).

Problem 13.9. Refine the estimate (13.70) in large field models by accounting for the fact that short waves exit the inflationary horizon later than modes of the present wavelength 1 Gpc.

The predicted amplitudes (13.70) are very small, so the relic gravity waves will hardly be directly detected in the near future. However, the situation is not hopeless, since interferometers in outer space may reach the required sensitivity for gravity

waves of wavelengths of order 10^{13} cm (the size of the Earth orbit). It is also not impossible that gravity waves of larger wavelengths, and hence larger amplitudes, may be detected by precision measurements of time variations of pulsar periods (pulsar timing). The detection of relic gravity waves in terrestrial experiments appears impossible, since the wavelengths to be probed are of order $\lambda_0 \sim 10^3$ km, and the predicted gravity wave amplitude at this wavelength is extremely small, $h_0 \lesssim 10^{-28}$. The whole situation is illustrated in Fig. 13.3. The vertical line in this Figure (k_{eq}) separates the modes that entered the horizon at radiation domination (higher frequencies) and matter domination. The former could in principle give sizeable contribution to the energy density at radiation domination, and hence are bounded by Big Bang Nucleosynthesis (“BBN bound”). They could also shift radiation-matter equality and hence affect CMB power spectrum and structure formation; this gives “M/R Eq.” bound. Short dashed lines show the predictions of inflationary models with flat tensor spectrum and various values of the inflation energy scale $E_{\text{inf}} \equiv V_{\text{inf}}^{1/4}$. According to our result (13.72), the predicted energy density is independent of frequency. The contribution of modes that entered the horizon at matter domination grows as frequency decreases. It is bounded by CMB

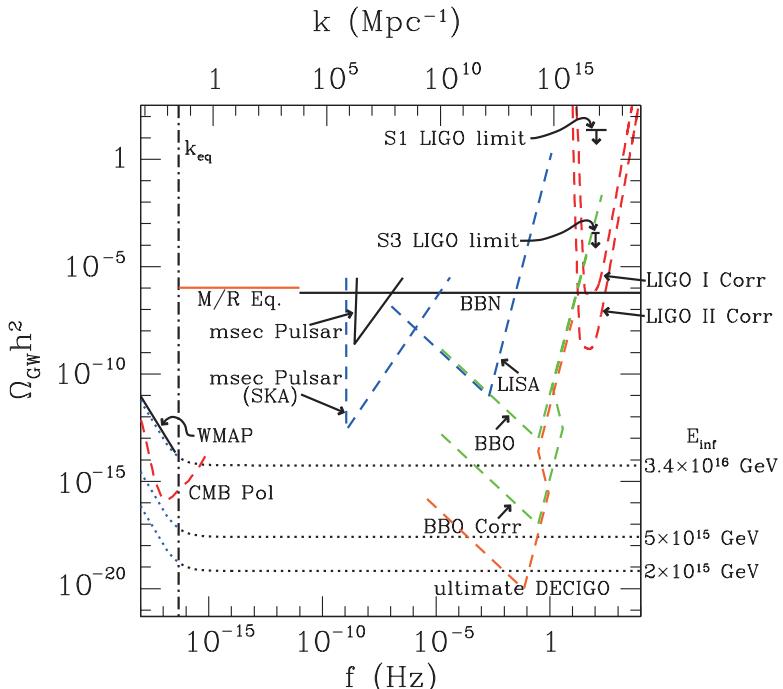


Fig. 13.3 Experimental and cosmological bounds (solid lines), projected sensitivities of planned experiments (long dashed lines) and predictions of the inflationary models with the flat primordial tensor spectrum, in terms of the relative present energy density of gravity waves [125]. Notations near the lines are explained in the text.

temperature anisotropy measurements (“WMAP bound”); the line “CMB Pol” shows the projected sensitivity of dedicated CMB polarization detectors. Direct limits come from existing terrestrial interferometers (“LIGO” are the bounds of two stages of LIGO). Long dashed lines show the projected sensitivities of terrestrial interferometers (“LIGO Corr”) and outer space interferometer projects (“LISA, BBO, DECIGO”). The limit obtained by pulsar timing technique is marked by “msec Pulsar” while projected sensitivity of future pulsar timing array is denoted by “SKA”.

13.4.3 Spectral tilts

Since the perturbations of different wavelengths exit the inflationary horizon at different times, and the inflaton field entering (13.61) and (13.64) evolves in time, scalar and tensor spectra are not exactly flat. Traditionally, scalar and tensor spectra are parameterized as follows,

$$\mathcal{P}_{\mathcal{R}}(k) = A_{\mathcal{R}} \left(\frac{k}{k_*} \right)^{n_s - 1}, \quad \mathcal{P}_T(k) = A_T \left(\frac{k}{k_*} \right)^{n_T}. \quad (13.74)$$

where k_* is a fiducial conformal momentum chosen at one’s convenience. To calculate the scalar tilt ($n_s - 1$), we recall that it is small in slow roll models, make use of the result (13.61) and write

$$\begin{aligned} \mathcal{P}_{\mathcal{R}}(k) &= \mathcal{P}_{\mathcal{R}}(k_*) \left\{ 1 + 2 \frac{d \log \Delta_{\mathcal{R}}}{d\phi} [\phi(\eta_k) - \phi(\eta_{k_*})] \right\} \\ &= \mathcal{P}_{\mathcal{R}}(k_*) \left\{ 1 + \left(3 \frac{V'}{V} - 2 \frac{V''}{V'} \right) [\phi(\eta_k) - \phi(\eta_{k_*})] \right\}, \end{aligned} \quad (13.75)$$

where $\phi(\eta_k)$ is the inflaton field at the time when mode of momentum k exits the inflationary horizon. We have anticipated that the field ϕ varies only slightly in the time interval between η_k and η_{k_*} . It remains to evaluate $[\phi(\eta_k) - \phi(\eta_{k_*})]$. To this end, we recall the definition of the horizon exit,

$$\frac{k}{a(\eta_k)} = H(\eta_k).$$

The scale factor rapidly changes in time, while the Hubble parameter evolves slowly. Neglecting the latter evolution, i.e., setting $H(\eta_k) = H(\eta_{k_*})$, we get

$$\frac{k}{k_*} = \frac{a(\eta_k)}{a(\eta_{k_*})} = e^{N_e(\phi_{k_*}) - N_e(\phi_k)}, \quad (13.76)$$

where $N_e(\phi_k) \equiv N_e(\eta_k)$ is still the number of e -foldings before the end of inflation, so that $a(\eta) \exp[N_e(\eta)] = a_e = \text{const.}$ Now,

$$N_e(\phi_k) - N_e(\phi_{k_*}) = \frac{dN_e}{d\phi} [\phi(\eta_k) - \phi(\eta_{k_*})] = \frac{8\pi V}{M_{Pl}^2 V'} [\phi(\eta_k) - \phi(\eta_{k_*})], \quad (13.77)$$

where we used (12.31) for deriving the last equality. Hence, the formula (13.76) gives

$$\phi(\eta_k) - \phi(\eta_{k_*}) = -\frac{M_{Pl}^2 V'}{8\pi V} \log \frac{k}{k_*}. \quad (13.78)$$

As a result, we obtain from (13.75)

$$\begin{aligned} \mathcal{P}_{\mathcal{R}}(k) &= \mathcal{P}_{\mathcal{R}}(k_*) \left[1 - \left(\frac{3M_{Pl}^2 V'^2}{8\pi V^2} - \frac{M_{Pl}^2 V''}{4\pi V} \right) \log \frac{k}{k_*} \right] \\ &= \mathcal{P}_{\mathcal{R}}(k_*) \left[1 + (2\eta - 6\epsilon) \log \frac{k}{k_*} \right], \end{aligned}$$

where ϵ and η are the slow roll parameters. On the other hand, we find from (13.74) that

$$\mathcal{P}_{\mathcal{R}}(k) = \mathcal{P}_{\mathcal{R}}(k_*) \left[1 + (n_s - 1) \log \frac{k}{k_*} \right].$$

The comparison of the last two formulas yields the final result

$$n_s - 1 = 2\eta - 6\epsilon. \quad (13.79)$$

We see that the scalar tilt depends on both ϵ and η . We recall that the tensor-to-scalar ratio r depends only on ϵ . Hence, the determination of the scalar tilt together with the measurement of scalar and tensor amplitudes will determine both slow roll parameters and hence will favor one or another inflationary model. As an example, the prediction of the large field models, that follows from (12.37) and (12.38), is

$$r = \frac{4n}{N_e}, \quad n_s - 1 = -\frac{n+2}{2N_e}. \quad (13.80)$$

Small field models without fine tuning give (see (12.52) and (12.57); we assume that the potential near its maximum is proportional to an integer power of the field ϕ)

$$r \simeq 0, \quad n_s - 1 = -(1-2) \cdot \frac{1}{N_e}.$$

The hybrid inflation models predict, again barring fine tuning,

$$r \simeq 0, \quad n_s - 1 > 0.$$

The regions in the plane (n_s, r) predicted in these classes of models are shown in Fig. 13.4.

We emphasize that the inflationary models considered in Section 12.2 are merely examples. Furthermore, the results we obtained are incomplete even in these models, since strong inequalities (12.51) and (12.60) may not hold. Hence, the parameters ϵ and η are to be considered as independent from the phenomenological viewpoint, see discussion in Section 13.5.

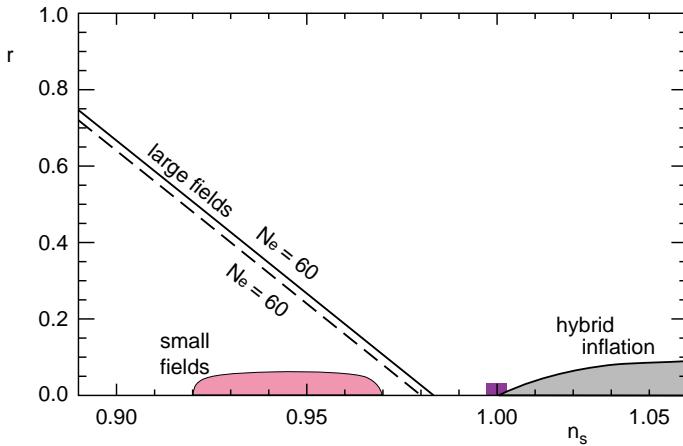


Fig. 13.4 Regions in the plane (n_s, r) predicted by the inflationary models of Section 12.2. Rectangle shows the Harrison–Zeldovich point: flat primordial scalar spectrum in the absence of tensor perturbations, $n_s = 1$, $r = 0$.

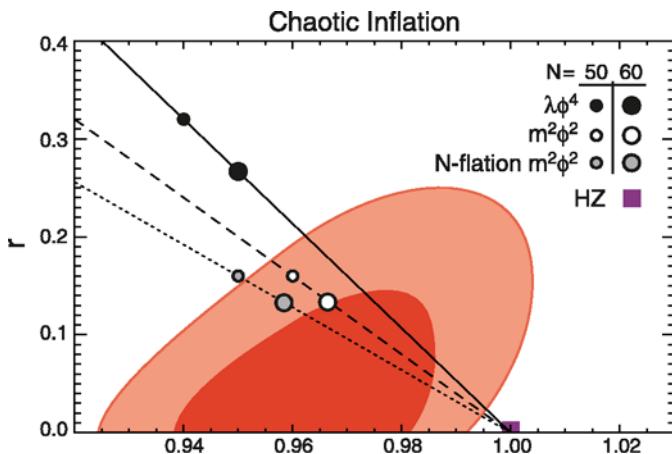


Fig. 13.5 Allowed region in (n_s, r) plane, at 68% and 95% C.L., together with predictions of large field inflation models [2]. The parameter r is defined at the present momentum $q_0 = 0.002 \text{ Mpc}^{-1}$. Rectangle shows the Harrison–Zeldovich point. Solid and long dashed lines show the predictions (13.80) of the large field models with quartic and quadratic potentials, for various values of N_e . Short dashed line is the prediction of a model with numerous inflaton fields with quadratic potential each (N -flation).

The result of the analysis of the existing CMB data together with the data on cosmic structure and SNe Ia is shown in Fig. 13.5. It points towards the negative scalar tilt or sizeable tensor amplitude, or both. Of course, the evidence is still weak; its confirmation or disproof is a matter of relatively near future.

We have not yet been able to obtain any model-independent relation between observables. Such a relation is of course very important for direct proof that the primordial cosmological perturbations were indeed generated at slow roll inflation by the mechanisms studied in this Chapter. One of the relations of this sort is obtained by studying the tensor tilt. In analogy to scalar perturbations, we write for the tensor power spectrum

$$\begin{aligned}\mathcal{P}_T(k) &= \mathcal{P}_T(k_*) \left\{ 1 + \frac{d \log \mathcal{P}_T}{d\phi} [\phi(\eta_k) - \phi(\eta_{k_*})] \right\} \\ &= \mathcal{P}_T(k_*) \left\{ 1 + \frac{V'}{V} [\phi(\eta_k) - \phi(\eta_{k_*})] \right\}.\end{aligned}$$

We now use (13.78) and obtain

$$\mathcal{P}_T(k) = \mathcal{P}_T(k_*) \left[1 - \frac{M_{Pl}^2 V'^2}{8\pi V^2} \log \frac{k}{k_*} \right] = \mathcal{P}_T(k_*) \left[1 - 2\epsilon \log \frac{k}{k_*} \right]. \quad (13.81)$$

Hence, we get finally

$$n_T = -2\epsilon. \quad (13.82)$$

We see that the slow roll inflation predicts the model-independent relation

$$n_T = -\frac{r}{8}. \quad (13.83)$$

Corrections to this relation are small in the slow roll parameters, i.e., they are within 10%.

The experimental test of (13.83) is a hard problem. One should not only detect the effect of the tensor perturbations but also determine their tilt. It goes without saying that testing (13.83) would be extremely important for gaining confidence in the inflationary theory.

Problem 13.10. *Find the value of the parameter M in the inflationary model with the potential (12.69) at which this model is realistic. Calculate the tensor-to-scalar ratio and scalar and tensor tilts in this model.*

To end this Section we note that the analysis of the observational data is often performed by allowing for the possibility that the primordial scalar spectrum deviates from power-law. The parameterization used in such an analysis is given in (5.36). To the lowest order in slow roll, the running index $(dn/d\log k)(k_*)$ in inflationary theory is obtained simply by differentiating the explicit expression for the tilt. It is convenient to make use of the relation between the differentials $d\phi$ and $d\log k$ at the point $\phi(\eta_{k_*}) = \phi_*$ which reads

$$\frac{d}{d\log k} = -\frac{M_{Pl}^2}{8\pi} \frac{V'}{V} \frac{d}{d\phi}.$$

This relation follows directly from (13.78). The formula (13.81) gives for the tensor tilt

$$n_T(k) = -\frac{M_{Pl}^2}{8\pi} \left(\frac{V'}{V} \right)^2(\phi),$$

so that we obtain

$$\frac{dn_T}{d \log k}(k_*) = 4\epsilon\eta - 8\epsilon^2.$$

The scalar tilt is given by (see (13.75))

$$n_s(k) - 1 = \frac{M_{Pl}^2}{4\pi} \left(\frac{V''}{V} - \frac{3}{2} \left(\frac{V'}{V} \right)^2 \right)(\phi),$$

hence

$$\frac{dn_s}{d \log k}(k_*) = 16\eta\epsilon - 24\epsilon^2 - 2\zeta^2,$$

where we introduced the notation

$$\zeta^2 \equiv \left(\frac{M_{Pl}^2}{8\pi} \right)^2 \frac{V'''}{V} \frac{V'}{V}(\phi_*).$$

The running tensor index is small in slow roll models, since it is bilinear in the slow roll parameters. The running scalar index contains the parameter ζ^2 which, generally speaking, does not have to be small. However, ζ^2 is small in simple models, $\zeta^2 \sim \eta^2, \epsilon^2$. This situation is inherent, e.g., in models with power law potentials.

Problem 13.11. Find the running scalar index in the inflationary models of Chapter 12.

13.5 Discussion

We conclude this Chapter with a few comments.

The first comment is on the choice of the inflaton potential. Even though the potentials considered in Section 12.2 may look natural, we have no handle on the true inflaton theory. Accordingly, the regions in the (n_s, r) plane shown in Fig. 13.4 cannot be viewed as ultimate predictions of the slow roll inflation. In fact, by adjusting the inflaton potential one can obtain any values of $(n_s - 1)$ and r ; the only prediction of slow roll inflation is that they are small.¹³ This is clear from the fact that $(n_s - 1)$ and r , as well as the scalar amplitude, are in one-to-one correspondence with the parameters V , ϵ and η , i.e., $V(\phi)$, $V'(\phi)$ and $V''(\phi)$ at $N_e \sim 60$ e -foldings before the end of inflation. So, the measurement of $A_{\mathcal{R}}$, $(n_s - 1)$ and r can be viewed as the reconstruction of $V(\phi)$, $V'(\phi)$ and $V''(\phi)$ at a certain value $\phi_* \equiv \phi(N_e)$. The relation (13.83) becomes then the self-consistency equation.

Once high quality observational data are available, one will be able to reconstruct the inflaton potential in more details (for a review see, e.g., Ref. [126]).

¹³If one gives up slow roll, the latter property no longer holds either.

One way would be to introduce a number of parameters into the inflaton potential (e.g., the coefficients of the Taylor expansion near $\phi = \phi_*$) and solve equations for the background, scalar and tensor perturbations. This approach is hard to pursue analytically, but the numerical analysis is straightforward.

Let us find the range of the inflaton field values where one would in principle be able to reconstruct the inflaton potential. This range is related to the range of present wavelengths ($\lambda_{0,min}, \lambda_{0,max}$) where the spectra of cosmological perturbations can be measured. According to (13.78), the relation is

$$\phi_{max} - \phi_{min} \equiv \Delta\phi = \frac{M_{Pl}^2 V'}{8\pi V} \log \frac{\lambda_{0,max}}{\lambda_{0,min}} = M_{Pl} \sqrt{\frac{\epsilon}{4\pi}} \log \frac{\lambda_{0,max}}{\lambda_{0,min}}.$$

We see that the inflaton potential can be reconstructed in a sizeable field range in models with fairly large ϵ , including the large field models of Section 12.2.1. Let us continue with discussing the latter models. Making use of (12.37), we find for the power-law inflaton potential

$$\Delta\phi = M_{Pl} \cdot \sqrt{\frac{n}{16\pi N_e}} \log \frac{\lambda_{0,max}}{\lambda_{0,min}}. \quad (13.84)$$

The optimistic wavelength interval available to cosmological observations extends from $\lambda_{0,min} \sim 100$ kpc to $\lambda_{0,max} \sim 1$ Gpc. Hence, the relevant field range is somewhat smaller than M_{Pl} , while the field itself is super-Planckian, see (12.35), (12.36). We conclude that the inflaton potential can in principle be reconstructed in a fairly narrow field interval. In other words, by comparing (13.84) to (12.34) we see that

$$\frac{\Delta\phi}{\phi_*} \simeq \frac{1}{2N_e} \log \frac{\lambda_{0,max}}{\lambda_{0,min}} \lesssim 0.1.$$

The potential can in principle be reconstructed via cosmological observations in the field interval which is smaller by an order of magnitude than the field itself. Note, though, that the idea of the inflaton potential in much wider field interval can in principle be obtained by measuring relic gravity waves of relatively short wavelengths, see the end of Section 13.3.

Problem 13.12. *In the large field models, estimate the relative change of the inflaton potential $\Delta V/V$ in the field interval (13.84).*

Problem 13.13. *Make similar estimates for the small field models of Section 12.2.2.*

Problem 13.14. *Assume that there was large field inflation with the power-law potential. In what range of field values this assumption could be confirmed or ruled out if, in addition to the cosmological observations, the gravity wave background*

were measured at wavelengths $\lambda \gtrsim 10^{13} \text{ cm}$? What is the number of e -foldings before the end of inflation at which the latter waves exit the horizon? Hint: Since the range in question is quite large, the approximate formula (13.77) cannot be used.

The second comment is that some properties of the primordial perturbations are strongly constrained in inflationary models without exotica. One example is the negative sign of the tensor tilt. This sign is a consequence of the properties that the amplitude of tensor perturbations is proportional to the Hubble parameter at the horizon exit, that the Hubble parameter decreases in time and that the shorter modes exit the horizon later. Note that this reasoning is independent of the slow roll; one only needs that gravity is described by General Relativity and that the energy density decreases in time. If observations show that the tilt is positive, then either the idea of inflation is irrelevant, or gravity deviates from General Relativity at inflation, or inflaton is phantom field with strongly negative effective equation of state, $p < -\rho$.

Our next comment concerns the level of Gaussianity of the scalar and tensor perturbations. Gaussianity is exact in linear quantum field theory. Hence, the non-Gaussianity is due to non-linear terms in the field equations, which we have discarded so far. One can get a rough idea of the level of non-Gaussianity by comparing non-linear and linear terms in the field equations at the time of the horizon exit at inflation. At $q \sim H$, the non-linear terms for tensor perturbations are estimated as $H^2 h^2$, while linear terms are of order $H^2 h$. Hence, the relative deviation from non-Gaussianity is roughly of order $h \lesssim 10^{-5}$. Similar estimate holds for scalar perturbations [127–129, 8]. This shows that non-Gaussianity is very small in the slow roll models with the inflaton-driven generation of scalar perturbations, so its observational detection is very difficult. Conversely, the observational discovery of non-Gaussianity would imply some other mechanism of the generation of the primordial perturbations. Possibilities of this sort are discussed in literature; we give a simple example in Section 14.2.1.

Our last comment is on the implicit assumption made when calculating the scalar and tensor spectra. We have chosen initial conditions for the quantum fields of inflaton and metric perturbations, implicitly assuming that the equations for these fields, Eqs. (13.16) and (2.58) are valid at arbitrarily high physical momenta. However, there is no guarantee that the properties of quantum fields, and even of space itself, are all the same down to very short distances, say, shorter than the Planck length. So, it is legitimate to ask whether or not the predictions of the inflationary theory are sensitive to physics at short, e.g., sub-Planckian distances [130]. This question is known in literature as the trans-Planckian problem. We have to make a qualification here. Usually, when one talks about short distance physics, one has in mind large (positive or negative) *4-dimensional momentum squared*, which is a Lorentz-invariant quantity. In particular, there is no doubt that quantum gravity effects are strong at $|Q^2| \sim M_{Pl}^2$. This has nothing to do with the trans-Planckian problem we discuss: the sub-horizon inflaton perturbations and

gravitons obey

$$Q^2 = \omega^2 - \mathbf{q}^2 \approx 0,$$

i.e., these fields are almost on mass shell. So, if the theory is exactly Lorentz-invariant (modulo space-time curvature effects, which are small for high momentum modes), high *spatial* momenta are not special, since they are related to low momenta by Lorentz boosts. Hence, trans-Planckian problem is a problem only if one gives up Lorentz-invariance. Nevertheless, this problem deserves the analysis, since there is no guarantee that Lorentz-invariance is an exact property of the theory at arbitrarily high spatial momenta (in other words, at velocities arbitrarily close to the speed of light).

Since theory, let alone experiment, does not tell the way Lorentz-invariance can be violated at ultra-short distances, one approaches the trans-Planckian problem by constructing *ad hoc* models in which Lorentz-invariance is broken at physical spatial momenta $q \gtrsim P_{LV}$, where P_{LV} is the scale of Lorentz-violation. A simple model of this sort is the subject of problem 13.15. In a given model, one studies the inflationary generation of the cosmological perturbations and evaluates the deviations from the predictions of the standard approach. Barring very exotic cases, the general properties of models studied in literature are as follows. First, the inflationary predictions for perturbations are unchanged for $P_{LV} \gg H$, where H is the Hubble parameter at the time of the horizon exit in inflation. More precisely, corrections to the standard results are suppressed by the ratio H/P_{LV} , although the strength of this suppression is model-dependent. Hence, the inflationary predictions are sensitive only to the properties of the fields at relatively low spatial momenta, $q \lesssim H$. Second, if the Hubble parameter H (and, possibly, the Lorentz-violation scale P_{LV}) depends on time weakly, then the power spectra are almost flat (almost independent of k) even for $P_{LV} \lesssim H$. The latter property reflects the invariance of the de Sitter space (inflating Universe with the time-independent H) under spatial dilatation supplemented with time shift. The properties like (13.65), (13.79) and (13.82) are not valid, generally speaking, for $P_{LV} \lesssim H$, and at finite ϵ and η and/or time-dependent P_{LV} the deviations of spectra from flat ones can be much stronger than in the standard theory. Hence, Lorentz-violation at spatial momenta comparable or smaller than H may in principle have observable consequences.

Problem 13.15. One *ad hoc* way to violate Lorentz-invariance is to modify the relativistic dispersion relation $\omega^2 = \mathbf{q}^2$ at high spatial momenta.

Consider a slow roll inflationary model with sufficiently flat inflaton potential. Let the gradient term in the inflaton action be

$$S_{grad} = \frac{1}{2} \int d^3x dt \sqrt{-g} \left[g^{00} (\partial_t \phi)^2 + f \left(-\frac{g^{kl} \partial_k \partial_l}{P_{LV}^2} \right) g^{ij} \partial_i \phi \partial_j \phi \right],$$

and the metric have the standard FLRW form. Here $f(q^2/P_{LV}^2)$ is some function with the property that $f \rightarrow 1$ as $q^2/P_{LV}^2 \rightarrow 0$. Find the power spectrum of the superhorizon inflaton perturbations φ at inflation within the simplified approach of Section 13.1.2. Illustrate the claims made in the text by choosing an appropriate form of the function f and relationship between the parameter P_{LV} and the inflationary Hubble parameter.

Chapter 14

*Further Aspects of Inflationary Theory

14.1 Eternal Inflation

We have seen in Section 13.1.2 that in the course of inflation, more and more modes of the inflaton field exit the horizon, freeze in and become essentially classical. Hence, different Hubble size regions have different values of the homogeneous (on the Hubble scale) inflaton field. As a result, the inflating patch of the Universe is divided into local universes which have their own values of the inflaton field and evolve independently. Under certain conditions which we quantify below, this pile-up of superhorizon perturbations is so strong that the values of the inflaton field in distant local universes are very different, and the Universe as a whole is strongly inhomogeneous. Considering for definiteness large field inflation of Section 12.2.1, we can say that inflation is close to its end in regions with smaller inflaton field, whereas it is still at full strength in regions with larger field. In the inflating regions, the division into local universes and pile-up of perturbations continue, and the whole picture is reproduced. In this way we arrive at the idea of the eternal, self-reproducing Universe [131, 132] divided into regions where inflation has already ended up and regions where inflation is still ongoing in the self-reproduction regime.

We emphasize that this picture of the Universe inhomogeneous at large by no means contradicts the homogeneity of the observed part of the Universe and the inflationary mechanism yielding this homogeneity. The regions considerably different from ours, including still inflating ones, are separated from us by distances exceeding the size of the observable Universe by many orders of magnitude. Hence, in the eternal inflation picture, the properties studied in the previous Chapters remain valid, albeit for finite but very large patch of the Universe. We comment on this aspect further at the end of this Section.

To quantify the above discussion, let us consider large field models of Section 12.2.1 with power-law potentials. Let us choose a region of the Hubble size H^{-1} (the interior of event horizon) in the inflating Universe at time t . Let ϕ_c be the average inflaton field in this region. In the Hubble time $\delta t \sim H^{-1}$, the

average field changes due to the classical evolution by

$$\delta\phi_c = \dot{\phi}_c \delta t \sim -\frac{V'}{H^2},$$

where we used the slow roll equation (12.11). At time $t + \delta t$ this region is divided into a few Hubble size regions where the average field differs from $\phi_c + \delta\phi_c$ due to the inflaton fluctuations that exit the horizon in the time interval δt . To estimate the latter effect, we observe that at times t and $t + \delta t$, superhorizon modes have momenta obeying

$$\frac{k}{a(t)} < H \quad \text{and} \quad \frac{k}{a(t + \delta t)} < H,$$

respectively. Hence, the interval of momenta of the modes that exit the horizon in the time interval δt is

$$\delta k \sim \frac{k}{a} \delta a \sim k H \delta t.$$

Their contribution to the fluctuation of the field ϕ is estimated as (see (13.26))

$$(\delta\varphi)^2 \sim \frac{H^2}{(2\pi)^2} \frac{\delta k}{k}.$$

For $\delta t \sim H^{-1}$ this gives

$$\delta\varphi \sim \frac{H}{2\pi}.$$

This fluctuation is homogeneous on the scale H^{-1} , since it is due to superhorizon modes. Hence, at time $(t + \delta t)$, Hubble size regions inside the original region may have any average fields from the interval

$$\phi_c + \delta\phi_c \pm \delta\varphi.$$

If $\delta\varphi \ll |\delta\phi_c|$, we come back to the picture known from the previous Chapters: the average field decreases in all regions, and the field fluctuations are small. The situation is different for

$$\delta\varphi \gtrsim |\delta\phi_c|. \quad (14.1)$$

Once (14.1) is satisfied, the field rolls *down* the inflaton potential in some regions and rolls *up* in others. Hence, (14.1) is the condition for the eternal inflation regime: inflation occurs at even higher expansion rate in the latter Hubble size regions.

Omitting numerical factors of order 1, we write the condition (14.1) in the following form,

$$H \gtrsim \frac{V'}{H^2}$$

or

$$\frac{V^{3/2}}{M_{Pl}^3 V'} \gtrsim 1. \quad (14.2)$$

It may well hold for energy densities well below the Planck density. As an example, for quadratic and quartic potentials, the ranges of the fields where eternal inflation occurs but quantum gravity effects are small ($V \ll M_{Pl}^4$) are

$$V = \frac{m^2}{2}\phi^2: \quad \sqrt{\frac{M_{Pl}}{m}}M_{Pl} \lesssim \phi \ll \frac{M_{Pl}}{m}M_{Pl}, \quad (14.3)$$

$$V = \frac{\lambda}{4}\phi^4: \quad \lambda^{-1/6}M_{Pl} \lesssim \phi \ll \lambda^{-1/4}M_{Pl}. \quad (14.4)$$

These ranges are sufficiently large for $m \ll M_{Pl}$ and $\lambda \ll 1$, respectively.

In the regime (14.2), pile-up of inflaton perturbations shifts the average field to *smaller* values in some Hubble size regions, and the condition (14.2) ceases to hold there. These regions then undergo “classical” inflation studied in Chapter 12. The number of e -foldings of “classical” inflation has to be recalculated, however. The initial inflaton value for “classical” inflation is obtained by requiring that the left hand side of (14.2) is of order 1. Making use of (12.33), we find the number of e -foldings in the two particular models,

$$\begin{aligned} V = \frac{m^2}{2}\phi^2: \quad N_e^{(tot)} &\sim \frac{M_{Pl}}{m}, \\ V = \frac{\lambda}{4}\phi^4: \quad N_e^{(tot)} &\sim \lambda^{-1/3}. \end{aligned}$$

These estimates are different from (12.41), (12.42), but the qualitative picture is the same: the total number of e -foldings is still extremely large, $N_e \sim 10^4 - 10^6$, and the scale of homogeneity $R \propto \exp(N_e)$ is still much larger than the size of the visible Universe.

A general comment is in order. The concept of eternal inflation cannot be tested experimentally, at least with the present understanding of the issue. The homogeneity scale after the “classical” inflation is way too large. At the same time, the eternal inflation picture, with gigantic patches of the Universe whose cosmology (and possibly even physical laws considered fundamental today) is dramatically different from ours, is of great heuristic value. In particular, it is often used for justifying so called anthropic principle, which we do not discuss in this book, see Refs. [133–136].

14.2 Generation of Scalar Perturbations by Curvaton Mechanism

14.2.1 Non-Gaussian adiabatic perturbations

The simplest mechanism of the generation of scalar perturbations, discussed in Chapter 13, makes use of the inflaton fluctuations. Among its predictions are the Gaussianity of the perturbations and the relation (13.83) between the scalar-to-tensor ratio and tensor tilt. We show in this Section that these predictions are not generic in the inflationary theory. To this end, we consider a simple class of models

where the scalar perturbations originate from an additional scalar field [137–141] called *curvaton*, rather than from the inflaton itself.

Let χ be a scalar field, curvaton, additional to the inflaton ϕ . Let us assume that this field has negligible energy density *at inflation*, as compared to the inflaton energy density, and that the coupling of χ to the inflaton field is negligible. Let us assume further, that the inflaton fluctuations give negligible contribution to the primordial scalar perturbations. The latter assumption is fairly strong; as an example, in the context of the large field inflationary models it means that the parameters in the inflaton potential are small compared to the right hand sides of (13.62), (13.63). We proceed under this assumption and concentrate on the curvaton field χ and its fluctuations. The idea is that the curvaton fluctuations may generate large enough scalar perturbations, even though the field χ is irrelevant from the viewpoint of the inflationary dynamics. If the field χ interacts sufficiently strongly with the usual matter, its fluctuations are reprocessed in the end into the density perturbations of hot plasma, i.e., they become the adiabatic perturbations.

For definiteness, we choose the scalar potential of the curvaton in the quadratic form (the results are not very sensitive to the form of the curvaton potential),

$$V_\chi = \frac{m_\chi^2}{2} \chi^2. \quad (14.5)$$

We assume that during the entire inflationary epoch

$$m_\chi^2 \ll H^2. \quad (14.6)$$

Let the field χ have non-zero classical part χ_c towards the end of inflation (before the last 60 *e-foldings*), which is homogeneous throughout the Universe visible today. The latter assumption is in fact quite natural, as we will see in Section 14.3.1, although the estimate for χ_c is very uncertain. We treat χ_c as a free parameter in this Section.

Once (14.6) is satisfied, the classical curvaton field χ_c stays constant in time at the last 60 *e-foldings* of inflation. Time-independence of χ_c at the late inflationary stage is one of the main requirements; if the potential is different from (14.5), its parameters should be chosen in such a way that this requirement is satisfied. Besides the homogeneous classical part, the field χ contains fluctuation part $\delta\chi$. The fluctuation part originates from the curvaton vacuum fluctuations, and its description is literally the same as that given in Section 13.1.2. The fluctuations $\delta\chi(\mathbf{x})$ are Gaussian, since the curvaton field equation is linear, and their power spectrum at the end of inflation is

$$\mathcal{P}_{\delta\chi} = \frac{H_k^2}{(2\pi)^2}, \quad (14.7)$$

where, like in (13.27), H_k is the Hubble parameter at the time when the mode of momentum k exits the inflationary horizon. Recall (see Section 13.1.2) that this spectrum is nearly flat, and that its tilt coincides with the tensor tilt (13.82).

As we know from Section I.4.8.1, the homogeneous (over the current horizon) field χ begins to oscillate near the minimum of its potential, $\chi = 0$, at the time t_{osc} when the inequality (14.6) ceases to hold, i.e.,

$$H(t_{osc}) \simeq m_\chi. \quad (14.8)$$

According to our assumption, the oscillations begin after the end of inflation. Let us assume for definiteness that this happens at radiation domination,¹ but at high enough temperature, so that all cosmologically interesting perturbations are still superhorizon. In fact, we will have to make a stronger assumption that the curvaton oscillations decay before the generation of the baryon asymmetry and freeze-out of dark matter density. These assumptions do not require fine tuning of parameters.

Under our assumptions, the curvaton field, perturbations included, oscillates in the same way as a homogeneous field. Its energy density just before the beginning of oscillations is

$$\rho_\chi(\mathbf{x}) = \frac{m_\chi^2}{2} \chi^2(\mathbf{x}) = \frac{m_\chi^2}{2} [\chi_c + \delta\chi(\mathbf{x})]^2,$$

where χ_c is still the initial homogeneous field, while $\delta\chi$ is characterized by the power spectrum (14.7). The average energy density is given by

$$\bar{\rho}_\chi = \frac{m_\chi^2}{2} [\chi_c^2 + \langle (\delta\chi)^2 \rangle].$$

Hence, the relative fluctuation of the curvaton energy density has the form

$$\delta_{\rho_\chi}(\mathbf{x}) \equiv \frac{\delta\rho_\chi(\mathbf{x})}{\bar{\rho}_\chi} = 2 \frac{\delta\chi(\mathbf{x})}{\chi_c} + \frac{[\delta\chi(\mathbf{x})]^2 - \langle (\delta\chi)^2 \rangle}{\chi_c^2}, \quad (14.9)$$

where we assume that $\langle (\delta\chi)^2 \rangle \ll \chi_c^2$. Since $\delta\chi(\mathbf{x})$ is the Gaussian random field, the first term here is also Gaussian. However, the second term is non-Gaussian; its correlation functions with the first term and with itself do not obey the Wick theorem. Hence, the field $\delta_{\rho_\chi}(\mathbf{x})$ is non-Gaussian; as an example, its three-point correlation function does not vanish. So, the model has the built-in mechanism of the generation of the non-Gaussianity of the density perturbations [139].

Note that this mechanism leads to a well-defined type of the non-Gaussianity: the density perturbations are local in the Gaussian field $\delta\chi(\mathbf{x})$, and the non-Gaussianity is due to the non-linear relationship between $\delta\rho_\chi$ and $\delta\chi$. It is precisely this type of the non-Gaussianity that we mentioned in the end of Section 5.4; we see that the experimental search for the non-Gaussianity of this type is reasonably well-motivated.

¹Another possibility is that the oscillations begin at the post-inflationary reheating epoch. In the latter case, the results are fairly similar to the ones discussed in the text.

Cosmological observations require that the non-Gaussianity in the primordial density perturbations be not too strong. In the model we consider, the second term in (14.9) is suppressed with respect to the first term by a factor of order

$$\frac{\Delta_{\delta\chi}}{\chi_c} \sim \frac{H_k}{2\pi\chi_c}, \quad (14.10)$$

where, as usual, $\Delta_{\delta\chi} = \sqrt{\mathcal{P}_{\delta\chi}}$. This quantifies the departure from the Gaussianity. In general, the suppression is not extremely strong (see the discussion after Eq. (14.13)), so the non-Gaussianity may be observable.

Problem 14.1. Let us introduce the following dimensionless quantity that estimates the departure from the Gaussianity,

$$d_3 = \frac{|D_3(\mathbf{x}, \mathbf{x}, \mathbf{x})_{\{k\}}|}{[D_2(\mathbf{x}, \mathbf{x})_{\{k\}}]^{3/2}},$$

where

$$\begin{aligned} D_2(\mathbf{x}_1, \mathbf{x}_2)_{\{k\}} &= \langle \delta_{\rho_\chi}(\mathbf{x}_1) \delta_{\rho_\chi}(\mathbf{x}_2) \rangle_{\{k\}} = \mathcal{P}_\delta(k) \\ D_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)_{\{k\}} &= \langle \delta_{\rho_\chi}(\mathbf{x}_1) \delta_{\rho_\chi}(\mathbf{x}_2) \delta_{\rho_\chi}(\mathbf{x}_3) \rangle_{\{k\}} \end{aligned}$$

are two- and three-point equal-time correlation functions, and the subscript $\{k\}$ denotes the contribution of the decimal interval of momenta around a given value k . Calculate d_3 in the model under discussion for $\Delta_\chi \ll \chi_c$. Is d_3 estimated by (14.10)?

Let us assume that the curvaton interactions with ordinary matter is strong enough.² As we discussed in Sections I.4.8.1 and 2.5, almost homogeneous oscillating field χ can be viewed as the condensate of χ -bosons of almost zero momenta. If χ -bosons are unstable, then the lifetime of the condensate is not larger than the lifetime of an individual χ -boson.³ The oscillations of the curvaton field are damped, and the field relaxes to its vacuum value $\chi = 0$ by the time t_{dec} , such that

$$H(t_{dec}) \simeq \Gamma_\chi, \quad (14.11)$$

where $\Gamma_\chi = \tau_\chi^{-1}$ is the inverse lifetime of the χ -condensate. In weakly coupled theories one has $\Gamma_\chi < m_\chi$. A particularly interesting possibility is that the decay of the oscillations occurs much later than their beginning, i.e., $\Gamma_\chi \ll m_\chi$ and hence $t_{dec} \gg t_{osc}$. This is natural if the curvaton interactions with the usual particles is weak enough.

As a result of the decay of χ -oscillations, the curvaton energy is transmitted to the relativistic hot gas. The perturbations of the curvaton become the perturbations

² Practically non-interacting field is cold dark matter, see Section 2.5, and perturbations of its energy density are in the CDM isocurvature mode. We study this possibility in Section 14.4 using axion as an example.

³ We will see in Chapter 15 that the lifetime of a scalar field condensate may be considerably shorter than that of an individual boson.

in the relativistic component, i.e., adiabatic perturbations. At this point we have to assume that the decay of χ -condensate occurs before freeze-out of dark matter density. In the opposite case, the dark matter perturbations would be absent, and we would have $\delta_{CDM} = 0$ in notations of Section 5.2, while $\delta_{CDM} = (3/4)\delta_{rad}$ for the superhorizon adiabatic mode. This would mean that the primordial scalar perturbations contain large admixture of the CDM isocurvature mode, anticorrelating with the adiabatic mode. This option is ruled out by the cosmological observations, see Section 5.4. For the same reason, the χ -condensate should decay before the generation of the baryon asymmetry.

Let us introduce the parameter r_χ as the ratio of the curvaton energy density to the energy density of the hot matter just before the decay of the curvaton oscillations,

$$\rho_\chi(t_{dec}) = r_\chi \rho_{rad}(t_{dec}). \quad (14.12)$$

The perturbations in the total energy density $\delta\rho$ after the decay of the χ -condensate are the same as the perturbations $\delta\rho_\chi$ before the decay, modulo numerical coefficient of order 1. Therefore, the relative perturbations after the decay are estimated as

$$\delta \equiv \frac{\delta\rho}{\rho} \simeq \frac{\delta\rho_\chi}{\rho_{rad} + \rho_\chi} \simeq \frac{r_\chi}{1 + r_\chi} \cdot \delta_{\rho_\chi}. \quad (14.13)$$

We make three comments concerning the result (14.13). First, at $r_\chi \ll 1$, the amplitude of the curvaton perturbations may not be very small despite the small value of $\delta \sim \mathcal{R}$. Hence, the non-Gaussianity parameter (14.10) need not be very small. Second, the statistical properties of the random field $\delta(\mathbf{x})$ coincide with those of the field $\delta_{\rho_\chi}(\mathbf{x})$, so the adiabatic perturbations have the same degree of the non-Gaussianity as the original curvaton perturbations. Finally, since the non-Gaussianity is small, we have $\delta_{\rho_\chi} = 2\delta\chi/\chi_c$ modulo small corrections, and because of (14.7), the dependence of the adiabatic power spectrum on k is the same as for the tensor spectrum. In terms of the spectral tilts introduced in Section 13.4.3 this means that

$$n_s - 1 = n_T. \quad (14.14)$$

We note that the model predicts red scalar spectrum, $n_s < 1$, and that the equality (14.14) is completely foreign to the inflaton mechanism of the generation of the scalar perturbations.

Let us now discuss the tensor-to-scalar ratio. To this end, we estimate the Hubble parameter H_k some 60 e -foldings before the end of inflation, required for obtaining the correct scalar amplitude. We first relate the initial curvaton field χ_c to the parameter r_χ . We assume for definiteness that $r_\chi \lesssim 1$, then the cosmological expansion at the epoch of the curvaton oscillations proceeds in the radiation domination regime, $a(t) \propto t^{1/2}$, $H(t) \propto t^{-1}$. Hence, Eqs. (14.8), (14.11) yield

$$\frac{a(t_{osc})}{a(t_{dec})} = \left(\frac{\Gamma_\chi}{m_\chi} \right)^{1/2}.$$

The energy density of the oscillating field χ decreases as a^{-3} , see Sections I.4.8.1 and 2.5, so that

$$\rho_\chi(t_{dec}) = \left(\frac{a(t_{osc})}{a(t_{dec})} \right)^3 \rho_\chi(t_{osc}) \sim \left(\frac{\Gamma_\chi}{m_\chi} \right)^{3/2} m_\chi^2 \chi_c^2.$$

Omitting numerical factors of order 1 and using the Friedmann equation, we find that at the decay of oscillations

$$\rho_{rad}(t_{dec}) \sim H^2(t_{dec}) M_{Pl}^2 \sim \Gamma_\chi^2 M_{Pl}^2.$$

Hence, the relation (14.12) gives

$$\chi_c^2 \sim r_\chi \left(\frac{\Gamma_\chi}{m_\chi} \right)^{1/2} M_{Pl}^2. \quad (14.15)$$

Note that our scenario requires $\chi_c \ll M_{Pl}$. Finally, we recall that $\delta \simeq \mathcal{R}$ for the adiabatic mode, make use of (14.7), (14.9), (14.15) and write (14.13) in the following form,

$$\frac{H_k}{2\pi M_{Pl}} \sim r_\chi^{-1/2} \left(\frac{\Gamma_\chi}{m_\chi} \right)^{1/4} \Delta_{\mathcal{R}}. \quad (14.16)$$

This is the desired expression for the inflationary Hubble parameter. Its alternative expression in terms of χ_c can be obtained from (14.15). Modulo numerical factor, the left hand side of (14.16) is the tensor amplitude Δ_T , see Section 13.3. We see that the tensor-to-scalar ratio $r = \Delta_T^2 / \Delta_{\mathcal{R}}^2 \propto r_\chi^{-1} (\Gamma_\chi / m_\chi)^{1/2}$ is determined by the interplay between the small parameters Γ_χ / m_χ and r_χ . Hence, there is no simple relationship between the scalar-to-tensor ratio and spectral tilts.

14.2.2 Isocurvature modes

Versions of the curvaton model may be used to illustrate the possibility of the generation of isocurvature modes that correlate or anticorrelate with the adiabatic mode [138–140, 142, 143]. We discuss this possibility using the CDM-mode as an example; the idea can be extended to the baryon isocurvature mode.

The first version assumes that dark matter particles are created only in the curvaton decays, and they never get into chemical equilibrium. This is possible if dark matter particles do not directly interact with the Standard Model particles, and their interaction with the curvaton is weak enough, so that their annihilation via curvaton exchange is negligible. Let us assume, in contrast to Section 14.2.1, that the energy density of the oscillating curvaton field dominates in the Universe before the curvaton decay, so that $r_\chi \gg 1$. Let us assume also that the curvaton decays into both dark matter and Standard Model particles. The latter decays generate entropy in the plasma. In this situation, the adiabatic mode of scalar perturbations is predominantly generated. To estimate the admixture of the CDM isocurvature mode, we note that after the curvaton decays, total entropy is the sum of entropy

that existed before the decays (s_{old}) and entropy generated in the decays (s_{dec}), with obvious relation between them, $s_{old}/s_{dec} \simeq r_\chi^{-3/4} \ll 1$. The ratio of the number density of dark matter particles n_{CDM} to s_{dec} is constant in space, even though both n_{CDM} and s_{dec} are inhomogeneous because of the initial curvaton perturbations. It is convenient to choose the gauge in which the spatial curvature of constant-time hypersurfaces vanishes. Under the above assumption that the only source of scalar perturbations are curvaton fluctuations, we have in this gauge $\delta s_{old} = 0$, $\Psi = \Phi = 0$, while the invariants ζ_λ are given by (5.29). The perturbation in n_{CDM}/s is

$$\delta\left(\frac{n_{CDM}}{s}\right) = \delta\left(\frac{n_{CDM}}{s_{dec} + s_{old}}\right) = \delta\left[\frac{n_{CDM}}{s_{dec}}\left(1 - \frac{s_{old}}{s_{dec}}\right)\right] = \frac{n_{CDM}}{s_{dec}} \frac{s_{old}}{s_{dec}} \frac{\delta s_{dec}}{s_{dec}}.$$

Since the strongest perturbations are adiabatic, we have $\delta s_{dec}/s_{dec} = 3/4 \cdot \delta_{rad} \approx 3\zeta \approx 3\mathcal{R}$. As a result, we obtain

$$\mathcal{S}_{CDM} \equiv \frac{\delta(n_{CDM}/s)}{n_{CDM}/s} \approx 3r_\chi^{-3/4}\mathcal{R}.$$

Hence, this mechanism generates the CDM isocurvature mode *completely correlated with the adiabatic mode*. Furthermore, these modes have the same shapes of the power spectra. The physics behind this result is that there is one and the same source of perturbations δn_{CDM} and δs , but the relative perturbations in entropy are smaller because of the presence of the homogeneous entropy density s_{old} . The admixture of the CDM mode is small, provided that this homogeneous entropy is subdominant, i.e., $r_\chi^{-1} \ll 1$.

The possibility of the *anticorrelation* between the isocurvature and adiabatic modes is more contrived. Yet, it cannot be discarded completely. Such a possibility is realized in the following version of the curvaton scenario. Let us assume that dark matter particles are produced both in the curvaton decay and independently, before this decay. Let us assume further that dark matter particles get out of chemical equilibrium with the hot component before the curvaton decay. Let us assume finally that r_χ is so large that entropy before the curvaton decay is negligibly small. Then after curvaton decay, the number density of dark matter particles is equal to $n_{CDM} = n_{dec} + n_{old}$, while the entropy density is s_{dec} . We assume that

$$n_{old} \ll n_{dec}; \quad (14.17)$$

the reason for this assumption will become clear soon. In the reference frame of zero spatial curvature of equal-time hypersurfaces, n_{old} is spatially homogeneous, while the perturbations δn_{dec} and δs_{dec} obey $\delta(n_{dec}/s_{dec}) = 0$. Hence, we find

$$\delta\left(\frac{n_{CDM}}{s}\right) = \delta\left(\frac{n_{dec} + n_{old}}{s_{dec}}\right) = -\frac{n_{old}}{n_{dec}} \frac{n_{dec}}{s_{dec}} \frac{\delta s_{dec}}{s_{dec}}.$$

Finally,

$$\mathcal{S}_{CDM} = \frac{\delta(n_{CDM}/s)}{n_{CDM}/s} \approx -\frac{n_{old}}{n_{dec}} \cdot 3\mathcal{R}.$$

Thus, the isocurvature mode exactly anticorrelates with the adiabatic mode. The physics is that there is again one and the same source of δn_{CDM} and δs , but the relative number density perturbations of dark matter particles are diluted due to the presence of the homogeneous part n_{old} . The assumption (14.17) must be valid, since the admixture of the isocurvature mode must be small.

Hence, introducing additional field(s) responsible for the generation of scalar perturbations opens up new possibilities, as compared to the model with the inflaton field only. These are sizeable non-Gaussianity of scalar perturbations, absence of any relationship between scalar and tensor amplitudes and their tilts, admixture of isocurvature modes. Hence, the precision study of the cosmological perturbations that can confirm or rule out these possibilities is of great interest.

14.3 Light Scalar Field in Inflating Universe

We discuss in this Section the behavior of a scalar field with relatively small mass in the inflating Universe. This aspect is of interest from the viewpoint of initial data for the post-inflationary evolution in models with relatively light scalar fields. One of these is the axion model studied in Section 14.4, another example is the Affleck–Dine model of the baryon asymmetry generation (see Section I.11.6). We first discuss free massive field and show that fluctuations of this field in the *de Sitter* space-time have well-defined and calculable amplitude, which is large for small mass. We show, however, that in realistic inflationary models, this amplitude strongly depends, generally speaking, on the evolution of the scale factor at the early inflationary stages. Hence, the predictions concerning the initial data for the post-inflationary evolution suffer from considerable uncertainties.

Since the fluctuation amplitude is large, free field approximation may be inadequate. For this reason, we also consider self-interacting scalar field. We find that the uncertainty in predictions is reduced, but the dependence on the details of inflation and the strength of the self-coupling does not go away.

We study the scalar field minimally coupled to gravity. It is worth noting that in the opposite case, when the Lagrangian of the scalar field σ contains the term $\xi_\sigma R\sigma^2$ with $\xi_\sigma \gtrsim 1$, the scalar field has large effective mass at inflation, $m_{eff}^2 \sim \xi_\sigma H^2$, and our analysis is not valid.

14.3.1 *Light field without self-interaction*

We begin with the model that contains, besides the inflaton, a free scalar field σ of mass m_σ . Its modes obey the equation (cf. Eq. (13.12))

$$\ddot{\sigma} + 3H\dot{\sigma} + \frac{k^2}{a^2}\sigma + m_\sigma^2\sigma = 0. \quad (14.18)$$

We consider inflation with the Hubble parameter varying slowly in time. We assume that the mass is small,⁴

$$H \gg m_\sigma,$$

and neglect the back reaction of the field σ on the expansion of the Universe. Like the inflaton fluctuations, the fluctuations of the field σ exit the horizon and then evolve very slowly. As a result, large fluctuation $\langle \sigma^2(\mathbf{x}) \rangle$ is produced. Our purpose is to find its value at time t . Clearly, interesting time is the time inflation ends and possibly somewhat earlier times corresponding to the last 60 e -foldings.

If we neglected the mass m_σ , we would obtain the power spectrum of superhorizon modes given by (13.27), i.e., $\mathcal{P}_\sigma = H^2/(2\pi)^2$. Hence, the field fluctuation would be given by the logarithmically divergent integral

$$\langle \sigma^2(\mathbf{x}) \rangle = \int_0^{q_{max}} \frac{dq}{q} \mathcal{P}_\sigma, \quad (14.19)$$

where $q_{max} \sim H$ is the maximum momentum of superhorizon modes. The divergency of this integral signals that we cannot neglect the mass m_σ when calculating $\langle \sigma^2(\mathbf{x}) \rangle$, i.e., when estimating the range of values the field σ takes in different patches of the inflating Universe. The physics is that the field σ evolves, albeit slowly, due to its mass, and its amplitude decreases gradually. To account for this fact, we neglect the momentum k in Eq. (14.18) for superhorizon modes, thus reducing it to the equation for a homogeneous field. The evolution proceeds in the slow roll regime, so σ obeys the equation analogous to Eq. (12.11),

$$\dot{\sigma} = -\frac{m_\sigma^2 \sigma}{3H}. \quad (14.20)$$

Problem 14.2. Show that the slow roll approximation (14.20) is indeed valid for $H \gg m_\sigma$.

Let $q(t)$ be the physical momentum of a given mode at time t . Let this mode exit the horizon at time t_q , so that $q(t_q) = H(t_q)$; for infrared modes of interest $t_q < t$. Then we have for the value of the physical momentum at later time

$$q(t) = H(t_q) \exp\left(-\int_{t_q}^t H dt'\right). \quad (14.21)$$

We find from (14.20) the following relation between the amplitude of the mode at the horizon exit and at time t ,

$$\sigma(t) = \sigma(t_q) \exp\left(-\int_{t_q}^t \frac{m_\sigma^2}{3H(t')} dt'\right). \quad (14.22)$$

⁴In the opposite case, $m_\sigma \gg H$, the field always stays near its vacuum value $\sigma = 0$.

We know the amplitude of $\sigma(t_q)$: it is equal to $H(t_q)/(2\pi)$. Hence, instead of (14.19) we find from (14.22) the expression valid for non-zero mass,

$$\langle \sigma^2(\mathbf{x}, t) \rangle = \int_0^{H(t)} \frac{dq}{q} \frac{H^2(t_q)}{(2\pi)^2} \exp\left(-\int_{t_q}^t \frac{2m_\sigma^2}{3H(t')} dt'\right). \quad (14.23)$$

Here t_q is related to q by (14.21). The latter relation can be used to change the integration variable from q to t_q . Neglecting the slow evolution of the Hubble parameter at time t_q , we find from (14.21) that

$$\frac{d \log q}{dt_q} = H(t_q),$$

and finally

$$\langle \sigma^2(\mathbf{x}, t) \rangle = \int_{t_i}^t dt_q \frac{H^3(t_q)}{(2\pi)^2} \exp\left(-\int_{t_q}^t \frac{2m_\sigma^2}{3H(t')} dt'\right). \quad (14.24)$$

The lower limit of integration t_i is written here quite formally; time t_i corresponds, roughly speaking, to the beginning of inflation, and we have essentially no idea of the Universe at that time.

Let us first apply the general formula (14.24) to the de Sitter Universe with $H = \text{const}$. The integrals are straightforwardly calculated, and we obtain [144–146]

$$\langle \sigma^2(\mathbf{x}) \rangle = \frac{3H^4}{8\pi^2 m_\sigma^2}. \quad (14.25)$$

The amplitude is finite in the de Sitter Universe, albeit large for small mass. Another derivation of the result (14.25) is given at the end of this Section, see (14.35).

In realistic theories of inflation, the Hubble parameter depends on time, and the situation is much more complicated. We will see that the result (14.25) is valid in rather narrow mass interval,

$$H^2 \gg m_\sigma^2 \gg \epsilon H^2, \quad (14.26)$$

where $\epsilon = |\dot{H}|/H^2$ is the slow roll parameter. The first inequality here is our basic assumption, while the second inequality is non-trivial. For $m^2 \ll \epsilon H^2$, the amplitude of the field σ is different from (14.25) and is determined by several effects.

In the first place, the field σ is not necessarily equal to zero at the beginning of inflation. As an example, the picture of chaotic initial state discussed in Section 12.2.1 suggests that the energy density of this field is of the order the Planck density,

$$m_\sigma^2 \sigma_i^2 \sim M_{Pl}^4. \quad (14.27)$$

This field quickly becomes spatially homogeneous due to inflation. The classical field obeys Eq. (14.20), whose solution is

$$\sigma_c = \exp\left(-m_\sigma^2 \int_{t_i}^t \frac{dt'}{3H(t')}\right) \cdot \sigma_i.$$

The initial value is irrelevant for

$$m_\sigma^2 \int_{t_i}^t \frac{dt'}{3H(t')} \gg 1, \quad (14.28)$$

otherwise the classical field σ_c does not change much from the beginning of inflation till time t . In some (but not all) inflationary models one has

$$\int_{t_i}^t \frac{dt'}{3H(t')} \sim \frac{1}{\dot{H}(t)}. \quad (14.29)$$

Then the relation (14.28) reduces to $m^2 \gg \dot{H}$, i.e., to the second inequality in (14.26). Hence, for $m_\sigma^2 \lesssim \epsilon H^2$, the field σ is partially determined by its initial value even at the end of inflation.

Note that the energy density of σ_c may start exceeding the inflaton energy density at some point. Then the field σ itself becomes the inflaton.

Problem 14.3. Show that the relation (14.29) is indeed valid at sufficiently late inflationary stage in the large field model with the inflaton potential $V(\phi) = (\lambda/4)\phi^4$ and in the power-law inflationary model, $a \propto t^\alpha$, $\alpha > 1$. Is it valid in the large field model with $V(\phi) = (m^2/2)\phi^2$? Hint: Make use of the results of problem 12.3.

Problem 14.4. Consider large field inflation with the inflaton potential $V(\phi) = (\lambda/4)\phi^4$ and initial data (14.27) for the classical field σ_c . Find the range of m_σ at given λ in which the energy density of the field σ_c never exceeds the inflaton energy density at the inflationary epoch. Making use of the results of Section 13.4.1, give numerical estimate for m_σ .

Let us come back to quantum fluctuations of the field σ . Since the equation for this field is linear, they evolve independently of the classical field σ_c . We continue to denote the field of quantum fluctuations by $\sigma(\mathbf{x}, t)$. An important point is as follows. For $H = \text{const}$, the integral (14.24) is saturated at

$$(t_q - t) \sim H/m_\sigma^2, \quad (14.30)$$

which is much larger than the Hubble time. The result (14.25) is valid only if the Hubble parameter is practically constant in this time interval. Hence, the domain of applicability of (14.25) is $\dot{H} \cdot H/m_\sigma^2 \ll H$, which is precisely the second inequality in (14.26).

Let us now turn to the opposite case,

$$m_\sigma^2 \ll \epsilon H^2. \quad (14.31)$$

According to (14.30), the relevant number of e -foldings is estimated as

$$N_e^{(\sigma)} \sim \frac{H^2}{m_\sigma^2}. \quad (14.32)$$

This is a large number, and it is not guaranteed that the total number of e -foldings at inflation, $N_e^{(tot)}$, is actually larger than H^2/m_σ^2 . For $N_e^{(tot)} < H^2/m_\sigma^2$, the result for

$\langle \sigma^2(\mathbf{x}) \rangle$ depends on the entire inflationary epoch (and even pre-inflationary epoch). Indeed, the contribution of modes that exit the horizon at inflation is estimated in the latter case as follows,

$$\langle \sigma^2(\mathbf{x}) \rangle = N_e^{(tot)}(t) \frac{\bar{H}^2}{(2\pi)^2}, \quad N_e^{(tot)} < H^2/m_\sigma^2, \quad (14.33)$$

where

$$\bar{H}^2 = \frac{1}{N_e^{(tot)}(t)} \int_0^{N_e^{(tot)}} H^2 dN_e$$

is the average of the Hubble parameter squared over all e -foldings, and $H(N_e)$ is the Hubble parameter at N_e e -foldings before the time t at which we calculate the variance $\langle \sigma^2(\mathbf{x}) \rangle$. As expected, the whole inflationary epoch contributes to (14.33), and the early stage with larger H is more relevant.

Problem 14.5. Obtain the estimate (14.33).

We see from (14.33) that in the case $N_e^{(tot)}(t) \ll H^2/m_\sigma^2$ we consider here, the fluctuations of the field σ are smaller than in the de Sitter space with the Hubble parameter \bar{H} : too few modes have exited the horizon since the beginning of inflation, and their total contribution has not yet been saturated at the de Sitter value. On the other hand, the result (14.33) shows that the generated fluctuation exceeds considerably the naive estimate $\sigma \sim H/(2\pi)$.

Even if the number of e -foldings is large, $N_e^{(tot)}(t) \gg H^2/m_\sigma^2$, the fluctuation of the field σ may be sensitive to the initial period of inflation, and hence it cannot be reliably calculated. In any case, the de Sitter formula (14.25) is, generally speaking, irrelevant unless (14.26) is satisfied.

Two comments are in order. First, we have encountered here an unusual phenomenon: the de Sitter approximation is not valid, even at the qualitative level, in describing the physics at the inflationary epoch. This is due to the fact that the naively logarithmic integral (14.23) is saturated at exponentially low momenta,

$$q \sim H \exp(-\text{const} \cdot H^2/m_\sigma^2). \quad (14.34)$$

The corresponding modes exit the inflationary horizon very early, hence the evolution of the Hubble parameter is by no means negligible.

Second, the spatial scale of the inhomogeneity of the field σ is very large. As an example, in the de Sitter space-time this scale is of order $l \sim H^{-1} \exp(\text{const} \cdot H^2/m_\sigma^2)$, as one can see from (14.34). For $m_\sigma^2 \ll H^2/60$ this scale exceeds the size of the visible Universe at the end of inflation. Hence, the field σ generated at inflation is spatially homogeneous over the entire visible Universe.

Problem 14.6. In the context of the large field inflationary model with the inflaton potential $V(\phi) = (\lambda/4)\phi^4$, assume that inflation begins at $V(\phi_i) \sim M_{Pl}^4$

(see Section 12.2.1) and find the range of m_σ in which the fluctuation $\langle \sigma^2 \rangle$ is saturated at the very early period of inflation. Give numerical estimate, using the results of Section 13.4.1. Estimate the fluctuation of the field σ for all masses smaller than the Hubble parameter H_e at the end of inflation. Hint: Change the integration variable in (14.24) to $\phi(t_q) \equiv \phi$, the classical inflaton field.

Problem 14.7. Consider the large field inflation with quadratic inflaton potential, $V(\phi) = (m_\phi^2/2)\phi^2$. Show that for $m_\sigma > m_\phi$ the fluctuation of the field σ is independent of the details of the early inflation. Find this fluctuation and compare it with (14.25).

Let us give the direct derivation of the result (14.25) in the framework of quantum field theory in the de Sitter space-time. The modes of the field σ obey

$$\sigma'' + 2\frac{d}{a}\sigma' + k^2\sigma + m_\sigma^2 a^2\sigma = 0.$$

We have $a = -1/(H\eta)$ for the de Sitter space-time, so this equation reads

$$\sigma'' - \frac{2}{\eta}\sigma' + k^2\sigma + \frac{m_\sigma^2}{H^2\eta^2}\sigma = 0.$$

The general solution to this equation is

$$\sigma = \text{const} \cdot |\eta|^{3/2} Z_\nu(k\eta),$$

where

$$\nu = \sqrt{\frac{9}{4} - \frac{m_\sigma^2}{H^2}} \approx \frac{3}{2} - \frac{m_\sigma^2}{3H^2},$$

and Z_ν is a linear combination of the Bessel functions. The mass m_σ can be neglected at large negative η , so we arrive at the situation described in Section 13.1.2. In particular, the quantum field $\chi = a\sigma$ is given by (13.20) at early times. We use the asymptotics of the Hankel functions $H_\nu^{(1,2)}$ at large argument, and write

$$\sigma(\mathbf{x}, \eta) = \int \frac{d^3k}{(2\pi)^{3/2}} \frac{\sqrt{\pi}}{2} H|\eta|^{3/2} [\mathrm{e}^{-i\mathbf{kx}} H_\nu^{(1)}(k\eta) A_\mathbf{k}^\dagger + \mathrm{e}^{i\mathbf{kx}} H_\nu^{(2)}(k\eta) A_\mathbf{k}],$$

where the operators $A_\mathbf{k}^\dagger$ and $A_\mathbf{k}$ differ from similar operators in (13.20) by irrelevant phase factors. The behavior of the Hankel functions at small $k\eta$ is as follows,

$$H_\nu^{(1,2)}(k\eta) = \mp i \frac{1}{\sin \nu \pi} \left(\frac{2}{k\eta} \right)^\nu \frac{1}{\Gamma(1-\nu)} \approx \mp i \sqrt{\frac{2}{\pi}} \frac{1}{(k\eta)^\nu},$$

where we set $\nu = 3/2$ wherever possible. As a result, the superhorizon modes of the field $\sigma(\mathbf{x}, \eta)$ are

$$\sigma(\mathbf{x}, \eta)_{\{\mathbf{k}\}} = \int d^3k \frac{1}{4\pi^{3/2}} \frac{H}{k^\nu} |\eta|^{3/2-\nu} [\mathrm{e}^{-i\mathbf{kx}} \tilde{A}_\mathbf{k}^\dagger + \mathrm{e}^{i\mathbf{kx}} \tilde{A}_\mathbf{k}].$$

Note the extra factor $|k\eta|^{3/2-\nu}$ in the integrand, as compared to (13.25). This immediately gives the desired result,

$$\langle \sigma^2(\mathbf{x}) \rangle = \int_0^{k_{max}} \frac{4\pi k^2 dk}{16\pi^3} \frac{H^2}{k^{2\nu}} |\eta|^{3-2\nu} = \frac{H^2}{(2\pi)^2} \frac{1}{3-2\nu} |k_{max} \eta|^{3-2\nu} = \frac{3}{8\pi^2} \frac{H^4}{m_\sigma^2}, \quad (14.35)$$

where $k_{max} \sim |\eta^{-1}|$ is the maximum momentum of superhorizon modes, and we have kept the leading order in the small parameter m^2/H^2 . Note that the integral is indeed saturated at conformal momenta

$$k \sim |\eta^{-1}| \cdot \exp\left(-\frac{1}{3-2\nu}\right) \sim \eta^{-1} \exp\left(-\frac{3H^2}{2m_\sigma^2}\right),$$

i.e., at exponentially small momenta (14.34).

14.3.2 Model with quartic potential

As we have seen in Section 14.3.1, the result for the fluctuation of the free light field is plagued by uncertainties: its fluctuation is often dominated by very large wavelength modes exiting the horizon at the early inflationary epoch. Also, there remains the dependence on the initial value of the field. The situation can be different in nonlinear theory. Consider as an example the theory of the field σ with quartic potential,

$$V_\sigma = \frac{\lambda_\sigma}{4} \sigma^4, \quad (14.36)$$

where the coupling λ_σ is not very small. We neglect the mass of the field σ ; we will see that this is legitimate provided that

$$m_\sigma^2 \ll 3\sqrt{\frac{\lambda_\sigma}{8\pi^2}} H^2. \quad (14.37)$$

The initial value σ_i can be different from zero in this model too. The analog of Eq. (14.20) for the classical field is

$$3H\dot{\sigma}_c = -\lambda_\sigma \sigma_c^3,$$

and its solution for large enough σ_i is given by

$$\sigma_c = \left(\lambda_\sigma \int_{t_i}^t \frac{2dt}{3H(t)} \right)^{-1/2}. \quad (14.38)$$

We will come back to this formula soon.

Problem 14.8. Consider large field inflation with the inflaton potential $V(\phi) = (\lambda/4)\phi^4$. What is the relationship between λ and λ_σ for which the field (14.38) gives negligible contribution to the energy density as compared to the inflaton? Is the slow roll condition $\ddot{\sigma}_c \ll 3H\dot{\sigma}_c$ satisfied? Hint: Make use of (14.29) and express the results in terms of the slow roll parameter ϵ .

Consider now quantum fluctuations of the field σ assuming that σ_c is negligible and the Hubble parameter is constant in time. Let $\bar{\sigma}$ be the superhorizon part of the field σ . It is homogeneous over the horizon, and it can be treated as a classical field. Let us denote by σ_k the quantum mode that exits the horizon and soon after adds up to $\bar{\sigma}$. In the presence of the classical field it obeys the equation

$$\ddot{\sigma}_k + 3H\dot{\sigma}_k + \frac{k^2}{a^2}\sigma_k + 3\lambda_\sigma\bar{\sigma}^2\sigma_k = 0. \quad (14.39)$$

We see that this equation involves the effective mass

$$m_{eff}^2 = 3\lambda_\sigma\bar{\sigma}^2. \quad (14.40)$$

In the stationary situation, the mean field generated by modes exiting the horizon coincides with $\bar{\sigma}$. On the other hand, this mean field is given by (14.25) with the effective mass m_{eff} . This yields the self-consistency condition⁵

$$\bar{\sigma}^2 \simeq \frac{3H^4}{8\pi^2 m_{eff}^2},$$

so that finally

$$\bar{\sigma}^2 \simeq \sqrt{\frac{1}{8\pi^2\lambda_\sigma}} H^2, \quad m_{eff}^2 = 3\sqrt{\frac{\lambda_\sigma}{8\pi^2}} H^2. \quad (14.41)$$

Our assumption that the time-dependence of the Hubble parameter can be neglected, as well as the slow roll approximation used throughout the derivation are valid only if (14.26) is satisfied for the effective mass m_{eff} . The inequality $m_{eff}^2 \ll H^2$ is valid for all $\lambda_\sigma \ll 1$, while the second inequality in (14.26) gives

$$\lambda_\sigma \gg \frac{8\pi^2}{9}\epsilon^2. \quad (14.42)$$

Hence, the applicability of the result (14.41) is limited by the condition (14.42).

Two comments are in order. First, once the condition (14.42) is satisfied, the classical field (14.38) is smaller than $\bar{\sigma}$ in many inflationary models. Consider, as an example, models where the relation (14.29) holds. In that case we obtain from (14.38) that $\sigma_c^2 \sim \dot{H}/\lambda_\sigma$, and the inequality $\bar{\sigma}^2 \gg \sigma_c^2$ is the same as (14.42) modulo a factor of order 1. Hence, if the condition (14.42) is satisfied, ambiguities due to the unknown initial value σ_i are irrelevant in many inflationary models.

The second comment has to do with the mass term of the field σ omitted in (14.36). The mass m_σ is indeed irrelevant for $m_\sigma^2 \ll m_{eff}^2$, which is equivalent to (14.37).

If the inequality (14.42) is valid up until the end of inflation, the estimate (14.41) is valid at all interesting times of the inflationary epoch. This situation is possible

⁵Since we are dealing with random fluctuations, our analysis is not rigorous. In fact, we do not even give accurate definition of $\bar{\sigma}$. Hence, our results should be treated as the estimates valid up to a factor of order 1.

in small field and hybrid inflation models, where ϵ is small until the end of inflation. The spatial distribution of the field σ at the end of inflation strongly depends on the coupling λ_σ . Indeed, we have seen in Section 14.3.1 that the field $\bar{\sigma}$ is dominated by modes of exponentially low momenta. Instead of (14.34), we now have a rough estimate

$$q \sim H \cdot \exp\left(-\text{const} \cdot \frac{H^2}{m_{\text{eff}}^2}\right) \sim H \cdot \exp\left(-\text{const} \cdot \sqrt{\frac{8\pi^2}{9\lambda_\sigma}}\right) \quad (14.43)$$

with constant of order 1. This determines the spatial scale of homogeneity at the end of inflation. If λ_σ is so small that this scale is larger than the visible Universe, we come back to the picture of practically homogeneous field discussed after Eq. (14.34). Otherwise the field σ is strongly inhomogeneous over the visible Universe, although the correlation length exceeds considerably the horizon size at the end of inflation.

If λ_σ is not very small (but still $\lambda_\sigma \lesssim 1$, so that the quantum theory is weakly coupled), it is quite possible that the condition (14.42) is satisfied early at inflation, but becomes violated later on. This situation is inherent in large field models, where ϵ is very small early at inflation and becomes of order 1 towards its end. In that case $\bar{\sigma}$ gets stuck at the value (14.41) at time t_f when the inequality (14.42) gets violated. Indeed, we have seen in Section 14.3.1 that for $m_{\text{eff}}^2 \ll \epsilon H^2$ the classical field ($\bar{\sigma}$ in our case) does not roll down considerably by the end of inflation, while the extra contribution of modes which enter the horizon after t_f is small compared to $\bar{\sigma}$. The homogeneity scale at time t_f is estimated according to (14.43); it stretches out by a factor $\exp[N_e(t_f)]$ towards the end of inflation.

Hence, unlike in free massive theory, the uncertainties related to the early inflationary epoch are absent for sufficiently large λ_σ . At the same time, the properties of the scalar field fluctuations depend strongly on the details of inflation towards its end and on the value of the coupling λ_σ . In particular, the field σ may be strongly inhomogeneous over the visible Universe at the end of inflation.

To end this Section, we make a comment concerning the formalism adequate for describing the fluctuating field σ . It is clear from the above discussion that the evolution of its superhorizon part $\bar{\sigma}$ is a sort of diffusion caused by “hits” due to modes which exit the horizon. The relevant formalism has indeed much in common with the diffusion theory; this formalism is presented, e.g., in the book [147].

14.4 Axion as Dark Matter Candidate: CDM Isocurvature Mode

We consider in this Section a mechanism that generates CDM isocurvature perturbations [148, 149] in a model with the oscillating axion field serving as dark matter. Unlike in models of Section 14.2.2, the CDM isocurvature mode is uncorrelated with the adiabatic mode.

The model has been studied in Section I.9.7.1. Let us recall its main properties. The dark matter energy density consistent with observations is obtained for the axion mass in the range

$$m_a = (10^{-5} - 10^{-6}) \text{ eV}.$$

The corresponding range of the Peccei–Quinn (PQ) scale is

$$f_{PQ} = (10^{11} - 10^{12}) \text{ GeV}. \quad (14.44)$$

Recall that the simple versions of the axion model assume that there exists a complex scalar field (PQ field) S whose Lagrangian is

$$\mathcal{L}_S = \partial_\mu S^* \partial^\mu S - \lambda_{PQ} \left(S^* S - \frac{f_{PQ}^2}{2} \right)^2 + \dots,$$

where omitted terms describe the interactions with other fields, including the Standard Model ones. We assume in most of this Section that the coupling λ_{PQ} is of order 1 and do not show the dependence on λ_{PQ} wherever possible (this is in fact unimportant for most of the results).

If Quantum Chromodynamics (QCD) is switched off, the theory has global symmetry $U(1)_{PQ}$ under which

$$S \rightarrow e^{i\beta} S.$$

The vacuum expectation value of the field S is non-zero, $\langle S \rangle = f_{PQ}/\sqrt{2}$. This breaks $U(1)_{PQ}$ spontaneously, and the theory possesses the Nambu–Goldstone boson, the phase of the field S . It parameterizes the field at energies below f_{PQ} , namely, $S = (f_{PQ}/\sqrt{2}) \exp[i\theta(x)]$. The canonically normalized axion field $a(x)$ is related to the phase in a way standard for the Nambu–Goldstone bosons,

$$a(x) = f_{PQ} \theta(x). \quad (14.45)$$

QCD effects give the axion a small mass⁶ at temperature $T \sim \Lambda_{QCD} \sim 200$ MeV; its vacuum value is estimated as $m_a \simeq (m_\pi f_\pi)/(2f_{PQ})$, where m_π and f_π are the pion mass and decay constant.

We described in Section I.9.7.1 the misalignment mechanism of the dark matter generation in this model. Its main property is that the phase θ is spatially homogeneous and constant in time when $T \gg \Lambda_{QCD}$. As the Universe cools down to roughly $T \sim \Lambda_{QCD}$, the axion field starts to oscillate, remaining almost homogeneous over space. This oscillating field plays the role of cold dark matter, see also Section 2.5.

It is worth mentioning that the picture for the adiabatic mode in this model is as follows. Before the beginning of oscillations, the phase θ_i is spatially homogeneous, while there are superhorizon perturbations in the hot gas of particles.

⁶The axion potential is not quadratic; it is periodic in θ with period 2π . The discussion of the corresponding corrections is beyond the scope of this book.

The axion oscillations begin when the temperature of this gas falls below a certain value, and the value of the axion field at any moment of time is in one-to-one correspondence with temperature. Hence, the axion field is homogeneous at equal-temperature hypersurfaces, which precisely means the adiabatic mode of perturbations, see Section 5.2.

Problem 14.9. *The above argument is not valid for modes which are subhorizon at the beginning of the axion oscillations. Estimate the maximum present wavelength of these modes. Are they cosmologically interesting?*

The mechanism finds its qualification in the inflationary theory. First, an assumption behind the entire misalignment mechanism is that the maximum temperature in the Universe is below f_{PQ} , otherwise the PQ symmetry would be restored at early times, and at $T \sim f_{PQ}$ the Universe would undergo the phase transition with PQ symmetry breaking. The phases $\theta(\mathbf{x})$ would be uncorrelated at length scales exceeding the Hubble size, so the assumption of the homogeneity of the initial phase θ_i would not be valid. In fact, the Kibble mechanism would give rise to the production of the axion strings, the topological extended objects. Their subsequent decays would also produce axions, but this would be another mechanism of their generation, see Refs. [150, 151] for reviews.

The inflationary theory indeed suggests that the maximum temperature in the post-inflationary Universe is not extremely high, see the discussion in Section 12.2, and, in particular, Eq. (12.30). We show in Chapter 15 that the reheating temperature T_{reh} is model dependent, so that it is quite possible that $T_{reh} < f_{PQ}$ for the PQ scale given in (14.44). Thus, the assumption under discussion is not entirely unnatural.

On the other hand, a new question arises of what happens to PQ symmetry at the reheating epoch between inflation and hot stage. We assume that PQ symmetry is broken at reheating. This is not an innocent assumption, but it is valid in many concrete realizations due to weak coupling of PQ field to other fields.

The behavior of PQ field S at inflation depends strongly on whether or not the scale f_{PQ} is small compared to the Hubble parameter. We begin with the case in which the Hubble rate is low⁷ towards the end of inflation (at the last 60 e-foldings),

$$H \ll f_{PQ}. \quad (14.46)$$

This of course restricts the class of inflationary models. We consider the opposite case towards the end of this Section.

The misalignment mechanism considered in Section I.9.7.1 does work under the above assumptions. Since the mass of the “radial” field $|S|$ is of order f_{PQ} (recall that we assume $\lambda_{PQ} \sim 1$), this field quickly, in a few Hubble times, rolls down to

⁷We assume throughout this Section that $H \gg \Lambda_{QCD}$ at inflation. This implies that QCD effects are negligible at that epoch; in particular, the axion field is massless. This assumption is valid in majority of inflationary models.

its vacuum value, and its fluctuations are negligibly small. The light field $\theta(x)$ is almost homogeneous over the visible Universe, see the discussion after Eq. (14.34). It can take any value; in view of its periodicity the initial value θ_i belongs to the interval $[0, 2\pi]$.

The phase θ is not exactly homogeneous in the visible Universe after the end of inflation, however, because of the inflationary generation of its perturbations. Since the axion field $a(x)$ has the standard kinetic term, the results of Section 13.1.2 apply to it. At interesting cosmological scales the axion field perturbations are Gaussian and have the power spectrum $\mathcal{P}_a = H^2/(2\pi)^2$, where we do not indicate the weak dependence on momentum, which is the same as in (13.27). The perturbation spectrum of the phase is

$$\mathcal{P}_\theta = \frac{H^2}{(2\pi)^2 f_{PQ}^2}, \quad (14.47)$$

while the average value is θ_i . Here, as before, H is the Hubble parameter at $N_e \sim 60$ e-foldings before the end of inflation. The perturbations of the phase $\theta(\mathbf{x})$ which are superhorizon down to $T \sim \Lambda_{QCD}$ do not evolve until the axion field starts to oscillate. Hence, at the beginning of oscillations, the axion energy density $\rho_a(x) \propto m_a^2 f_{PQ}^2 \theta^2(x)$ is spatially inhomogeneous, and its relative fluctuations are

$$\delta_a = \frac{\delta\rho_a}{\bar{\rho}_a} = 2 \frac{\delta\theta}{\theta_i}.$$

Unlike the adiabatic perturbations discussed above, these are CDM isocurvature perturbations with $\mathcal{S}_{CDM} = \delta_a$, uncorrelated with the adiabatic mode. The power spectrum of the CDM mode is given by

$$\mathcal{P}_{\mathcal{S}_{CDM}} = \mathcal{P}_{\delta_a} = \frac{H^2}{(\pi f_{PQ} \theta_i)^2}. \quad (14.48)$$

This result shows that the CDM mode may be quite strong. The observational limit on the CDM mode given in Section 5.4, $r_{CDM} \equiv \mathcal{P}_{\mathcal{S}_{CDM}}/\mathcal{P}_{\mathcal{R}} < 0.07$, implies the bound on the inflationary Hubble parameter,

$$H \lesssim \pi \theta_i f_{PQ} r_{CDM}^{1/2} \Delta_{\mathcal{R}} \quad (14.49)$$

$$\sim (10^7 - 10^8) \text{ GeV}, \quad (14.50)$$

where the numerical estimate is obtained for $f_{PQ} = 10^{11} - 10^{12}$ GeV and $\theta_i \sim \pi$.

Hence, under the assumption (14.46), the misalignment mechanism of the generation of the axion dark matter does not contradict the observational data only if inflation occurred at low rate, and hence at low energy scale,

$$E_{inf} \equiv V^{1/4} = \left(\frac{3}{8\pi} H^2 M_{Pl}^2 \right)^{1/4} \lesssim (10^{12} - 10^{13}) \text{ GeV}. \quad (14.51)$$

If the scale E_{inf} is close to the upper limit, the CDM isocurvature perturbations may be detectable.

We note that the bound (14.50) is not completely model-independent. The interaction of the field S with the inflaton field ϕ may break PQ symmetry explicitly, e.g., it may have the form $(\gamma\phi^2 S^2 + \text{h. c.})$, where γ is a complex coupling. Then the axion has large mass at inflation, $m_a^{(\text{eff})} \simeq \sqrt{|\gamma|}\phi$. The axion fluctuations are suppressed if $m_a^{(\text{eff})} \gg H$ at inflationary epoch. Another possibility is that the expectation value of $|S|$ at inflation is much greater than its vacuum value f_{PQ} . These and other possibilities are considered, e.g., in Refs. [152, 153].

The bound (14.50) can be refined by making use of the results of Section I.9.7.1. It has been shown there that the misalignment mechanism yields the correct present energy density of the axion dark matter, provided that the initial phase obeys

$$\theta_i^2 \cdot \left(\frac{4 \cdot 10^{-6} \text{ eV}}{m_a} \right)^{1.2} \simeq 1.$$

Now, the axion mass is given in terms of PQ scale as follows,

$$m_a \approx m_\pi \cdot \frac{f_\pi}{2f_{PQ}} \approx 6 \cdot 10^{-6} \text{ eV} \cdot \left(\frac{10^{12} \text{ GeV}}{f_{PQ}} \right).$$

Inserting these expressions, together with the numerical value of Δ_R , into (14.49), we find

$$H < 5 \cdot 10^7 \cdot \left(\frac{f_{PQ}}{10^{12} \text{ GeV}} \right)^{0.4} \cdot \left(\frac{r_{CDM}}{0.07} \right)^{0.5} \text{ GeV}.$$

We see that the dependence on f_{PQ} is fairly weak, so the conservative bound on the inflationary Hubble parameter is

$$H < 1 \cdot 10^8 \text{ GeV}.$$

This gives the conservative bound on the energy scale of inflation, $E_{inf} < 2 \cdot 10^{13} \text{ GeV}$.

The previous analysis, including the bounds (14.50), (14.51), is valid under the assumption (14.46). Clearly, of interest is also the opposite case

$$H \gtrsim f_{PQ}. \quad (14.52)$$

In this case the field S is light at inflation, and the amplitude of its fluctuations is large compared to the vacuum expectation value $f_{PQ}/\sqrt{2}$. To study this situation, it is convenient to consider real and imaginary parts $S_1 = \sqrt{2} \operatorname{Re} S$ and $S_2 = \sqrt{2} \operatorname{Im} S$, rather than the absolute value and phase of the field S . Once the relation (14.52) holds, the quadratic term in the potential for the field S can be neglected, and we arrive at a theory of two effectively massless scalar fields with quartic self-interaction. The results of Section 14.3.2 apply to this theory directly; the modes of the fields $S_{1,2}$ obey Eq. (14.39), where the effective mass is now given by

$$m_{\text{eff},(1,2)}^2 = \lambda_{PQ}(\bar{S}_1^2 + \bar{S}_2^2) + 2\lambda_{PQ}\bar{S}_{(1,2)}^2.$$

The analysis leading to the estimate (14.41) is valid, and modulo a factor of order one we have

$$\bar{S}_1^2 \sim \bar{S}_2^2 \sim \sqrt{\frac{1}{8\pi^2 \lambda_{PQ}}} H^2. \quad (14.53)$$

The estimates for the correlation length coincide with the estimates obtained in Section 14.3.2, see Eq. (14.43) and discussion after that formula. If the correlation length $l \sim q^{-1}$ is smaller than the size of the visible Universe at the end of inflation, the phase of the field S is strongly inhomogeneous, and the misalignment mechanism does not work. This is the case for large enough λ_{PQ} . Instead, after regions of initial size l enter the horizon, the Kibble mechanism starts to operate, and axion strings get formed, see Ref. [154] for details. So, the validity of the misalignment mechanism requires that the correlation length is very large at the end of inflation, and the field S is almost homogeneous over the visible Universe. The fluctuations of the phase are calculated in the same way as above, the only difference is that the role of PQ scale f_{PQ}^2 in (14.47) is played by the expectation value of $(\bar{S}_1^2 + \bar{S}_2^2) \equiv |\bar{S}|^2$. We repeat the analysis leading to (14.48), use (14.53) and obtain that the power spectrum of CDM isocurvature mode is given in the regime (14.52) by

$$\mathcal{P}_{S_{CDM}} \sim \frac{\sqrt{\lambda_{PQ}}}{\theta_i^2}.$$

Recalling the above bound on r_{CDM} , we conclude that the misalignment mechanism is consistent with the data in the case (14.52) only if $\lambda_{PQ} \lesssim 10^{-10}$. This is a very implausible possibility.

Hence, the study of the axion perturbations at inflation shows the possibility of the generation of sizeable admixture of the CDM isocurvature mode uncorrelated with the adiabatic mode. Consistency with observations places strong constraints on the models with the misalignment mechanism, where the role of dark matter is played by the coherently oscillating, almost homogeneous axion field. With reservations mentioned above, the least contrived possibility is the model with low inflationary scale that obeys (14.50) and (14.51). If future observations detect tensor perturbations, this possibility will be ruled out. There still remains an option that the axion field is strongly inhomogeneous by the end of inflation, and the axion dark matter is produced in decays of axion strings. This is the case for sufficiently high inflation scale, see (14.52), and large enough λ_{PQ} . The same mechanism operates if PQ phase transition occurs after the inflationary epoch, which also requires high enough inflation scale.

This page is intentionally left blank

Chapter 15

*Preheating after Inflation

The reheating epoch occurs after inflation and ends with the hot stage. In slow roll inflation models, it begins when the slow roll conditions get violated. The inflaton field rolls down its potential and then oscillates around the minimum of its potential. The energy of the oscillations is gradually reduced, and the oscillations get damped due to particle creation by the strongly non-stationary inflaton field. The main purpose of this Chapter is to study various aspects of the latter process. The stage of the decay of the inflaton oscillations and particle creation is often called *preheating*, while the term reheating is used for the entire epoch between inflation and thermal equilibrium.

We note in this regard that in multi-field models, including hybrid inflation, the oscillating field that creates particles may be different from inflaton; in notations of Section 12.2.3 this is the field χ . The processes we discuss in this Chapter are inherent in this case too, although the parameters characterizing the field χ (say, its mass) may be quite different from those of inflaton. We will be talking about inflaton for definiteness, having in mind any homogeneous scalar field whose oscillations occur right after inflation.

Particles created in the decay of the inflaton oscillations gradually come to thermal equilibrium. The equilibration process is quite complicated, and effects related to it can hardly be detected. Therefore, we do not discuss it in any detail.

15.1 Inflaton Decay in Weakly Coupled Models

It is of interest for the study of processes occurring at the hot epoch to estimate the maximum temperature of relativistic matter after inflation and the temperature at which the relativistic matter begins to dominate the cosmological expansion. We denote them by T_{max} and T_{reh} , respectively; clearly, $T_{reh} \leq T_{max}$. We leave aside the question of whether particle interactions are strong enough to thermalize the relativistic fraction: our estimates are literally valid for the energy densities ρ_{max} and ρ_{reh} of the relativistic matter, rather than for its temperatures. Yet we prefer to talk about temperatures for convenience.

The observational results quoted in Section 5.4 give a model-independent bound on the temperature T_{max} . Namely, the Friedmann equation applied to the Universe filled with the equilibrated gas of relativistic particles at temperature T and possibly other matter gives

$$\frac{\pi^2}{30} g_* T^4 \leq \frac{3}{8\pi} M_{Pl}^2 H^2.$$

The Hubble parameter decreases in time, hence this implies the bound on T_{max} ,

$$\frac{T_{max}}{M_{Pl}} < \left(\frac{90}{8\pi^3 g_*} \right)^{1/4} \left(\frac{H_e}{M_{Pl}} \right)^{1/2}, \quad (15.1)$$

where H_e is the Hubble parameter at the end of inflation. Now, $H_e < H_k$, where H_k is the Hubble parameter at 50–60 e -foldings before the end of inflation. The latter is bounded from above, $H_k < 1 \cdot 10^{-5} M_{Pl}$, see (13.68). Let us assume that the effective number of degrees of freedom g_* is not smaller than that in the Standard Model at high temperature, $g_{*SM} = 106.75$, see Section I.B.7. In this way we arrive at the model-independent bound,

$$T_{max} < 0.8 \cdot 10^{-3} M_{Pl} = 1 \cdot 10^{16} \text{ GeV}.$$

This bound is stronger in concrete inflationary models, since the limit (13.68) refers to the Hubble parameter at 50–60 e -foldings before the end of inflation, rather than to H_e . Consider, e.g., the large field model with the inflaton potential $V(\phi) = (m^2/2)\phi^2$. Inflation ends when the slow roll parameters ϵ and η become of order 1 (this happens simultaneously in the model we discuss),

$$\epsilon \equiv \frac{V'^2 M_{Pl}^2}{16\pi V^2} \simeq 1.$$

The inflaton field at this time is given by

$$\phi_e^2 \simeq \frac{M_{Pl}^2}{4\pi}, \quad (15.2)$$

so the Hubble parameter is of the order of the inflaton mass,

$$H_e^2 = \frac{8\pi}{3M_{Pl}^2} \frac{m^2}{2} \phi_e^2 \simeq \frac{m^2}{3}. \quad (15.3)$$

This is in accordance with the result of Section I.4.8 that the oscillations of a massive scalar field begin when the Hubble parameter becomes of the order of the mass. We know from Section 13.4.1 that the inflaton mass in this model is $m \simeq 10^{-6} M_{Pl}$, see (13.62). The inequality (15.1) gives now

$$T_{max} < 0.2 \cdot 10^{-3} M_{Pl} = 2 \cdot 10^{15} \text{ GeV}. \quad (15.4)$$

The bounds are at the same level in other large field models, and are even stronger in the models of Sections 12.2.2 and 12.2.3.

Problem 15.1. Find the bounds analogous to (15.4) in the models of Sections 12.2.2, 12.2.3 and in the large field model with the inflaton potential $V(\phi) = (\lambda/4)\phi^4$.

Note that these bounds are very conservative, and yet they show that the temperatures at the hot stage were never as high as the Grand Unification scale (see Section I.11.2.2) $M_{GUT} \simeq 10^{16}$ GeV.

We have seen in Section 13.4.1 that the inflaton self-coupling must be typically very small. This suggests that the inflaton is a weakly coupled field. So, it is quite possible that its couplings to all other fields are also very small. This is not necessarily true, of course; the opposite possibility is considered in forthcoming Sections. With this reservation, let us see that in the case of very weakly coupled inflaton, the maximum temperature in the Universe is much lower than the upper bound in (15.4). To illustrate the situation, consider again the large field inflationary model, now including the inflaton interactions with other particles. At the beginning of the reheating epoch, the Universe is filled with the oscillating inflaton field whose frequency is $\omega = m$ where m is the inflaton mass. This field has the effective equation of state $p = 0$, see Sections I.4.8.1 and 2.5, so the expansion law is $a \propto t^{2/3}$. A simple estimate for the maximum temperature, valid both for very weakly coupled inflaton and in a number of other cases, is obtained by treating the inflaton oscillations as a collection of particles at rest and assuming that these particles decay independently and their decay products quickly thermalize. We have seen in Section I.5.3 that the maximum temperature in this situation is reached immediately (in a few Hubble times) after the beginning of oscillations, and that the maximum energy density of the hot component is

$$\rho_{rad, max} \simeq \frac{\Gamma}{10\pi G t_e} \simeq \frac{3}{20\pi} \Gamma M_{Pl}^2 H_e,$$

where Γ is the inflaton decay width, $t_e = 2/(3H_e)$ is the time of the beginning of oscillations which we identify with the time when inflation terminates. This gives the estimate for the maximum temperature,

$$T_{max} \simeq \left(\frac{9}{2\pi^3 g_*} \cdot \Gamma H_e M_{Pl}^2 \right)^{1/4}.$$

For $\Gamma \ll H_e$, i.e., $\Gamma \ll m$ (see (15.3)), this temperature is much lower than the right hand side of (15.4). To estimate the plausible range of maximum temperatures, we write for the inflaton decay width,

$$\Gamma \sim \frac{h^2}{8\pi} N_d m, \quad (15.5)$$

where N_d is the number of decay channels, h is a typical inflaton coupling to other fields (say, the Yukawa coupling to fermions $h\phi\bar{\psi}\psi$). Assuming no hierarchy between

the parameters, we write $N_d \sim g_*$, $h \sim \sqrt{\lambda}$ where λ is the inflaton self-coupling.¹ As a result we obtain

$$T_{max} \sim \frac{1}{2\pi} (hmM_{Pl})^{1/2}, \quad (15.6)$$

where we used $H_e \sim m$. We know from Section 13.4.1 that $\lambda \lesssim 10^{-13}$, $m \lesssim 10^{-6} M_{Pl}$, and with $h \sim \sqrt{\lambda}$ we find numerically

$$T_{max} \lesssim 10^{-7} M_{Pl} \sim 10^{12} \text{ GeV}.$$

Hence, the plausible range of the maximum temperatures in the large field inflationary models is

$$T_{max} \sim 10^{12} - 10^{15} \text{ GeV}.$$

Other models typically have lower maximum temperature.

Within this simple approach to the reheating process, the temperature in the Universe decreases during the reheating epoch, see Section I.5.3. Reheating ends when $H \sim \Gamma$: most particles in the inflaton condensate decay precisely at that time, and the Universe enters the radiation domination epoch. Under the above assumptions, the temperature at that time is

$$T_{reh} \simeq (\Gamma M_{Pl}^*)^{1/2} \simeq \frac{N_d^{1/2}}{4\pi g_*^{1/4}} (h^2 m M_{Pl})^{1/2}.$$

This temperature is much lower than the maximum temperature (15.6) for small couplings h ; for $N_d \sim g_* \sim 100$, $h^2 \sim 10^{-13}$, $m \sim 10^{-6} M_{Pl}$ we have

$$T_{reh} \sim 10^{-11} M_{Pl} \sim 10^8 \text{ GeV}. \quad (15.7)$$

We see that in the case of very weakly coupled inflaton, it oscillates for a long time.

As we have already noticed, the inflaton couplings to other fields may not be as small as we have assumed.² If so, the above treatment may not be valid. In particular, if the inflaton decays into pairs of bosons, the latter have one and the same energy. Their phase space density may be large, and hence the inflaton decays may be Bose-enhanced [155–160]. Another property that may be important for both bosons and fermions is the possible dependence of the effective mass of decay products on the inflaton field. These and other aspects of preheating are discussed in the forthcoming Sections.

¹Recall that the amplitudes of scattering processes $\phi + \phi \rightarrow \phi + \phi$ and $\phi + \psi \rightarrow \phi + \psi$ are proportional to λ and h^2 , respectively. So, the assumption $h \sim \sqrt{\lambda}$ means that the inflaton interaction with other fields is of the same order as its self-interaction.

²According to the discussion in Section 12.2.4, the inflaton effective potential obtains quantum corrections due to interactions with other fields, which are large for relatively large couplings of these interactions. We assume throughout this Chapter that these corrections somehow cancel out, and the inflaton effective potential remains sufficiently flat. This is required for consistency with observations, see Section 13.4.1.

15.2 Inflaton Decay in a Model with Quadratic Potential

Let us study in more details the inflaton decay into bosons in a model with quadratic inflaton potential,

$$V(\phi) = \frac{m^2}{2}\phi^2. \quad (15.8)$$

This study is relevant not only for the large field models but also for other models at the stage when inflaton oscillates near the minimum of its potential, and the non-linearity of oscillations is negligible. In fact many (albeit not all) results of this Section are valid for other potentials; peculiarities of quartic potential $\lambda\phi^4$ are discussed in Section 15.3. Let the inflaton interact with another scalar field χ , whose action is

$$S_\chi = \int d^4x \sqrt{-g} \left(\frac{1}{2}g^{\mu\nu}\partial_\mu\chi\partial_\nu\chi - \frac{m_\chi^2}{2}\chi^2 - \frac{1}{2}g^2\chi^2\phi^2 \right), \quad (15.9)$$

where g is the dimensionless coupling which we assume to be small. We are interested in the decay of the inflaton oscillations into χ -bosons. We neglect other interactions of the field χ ; this approximation is valid if these interactions are weak, and the number density of χ -bosons is not yet so high that its interactions begin to thermalize the system. Furthermore, we neglect the back reaction of the created χ -bosons on the inflaton oscillations, i.e., we treat the inflaton field as background. These assumptions are valid at least in the beginning of the inflaton oscillations; overall, our analysis is sufficient to clarify the general picture of the inflaton decay and uncover the phenomena that may show up in this process.

In our approximation, the homogeneous inflaton field oscillates as follows,

$$\phi(t) = \phi_0(t) \cos mt,$$

where the amplitude ϕ_0 decreases as $\phi_0(t) \propto a^{-3/2}(t)$, see Section 2.5. At the epoch of oscillations, one has $H \ll m$, so the amplitude ϕ_0 is a slowly varying function in the sense that $\dot{\phi}_0/\phi_0 \ll m$. The field χ obeys the following equation, in the momentum representation,

$$\ddot{\chi} + 3H\dot{\chi} + [\varepsilon^2 + g^2\phi^2(t)]\chi = 0, \quad (15.10)$$

where

$$\varepsilon^2 = \frac{k^2}{a^2} + m_\chi^2$$

is the frequency squared of χ -oscillations, i.e., the energy squared of χ -quanta, and k is time-independent conformal momentum. The parameter ε also varies slowly, $\dot{\varepsilon}/\varepsilon \ll m$. We change variables,

$$\chi = \frac{1}{a^{3/2}}\sigma,$$

and cast Eq. (15.10) into the form

$$\ddot{\sigma} + \left(\varepsilon^2 + g^2 \phi^2 - \frac{3}{4} H^2 - \frac{3}{2} \frac{\ddot{a}}{a} \right) \sigma = 0.$$

The last two terms in parenthesis are negligible at the epoch we discuss. Indeed, the energy of each of the χ -quanta created in the inflaton decay is at least of order $m \gg H, \sqrt{|\ddot{a}/a|}$; we will see that the typical energy ε is sometimes even higher. So, we arrive at the equation

$$\ddot{\sigma} + (\varepsilon^2 + g^2 \phi_0^2 \cos^2 mt) \sigma = 0, \quad (15.11)$$

where the parameters ε and ϕ_0 vary in time adiabatically slowly.

Equation (15.11) is the equation for an oscillator with time-dependent frequency,

$$\omega_\chi^2(t) = \varepsilon^2 + g^2 \phi_0^2 \cos^2 mt. \quad (15.12)$$

At the classical level, this equation describes the parametric excitation of the oscillator. At the quantum level, this is the process of χ -particle creation in the background time-dependent inflaton field. We use mainly quantum field theory language, and speak about particle creation. The general method of studying particle creation in time-dependent backgrounds is described in Appendix E; its key element is the calculation of the Bogoliubov coefficients.

We assume here that the average field χ is zero by the end of inflation. However, within the picture of chaotic beginning of inflation outlined in Section 12.2.1 one would think that the initial value of the field χ is large in the model (15.8), (15.9). Nevertheless, under fairly mild assumptions, the field χ rolls down to the minimum $\chi = 0$ of its potential long before the end of inflation. So, the situation we discuss is fully realistic. Indeed, for large values of ϕ and χ , the total scalar potential is dominated by their interaction. So, the slow roll equations for these fields read

$$\begin{aligned} 3H\dot{\phi} &= -g^2 \chi^2 \phi \\ 3H\dot{\chi} &= -g^2 \phi^2 \chi. \end{aligned}$$

If the initial condition is such that $\phi > \chi$ (strong inequality is not required), the force acting on χ is greater than the force acting on ϕ , and χ rolls down faster than ϕ . Hence, χ reaches zero well before the end of inflation.

Problem 15.2. *By analyzing the complete system of equations for the homogeneous fields ϕ and χ (including both mass terms) find the range of parameters in the model (15.8), (15.9) and the range of initial data for ϕ and χ , in which the field χ reaches zero well before the end of inflation. [Hint: Assume that the potential energy density at the beginning of inflation is of order $V(\phi, \chi) \sim M_{Pl}^4$.] Consider the fluctuations of χ generated at inflation via the mechanism of Section 14.3.1 and find the parameter range where they are negligible by the end of inflation.*

15.2.1 Decay of large amplitude oscillations

The inflaton oscillation amplitude at the early preheating epoch may well be so large that the parameters in Eq. (15.12) obey

$$g\phi_0 \gg m. \quad (15.13)$$

As an example, in the large field inflation model with the potential (15.8) the initial oscillation amplitude is estimated as $\phi_0 \simeq M_{Pl}/\sqrt{4\pi}$, see (15.2), while the inflaton mass is much smaller, $m \simeq 10^{-6}M_{Pl}$. Hence, the inequality (15.13) is satisfied for $g \gg 3 \cdot 10^{-6}$.

In the case (15.13), the dynamics of the χ -boson creation [161, 162] is quite different from that outlined in Section 15.1. Let us consider this case, assuming that the mass m_χ is negligible. The class of processes studied here and in Section 15.3 is called *wide resonance* in literature; for reasons that become clear later, the subclass considered in this Section (but not in Section 15.3) is dubbed *stochastic resonance*. There are two properties of wide resonance which are of particular interest. First, in the naive picture, χ -particles can be created only if their mass does not exceed the inflaton mass. In fact, as we show in this Section, the creation of particles of much larger masses is possible. Second, the amplitude of the coherent field χ may be quite large. Some possible physical applications of these properties are considered in Sections 15.5.1 and 15.5.2.

Once the inequality (15.13) is satisfied, the evolution of the frequency (15.12) is adiabatic most of the time. The reason is that the oscillation frequency of the field χ is almost always high compared to the inflaton oscillation frequency,

$$\omega_\chi \gg m. \quad (15.14)$$

Hence, the adiabaticity parameter is

$$\left| \frac{\dot{\omega}_\chi}{\omega_\chi^2} \right| = \left| \frac{g^2 \phi_0^2 m \cos mt \sin mt}{\omega_\chi^3} \right| \lesssim \left| \frac{m}{g\phi_0} \frac{\sin mt}{\cos^2 mt} \right|. \quad (15.15)$$

This parameter is indeed small at all t except for narrow time intervals near $t_n = (\pi/2 + \pi n)/m$, $n = 0, 1, \dots$, when the inflaton field vanishes. Hence, the creation of χ -bosons occurs only during these intervals, while the number of χ -bosons is conserved in comoving volume between these intervals. Consider particle creation at the first of these intervals, assuming that χ -bosons are absent initially. As described in Section E.1, the number of created particles in that case is determined by the Bogoliubov coefficient β . To calculate it, we study Eq. (15.11) near $t_0 = \pi/(2m)$. We make use of the approximation $\cos(\pi/2 + m\tau) = -m\tau$, valid at small $\tau = t - t_0$, and write this equation in the following form,

$$\ddot{\sigma} + (\varepsilon^2 + g^2 \dot{\phi}^2 \tau^2) \sigma = 0, \quad (15.16)$$

where

$$\dot{\phi}^2 \equiv \dot{\phi}^2(\tau = 0) = m^2 \phi_0^2.$$

According to Section E.1, we have to solve this equation with the initial condition that σ is a positive-frequency function at large *negative* τ (formally, as $\tau \rightarrow -\infty$),

$$\sigma^{(+)} = \frac{1}{\sqrt{2\omega_\chi}} e^{i \int \omega_\chi d\tau}, \quad \tau \rightarrow -\infty.$$

The solution behaves at large *positive* τ as follows,

$$\sigma^{(+)} = \alpha \cdot \frac{1}{\sqrt{2\omega_\chi}} e^{i \int \omega_\chi d\tau} + \beta^* \cdot \frac{1}{\sqrt{2\omega_\chi}} e^{-i \int \omega_\chi d\tau}, \quad \tau \rightarrow +\infty, \quad (15.17)$$

where β is precisely the Bogoliubov coefficient we are after.

Equation (15.16) is formally equivalent to the stationary Schrödinger equation with “coordinate” τ for a particle of “mass” $1/2$ with “energy” ε^2 in the “potential”

$$U(\tau) = -g^2 \phi_0^2 m^2 \tau^2. \quad (15.18)$$

In this quantum mechanical analog, the coefficient β in (15.17) corresponds to the over-barrier reflection off negative quadratic potential (note that the “energy” is positive). The solution to this quantum mechanical problem is well-known, see the book [61]. We recall it here, adapting the derivation to the particle creation language. The frequency at large negative τ is

$$\omega_\chi = \sqrt{\varepsilon^2 + g^2 \phi_0^2 m^2 \tau^2} \approx -\left(g\phi_0 m \tau + \frac{\varepsilon^2}{2g\phi_0 m \tau}\right),$$

so the positive frequency function has the following form,

$$\sigma^{(+)} = \frac{1}{2g\phi_0 m} \cdot z^{-\frac{1}{2} - i \frac{\varepsilon^2}{2g\phi_0 m}} \cdot e^{-\frac{i}{2} g\phi_0 m z^2}, \quad (15.19)$$

where $z = -\tau \rightarrow +\infty$. The negative frequency part in (15.17) (the second term) is obtained by the analytical continuation of the function (15.19) in the lower half-plane to the region $z = \rho e^{-i\pi}$, $\rho > 0$, which gives the desired result

$$\beta^* = -i \exp\left(-\frac{\pi\varepsilon^2}{2g\phi_0 m}\right). \quad (15.20)$$

It shows that the number density in the physical momentum volume $d^3 q$ right after the creation event is, see (E.22),

$$n(q)d^3 q = |\beta(q)|^2 \frac{d^3 q}{(2\pi)^3} = \exp\left(-\frac{\pi\varepsilon^2}{g\phi_0 m}\right) \frac{d^3 q}{(2\pi)^3}. \quad (15.21)$$

Remarkably, the energies of the created particles are as high as

$$\varepsilon \sim \sqrt{g\phi_0 m}, \quad (15.22)$$

which is much greater than the inflaton mass. This property is due to the fact that at the time the inflaton field crosses zero, the adiabaticity parameter $\dot{\omega}_\chi/\omega_\chi^2$ is of order 1 for particles of this energy, see (15.15).

Problem 15.3. Estimate the maximum value of the adiabaticity parameter $\dot{\omega}_\chi/\omega_\chi^2$ for $\varepsilon \sim \sqrt{g\phi_0 m}$ and $\varepsilon \gg \sqrt{g\phi_0 m}$.

Problem 15.4. Find the relationship between the Bogoliubov coefficients α and β entering (15.17) and the transmission and reflection coefficients in quantum mechanics of a particle in the potential (15.18). What is the quantum mechanical analog of the identity $|\alpha|^2 - |\beta|^2 = 1$ known from Section E.1?

For $m_\chi \ll \sqrt{g\phi_0 m}$, the total number density of χ -bosons after the creation event is

$$n_\chi = \int d^3 q n(q) = \frac{1}{8\pi^3} (g\phi_0 m)^{3/2}. \quad (15.23)$$

As the inflaton field grows from nearly zero to $\phi \sim \phi_0$, the energy of each χ -particle increases towards $\omega_\chi \sim g\phi_0$, and the energy density is estimated as

$$\rho_\chi = n_\chi \omega_\chi \sim n_\chi g\phi_0 \sim \frac{1}{8\pi^3} \left(\frac{g^5 \phi_0}{m} \right)^{1/2} m^2 \phi_0^2. \quad (15.24)$$

We see that at large enough coupling g , the energy density of the created particles may be sizeable compared to the inflaton energy density $\rho_\phi = m^2 \phi_0^2 / 2$ and may even become of the order of the latter. The back reaction of the created particles on the inflaton field cannot be neglected for $\rho_\chi \sim \rho_\phi$, but it is clear that the inflaton oscillation amplitude gets strongly damped in one or a few oscillations in that case.

We studied until now the creation of χ -bosons at the *first* event when the inflaton crosses zero. If the created particles have low energy density, $\rho_\chi \ll \rho_\phi$, the inflaton oscillations proceed with almost the same amplitude as before. Further dynamics depends on the properties of χ -particles. If they have large probability to scatter or decay in half period of the oscillations, then the details of the previous creation process are forgotten by the next time the inflaton crosses zero, and the particle creation proceeds without much change. In the opposite case, the mode σ evolves according to Eq. (15.11); then the χ -boson creation is stochastic in a certain sense. The number of particles grows exponentially,

$$n_k(t) = c_k \exp\left(\frac{\nu_k m t}{\pi}\right), \quad (15.25)$$

where the coefficients ν_k and c_k depend on conformal momentum k in an irregular way. These coefficients are of order 1 for $\beta(k) \sim 1$.

If χ -particle scattering and decay are negligible, their creation occurs as a coherent process. To describe it, we continue solving Eq. (15.11) for the function $\sigma^{(+)}$ and find the Bogoliubov coefficients after subsequent inflaton zero crossings. Let $\sigma^{(+)}$ before the n -th zero crossing (but at the time when the adiabatic approximation still works for Eq. (15.11)) be

$$\sigma^{(+)}(t) = \alpha_n e^{i \int_{t_n}^t \omega_\chi dt} + \beta_n e^{-i \int_{t_n}^t \omega_\chi dt}, \quad (15.26)$$

where t_n is the time of zero crossing, and we omit the irrelevant factor $(2\omega_\chi)^{-1/2}$. Then immediately *after* the n -th crossing it is given by

$$\sigma^{(+)}(t) = \alpha_n(\alpha e^{i \int_{t_n}^t \omega_\chi dt} + \beta^* e^{-i \int_{t_n}^t \omega_\chi dt}) + \beta_n(\alpha^* e^{-i \int_{t_n}^t \omega_\chi dt} + \beta e^{i \int_{t_n}^t \omega_\chi dt}),$$

where β is the known Bogoliubov coefficient (15.20) and α is the second Bogoliubov coefficient entering (15.17). After half period of inflaton oscillations, just before the $(n+1)$ -th zero crossing, the solution $\sigma^{(+)}$ has the form similar to (15.26), but with the coefficients

$$\alpha_{n+1} = (\alpha_n \alpha + \beta_n \beta) e^{i \theta_n}, \quad (15.27)$$

$$\beta_{n+1} = (\beta_n \alpha^* + \alpha_n \beta^*) e^{-i \theta_n}, \quad (15.28)$$

where

$$\theta_n = \int_{t_n}^{t_{n+1}} \omega_\chi dt.$$

Equations (15.27) and (15.28) determine the evolution of the Bogoliubov coefficients. The phase θ_n entering these equations is a random function of n and k . Indeed, for the interesting case $\varepsilon \ll g\phi_0$, see (15.22) and (15.13), this phase is estimated as

$$\theta_n \simeq \frac{2}{m} g\phi_0(t_n),$$

where we recalled that the half period of oscillations is

$$t_{n+1} - t_n = \pi/m. \quad (15.29)$$

Since $\phi_0(t) \propto a^{-3/2}(t)$, the phase changes in half period by

$$\theta_{n+1} - \theta_n \simeq \frac{2\pi}{m^2} g\dot{\phi}_0 \sim -3\pi \frac{g\phi_0}{m} \frac{H}{m}.$$

If the parameter H/m is not particularly small, which is the case at early preheating, the absolute value of this change may be much greater than π , which means that θ_n takes random values. As a result, the phases of the Bogoliubov coefficients α_n and β_n depend irregularly on both n and conformal momentum.

This stochasticity of the coefficients α_n and β_n is reflected by the number of particles, which is proportional to $|\beta_n|^2$. It follows from Eqs. (15.27) and (15.28) that $|\beta_n|^2$ obey

$$|\beta_{n+1}|^2 - |\beta_n|^2 = 2|\beta_n|^2 |\beta|^2 + 2|\beta_n| |\beta| \sqrt{(1 + |\beta_n|^2) \cdot (1 + |\beta|^2)} \cos(\psi_n - \psi) + |\beta|^2, \quad (15.30)$$

where $\psi_n = \text{Arg}(\alpha_n^* \beta_n)$ and $\psi = \text{Arg}(\alpha^* \beta)$. When deriving this equation, we used the identities, see (E.21),

$$|\alpha|^2 - |\beta|^2 = 1 \quad (15.31)$$

$$|\alpha_n|^2 - |\beta_n|^2 = 1. \quad (15.32)$$

Since the right hand side of Eq. (15.30) contains the random factor $\cos(\psi_n - \psi)$, the number of χ -particles can either increase or decrease at each crossing of zero, and the evolution of this number is irregular. This is the origin of the term “stochastic resonance”. This is indeed a resonance, since the number of particles exponentially grows in time; the phenomenon we discuss is thus similar to parametric resonance, see Section 15.2.2. Indeed, the right hand side of Eq. (15.30) is proportional to $|\beta_n^2|$ at large β_n , with positive *on average* proportionality coefficient. Recalling (15.29), we see that the number of particles is indeed given by (15.25). The dependence of ν_k on conformal momentum is due to the fact that $\cos(\psi_n - \psi)$ is different for different modes, see Ref. [162] for details.

Problem 15.5. Making use of Eqs. (15.27), (15.28), show that the identity (15.31) yields (15.32) for all n .

We see from (15.20) that the occupation numbers $N_{\mathbf{q}} = |\beta(q)|^2$ are smaller than 1 at the initial stage of the process (here we consider the system in large but finite spatial volume and define the occupation number $N_{\mathbf{q}}$ in the same way as in Section E.1). At later stages, the occupation numbers become large. In that case, the analysis can be carried out within *classical* field theory. Indeed, quantum nature of the creation and annihilation operators $A_{\mathbf{q}}^\dagger, A_{\mathbf{q}}$ is reflected by their non-vanishing commutator $[A_{\mathbf{q}}, A_{\mathbf{q}}^\dagger] = 1$. However, if the average value $\langle A_{\mathbf{q}}^\dagger A_{\mathbf{q}} \rangle = N_{\mathbf{q}}$ is much greater than 1, then the commutator is negligible as compared to this average value, so $A_{\mathbf{q}}$ can be treated as classical Gaussian random field with zero average value and fluctuation $\langle A_{\mathbf{q}}^* A_{\mathbf{q}} \rangle = N_{\mathbf{q}}$. Hence, the quantum field $\chi(\mathbf{x}, t)$ is effectively the classical Gaussian field. According to Section E.1, the particle number density and energy density are given in terms of this field as follows,

$$n_\chi \sim \omega_\chi \langle \chi^2(\mathbf{x}) \rangle, \quad \rho_\chi \sim \omega_\chi^2 \langle \chi^2(\mathbf{x}) \rangle,$$

where ω_χ is a typical energy of χ -particle (concerning exact expressions, see problem C.22).

As we point out in Section C.2, any linear quantum field whose initial state is vacuum can *always* be treated as classical Gaussian random field, so the above discussion is not particularly interesting in the linearized theory. However, it becomes very useful for studying the non-linear evolution. The idea [164] is that one can start describing the system in terms of random classical fields at the linear stage, when the occupation numbers $N_{\mathbf{q}}$ are already large, but non-linear effects are still negligible. This linear random field then serves as the initial condition for further classical evolution. This evolution eventually enters the non-linear regime, and can be analyzed by solving non-linear, but classical field equations. The latter problem is tractable numerically. The classical solution obtained in this way corresponds to the non-linear evolution of *quantum* fields with full account of particle scattering, back reaction on the inflaton, etc.; all Bose factors are taken care of automatically. Of course, this approach has its limits of applicability, which we do not discuss here. It is worth noting that this classical field theory approach is in fact quite general, see also the discussion in Section I.10.3.

If scattering and decays of χ -bosons are negligible, the quasi-linear regime³ lasts in the model (15.8), (15.9) until the energy of the field χ becomes comparable to the energy of inflaton oscillations. It is clear from (15.25) that the duration of the quasi-linear regime is equal to π/m modulo a logarithmic factor. Hence, the initial stage of the decay of the inflaton oscillations is quite short. Towards the end of this stage, the amplitude of the field χ can be estimated by equating its energy density $\omega_\chi^2 \chi^2 \simeq g^2 \phi_0^2 \chi^2$, to the inflaton energy density $m^2 \phi_0^2/2$. We find

$$\chi^2 \sim \frac{m^2}{g^2}. \quad (15.33)$$

We see that the contribution of the field χ to the effective inflaton mass squared $m_{eff}^2 = m^2 + g^2 \chi^2$ becomes of order m^2 at that time. This is another manifestation of the fact that the back reaction of the field χ becomes non-negligible at the

³We call this regime quasi-linear, since the evolution of the field χ is linear, but it is affected by the interaction with the inflaton; it is this interaction that is responsible for the χ -particle creation.

end of the quasi-linear stage. The number density of χ -particles of all momenta is estimated as

$$n_\chi \sim \omega_\chi \chi^2 \sim \frac{m^2 \phi_0}{g}. \quad (15.34)$$

We note that these estimates are also valid at fairly large coupling, when the particle creation is rapid and the quasi-linear regime lasts a few periods of the inflaton oscillations, see the discussion of the formula (15.24). In any case, the amplitude of the field χ may become even larger at the non-linear stage, as compared to the estimate (15.33). The non-linear evolution and subsequent thermalization are very complex and model-dependent phenomena, whose analysis is beyond the scope of this book. We only note that the final thermalization proceeds through the energy transfer from low- to high-momentum modes. It is a slow process that occurs in the regime of turbulence [163].

15.2.2 Intermediate coupling

In a fairly narrow class of inflationary models with small field and/or weak coupling, of importance is a phenomenon [157, 158], similar to parametric resonance known from classical mechanics. In the context of preheating it is called *narrow resonance*. In quantum field theory language, it is the Bose-enhancement of perturbative particle creation. In the model (15.8), (15.9), χ -particles are produced in processes $\phi + \phi \rightarrow \chi + \chi$, where ϕ is the inflaton particle at rest. We will see that the narrow resonance occurs when the pair creation is kinematically allowed, $m_\chi < m$, and at the same time

$$g^2 \phi_0^2 \ll m^2. \quad (15.35)$$

In that case the energy of each of the created χ -particles is close to m . The Bose-enhancement of the creation process is due to the presence of χ -particles created earlier.

Essentially the same effect exists in a model where the action for the field χ is

$$S_\chi = \int d^4x \sqrt{-g} \left(\frac{1}{2} g^{\mu\nu} \partial_\mu \chi \partial_\nu \chi - \frac{m_\chi^2}{2} \chi^2 - \frac{1}{2} h \phi \chi^2 \right), \quad (15.36)$$

where h is a small parameter of dimension of mass. Here the inflaton oscillations get damped because of decays $\phi \rightarrow \chi + \chi$. Instead of (15.11), we have the following equation for the modes of the field χ ,

$$\ddot{\sigma} + (\varepsilon^2 + h\phi_0 \cos mt) \sigma = 0, \quad (15.37)$$

where $\phi_0 \propto a^{-3/2}$ is again a slowly varying amplitude of inflaton oscillations, and $\varepsilon^2 = k^2/a^2 + m_\chi^2$. The condition similar to (15.35) is

$$h\phi_0 \ll m^2. \quad (15.38)$$

The decay $\phi \rightarrow \chi + \chi$ is kinematically allowed for $m_\chi < m/2$.

Since ϕ_0 evolves slowly, Eqs. (15.11) and (15.37) are basically the same: in view of the relation $\cos^2 mt = \frac{1}{2}(1 + \cos 2mt)$, they coincide modulo notations. We consider Eq. (15.37) for definiteness; our results are straightforwardly translated to the model (15.9). For time-independent ε and ϕ_0 , Eq. (15.37) is nothing but the Mathieu equation, while the adiabatic dependence of ε and ϕ_0 on time is not a serious problem. The solutions to the Mathieu equation in the regime (15.38) have the following properties, see, e.g., Ref. [165]. In a narrow interval of energies near $\varepsilon = m/2$, such that

$$\left| \varepsilon - \frac{m}{2} \right| \leq \frac{\hbar\phi_0}{2m}, \quad (15.39)$$

there are exponentially growing solutions,

$$\sigma = e^{\int \nu dt} \cos \left[\frac{mt}{2} + \varphi(t) \right], \quad (15.40)$$

where

$$\nu(t) = \sqrt{\frac{\hbar^2 \phi_0^2(t)}{4m^2} - \left[\varepsilon(t) - \frac{m}{2} \right]^2},$$

and the phase $\varphi(t)$ is irrelevant for our purposes. There are other exponentially growing solutions near $\varepsilon = (m/2) \cdot n$, $n = 2, 3, \dots$, which can be interpreted as multiparticle processes $\phi + \phi + \dots \rightarrow \chi + \chi$, but these solutions grow much slower and hence are not interesting for us. It follows from the invariance of Eq. (15.37) under reflection of time, $t \rightarrow -t$, that there are also exponentially decaying solutions in the region (15.39),

$$\sigma = e^{-\int \nu dt} \cos \left(\frac{mt}{2} - \varphi \right). \quad (15.41)$$

As the Universe expands, a mode of a given conformal momentum enters the interval (15.39) from above, and then exits from this interval. The function of interest, $\sigma^{(+)}$, is a linear combination of the solutions (15.40) and (15.41), and in the resonance region it grows,

$$\sigma^{(+)} = c \cdot e^{\int \nu dt} \cos \left[\frac{mt}{2} + \varphi(t) \right],$$

where the complex parameter c is determined by matching to the positive frequency solution at early times. This immediately gives the estimate for the number of χ -particles created in a mode of conformal momentum k ,

$$|\beta|^2 = \text{const} \cdot e^{2 \int_{t_-}^{t_+} \nu(t) dt}, \quad (15.42)$$

where t_- and t_+ are the moments of time when the mode enters and exits the interval (15.39).

Problem 15.6. Find the approximate solution $\sigma^{(+)}$, including the pre-exponential factor, to Eq. (15.37) with slowly varying $\varepsilon(t)$ and $\phi_0(t)$ in the regime (15.38). Refine the estimate (15.42) in this way. Hint: The non-trivial part of evolution is when $|\varepsilon - m/2| \ll m$. Outside this region, the oscillating term in Eq. (15.37) is negligible, and the solution is a linear combination of positive- and negative-frequency WKB solutions to the equation

$$\ddot{\sigma} + \varepsilon^2 \sigma = 0. \quad (15.43)$$

The solution in question is positive-frequency at early times. The solution in the region $|\varepsilon - m/2| \ll m$ can be searched for in the following form,

$$\sigma^{(+)} = a_0(t)e^{imt/2} + b_0(t)e^{-imt/2} + a_1(t)e^{i3mt/2} + b_1(t)e^{-i3mt/2} + \dots,$$

where $a_0(t), \dots, b_1(t)$ evolve slowly, $\dot{a}_0/a_0 \ll m$ etc. Prove that the coefficients of higher harmonics a_1, b_1 , etc., are small compared to a_0 and b_0 . Using this observation, derive a closed system of equations for $a_0(t)$ and $b_0(t)$. The latter equations can be solved in the WKB approximation, and useful analogy is quantum mechanics of a particle penetrating the potential barrier. The complete solution is constructed by matching, in the region $g\phi_0 \ll |\varepsilon - m/2| \ll m$, the solutions obtained in this way with the WKB solutions to Eq. (15.43).

Let us find the exponent in (15.42). Let us assume for simplicity that m_χ and m are not degenerate, i.e., $m_\chi \ll m$, so we can set $m_\chi = 0$. As the mode k passes through the resonance zone, the dependence of ϕ_0 and the Hubble parameter on time can be neglected. Hence, the exponent is

$$2 \int_{t_-}^{t_+} \nu(t) dt = \frac{h\phi_0}{m} \int_{t_-}^{t_+} \sqrt{1 - z^2(t)} dt = \frac{h\phi_0}{m} \int_1^{-1} \sqrt{1 - z^2} \frac{dz}{\dot{z}},$$

where

$$z = \frac{2m\varepsilon(t) - m^2}{h\phi_0}.$$

Since $\varepsilon = k/a(t)$, we have $\dot{z} = -2m\varepsilon H/(h\phi_0)$, and in the resonance zone $\dot{z} = -m^2 H/(h\phi_0)$. This gives

$$2 \int_{t_-}^{t_+} \nu(t) dt = \frac{\pi}{2} \frac{h^2 \phi_0^2}{m^3 H}.$$

We consider the situation where the energy density of the inflaton oscillations dominates in the Universe. This energy density is equal to $m^2 \phi_0^2/2$, so that

$$H = \sqrt{\frac{4\pi}{3}} \frac{m\phi_0}{M_{Pl}}.$$

We obtain finally

$$2 \int_{t_-}^{t_+} \nu(t) dt = \sqrt{\frac{3\pi}{16}} \left(\frac{h\phi_0}{m^2} \right)^2 \frac{M_{Pl}}{\phi_0}. \quad (15.44)$$

We recall that the amplitude of oscillations ϕ_0 is to be taken at the time when the mode of conformal momentum k has energy $\varepsilon = k/a = m/2$.

It follows from (15.44) that χ -particle creation is most efficient in the beginning of the inflaton oscillations, when ϕ_0 is large. The exponent (15.44) substantially exceeds 1 and at the same time the inequality (15.38) is satisfied only in models with $m^2 \ll hM_{Pl}$, and only when the amplitude of oscillations obeys

$$\frac{m^4}{hM_{Pl}} \ll h\phi_0 \ll m^2. \quad (15.45)$$

Hence, the narrow resonance regime occurs only in a rather special range of parameters. If the opposite inequality holds, $\phi_0 \ll m^4/(h^2 M_{Pl})$ (and $\phi_0 \ll m^2/h$), inflaton decay proceeds without the Bose-enhancement, i.e., in the regime studied in Section 15.1.

Problem 15.7. *In the limit of small coupling h , find the number density of created χ -particles by making use of the Bogoliubov transformation method. Consider the time interval obeying $H^{-1} \gg \Delta t \gg m^{-1}$. Show that this number density is $\Delta n_\chi = \Gamma n_\phi \Delta t$, where n_ϕ is the number density of ϕ -particles in the inflaton condensate, $n_\phi = m\phi_0^2/2$, and Γ is the width of the decay $\phi \rightarrow \chi + \chi$. This reiterates the last statement in the text.*

Problem 15.8. *In the case (15.38), obtain the result on the exponential Bose-enhancement of χ -particle creation and estimate the exponent by quantum field theory methods. Hints: Use the analogy with induced photon emission, see Ref. [31]. Recall that the inflaton decays into two χ -particles of opposite 3-momenta.*

Problem 15.9. *Consider the model (15.9). Find the exponent in the number of created particles in the regime (15.35). Show that this exponent can substantially exceed 1 only for $\phi_0 \ll M_{Pl}$. Find the analogs of inequalities (15.45).*

15.3 Peculiarities of ϕ^4

Let us consider for completeness the large field inflationary model with the inflaton potential

$$V(\phi) = \frac{\lambda}{4}\phi^4. \quad (15.46)$$

We assume for the time being that the inflaton mass term is absent. There is a possibility that the inflaton oscillations create the quanta of the inflaton field itself. However, we consider another possibility, namely the decay of the inflaton oscillations due to interaction with another field χ whose action is given by (15.9).

At $m_\chi = 0$, this two-field theory has conformal symmetry in the Minkowski space-time. In particular, the action is invariant under space-time dilatations supplemented by field transformations,

$$\phi(x^\mu) \rightarrow \alpha\phi(\alpha x^\mu), \quad \chi(x^\mu) \rightarrow \alpha\chi(\alpha x^\mu).$$

The peculiarities of the decay of the inflaton oscillations [166] are traced back to this symmetry, even though it is not exact in the expanding Universe.

We note in the first place, that the inflaton oscillations occur in the regime

$$\lambda\phi_0^2 \gg H^2, |\dot{H}|, \quad (15.47)$$

where ϕ_0 is the amplitude of oscillations.

Problem 15.10. Show that in the opposite case, $\lambda\phi^2 \ll H^2$, inflaton does not oscillate and evolves in the slow roll regime.

To describe the inflaton oscillations, one conveniently introduces the field $\tilde{\phi}$ by

$$\phi = \frac{\tilde{\phi}}{a(t)} \quad (15.48)$$

and works with conformal time η . Then the field equation for the homogeneous inflaton is

$$\tilde{\phi}'' - \frac{a''}{a}\tilde{\phi} = -\lambda\tilde{\phi}^3,$$

where prime again denotes the derivative with respect to η . It follows from (15.47) that the second term in the left hand side of this equation is negligible as compared to the right hand side, so the field $\tilde{\phi}$ obeys the same equation as the homogeneous field in the Minkowski space-time,

$$\tilde{\phi}'' = -\lambda\tilde{\phi}^3. \quad (15.49)$$

This equation describes non-linear oscillations in quartic potential with constant amplitude $\tilde{\phi}_0$. The change of variables $\eta = \zeta/(\sqrt{\lambda}\tilde{\phi}_0)$ yields the equation with no parameters for the function $\tilde{\phi}/\tilde{\phi}_0$ that oscillates with unit amplitude. Hence, the period of oscillations in conformal time η is given by

$$T_\phi = \frac{C}{\sqrt{\lambda}\tilde{\phi}_0}, \quad (15.50)$$

where C is a numerical constant, which turns out to be equal to $C = \sqrt{2\pi}\Gamma(1/4)/\Gamma(3/4) \approx 7.4$. It is important that in the absence of the back reaction of χ -particles on the inflaton oscillations, neither the amplitude of oscillation $\tilde{\phi}_0$ nor their period depend on the scale factor, and hence on time. The reason behind this property is the approximate conformal invariance of this model. The amplitude of the original inflaton field ϕ decays as a^{-1} , as opposed to $a^{-3/2}$ for the mass dominated oscillations.

Problem 15.11. Find the solution to Eq. (15.49) in terms of elliptic functions. Show that the constant C is indeed equal to 7.4.

Problem 15.12. Show that the energy density of inflaton oscillations decreases as a^{-4} . Find the energy-momentum tensor averaged over a period and show that $p = \rho/3$. Hence, from the viewpoint of the cosmological expansion, the oscillating inflaton field behaves like relativistic matter. This property is also due to the approximate conformal symmetry.

To study the creation of χ -particles, let us change the variables in a way analogous to (15.48), namely, $\chi = \tilde{\chi}/a$. Then the equation for a mode of conformal momentum k reads

$$\tilde{\chi}'' + (k^2 + m_\chi^2 a^2(\eta) + g^2 \tilde{\phi}^2) \tilde{\chi} = 0,$$

where we again neglected the term proportional to a''/a . Let us assume that the coupling g is relatively large,

$$g^2 \gg \lambda. \quad (15.51)$$

This is quite a realistic assumption, since the coupling λ must be very small, see (13.63). Then the mechanism of particle creation is in some respect similar to that studied in Section 15.2.1. Indeed, away from zero value of the inflaton field, the conformal frequency squared of the field $\tilde{\chi}$ is estimated as $\Omega_\chi^2 \sim k^2 + m_\chi^2 a^2 + g^2 \tilde{\phi}_0^2$, and recalling (15.50) we have (cf. (15.14)) $\Omega_\chi \gg T_\phi^{-1}$. Hence, the mode of the field χ is almost always in the adiabatic regime. Adiabaticity breaks down when $\tilde{\phi}$ is close to zero. In the latter case, the equation for $\tilde{\chi}$ has the form (15.16) with an obvious change of notations, and the role of the parameter $\dot{\phi}$ is now played by $\tilde{\phi}'$ calculated at the time when $\tilde{\phi} = 0$. We find the latter parameter by making use of the integral of motion for Eq. (15.49),

$$\frac{1}{2}(\tilde{\phi}')^2 + \frac{\lambda}{4}\tilde{\phi}^4 = \text{const.}$$

We obtain

$$|\tilde{\phi}'|_{\tilde{\phi}=0} = \sqrt{\frac{\lambda}{2}}\tilde{\phi}_0^2.$$

The calculation yielding (15.20) is reproduced word by word, modulo the change of notations. The result for the Bogoliubov coefficient is

$$\beta^* = -i \exp\left(-\frac{\pi(k^2 + m_\chi^2 a^2)}{g\sqrt{2\lambda}\tilde{\phi}_0^2}\right) = -i \exp\left(-\frac{\pi\varepsilon^2}{g\sqrt{2\lambda}\phi_0^2}\right), \quad (15.52)$$

where $\varepsilon^2 = k^2/a^2 + m_\chi^2$. We see that efficient particle creation occurs up to high momenta and masses,

$$\frac{k}{a}, \quad m_\chi \sim \lambda^{1/4} g^{1/2} \phi_0.$$

This relation is the analog of (15.22).

Until this point our analysis repeated that of Section 15.2.1. The peculiarity of the model we consider here is the subsequent evolution of the Bogoliubov coefficients in the situation where scattering and decays of χ -particles are negligible, and the mass m_χ is small. In that case, the oscillation frequencies of $\tilde{\phi}$ and $\tilde{\chi}$ are constant in conformal time. Hence the evolution of modes of the field χ is not stochastic. Indeed, for $m_\chi = 0$, the change of the Bogoliubov coefficients in half period of the inflaton oscillations is given by Eqs. (15.27), (15.28) with

$$\theta_n = \int_{nT_\phi/2}^{(n+1)T_\phi/2} \Omega_\chi d\eta \equiv \theta, \quad (15.53)$$

where $\Omega_\chi^2(\eta) = k^2 + g^2 \tilde{\phi}^2(\eta)$. Unlike in Section 15.2.1, this phase is not a random function of n and conformal momentum, and in the absence of the back reaction of created χ -particles on the inflaton oscillations it is independent of n . Equations (15.27), (15.28) describe fully deterministic evolution of the Bogoliubov coefficients, and hence the number of particles in each mode.

Let us find the number of particles in a given mode in the situation where this number is already large, and the Bogoliubov coefficients obey

$$|\alpha_n| = |\beta_n|.$$

The latter relation follows from the general identity $|\alpha_n|^2 - |\beta_n|^2 = 1$ and $|\beta_n| \gg 1$. Since $\theta_n = \theta$ is independent of n , the solution to Eqs. (15.27), (15.28) is

$$\begin{aligned} \alpha_n &= e^{i\varphi} c_n, \\ \beta_n &= e^{i(\varphi+\psi)} c_n, \end{aligned}$$

where the amplitudes c_n are real, and the phases φ and ψ are independent of n . Indeed, we insert this Ansatz into (15.27) and obtain two equations,

$$\begin{aligned} |c_{n+1}| &= e^\nu |c_n|, \\ \text{Im}[(\alpha + e^{i\psi} \beta) e^{i\theta}] &= 0, \end{aligned} \quad (15.54)$$

where

$$e^\nu = |\alpha + e^{i\psi} \beta| = \sqrt{|\alpha|^2 + |\beta|^2 + 2\text{Re}(\beta\alpha^* e^{i\psi})}. \quad (15.55)$$

The number of particles grows for $\nu > 0$. This growth is exponential in conformal time,

$$|\beta_n|^2 = \text{const} \cdot e^{2(n+1)\nu} = \text{const} \cdot e^{4\nu\eta/T_\phi}.$$

Hence, we are dealing with the resonance situation. The resonance is not stochastic, the number of particles depends on time in a regular way.

In view of Eq. (15.54) and the relation $|\beta|^2 = |\alpha|^2 - 1$, the exponent ν in (15.55) is positive provided that

$$|\text{Re}(\alpha e^{i\theta})| > 1. \quad (15.56)$$

The maximum growth occurs for

$$\text{Arg}(\alpha e^{i\theta}) = 0 \pmod{\pi}. \quad (15.57)$$

In that case

$$e^\nu = |\alpha| + \sqrt{|\alpha|^2 - 1}.$$

Since both α and θ depend on conformal momentum, the relation (15.56) determines the resonance zones, i.e., the ranges of momenta where the number of χ -particles grow exponentially. The momenta of the fastest growth are determined by Eq. (15.57).

Let us describe the general properties of the solutions to Eq. (15.57). We are interested in momenta for which the coefficient β is not very small, i.e., (see (15.52))

$$k^2 \lesssim k_0^2 = \sqrt{\frac{\lambda}{2}} g \tilde{\phi}_0^2.$$

The coefficient α does not depend strongly on momentum in this range. On the other hand, the phase θ changes quite rapidly with k . Indeed, this phase is given by (15.53) with $\Omega_\chi^2 = k^2 + g^2 \tilde{\phi}^2(\eta)$. The dependence of the integral in (15.53) on k emerges due to the integration region where $|\tilde{\phi}(\eta)|$ is small, so we find to the leading logarithmic approximation

$$\theta(k) = \theta(k=0) + \frac{k^2}{k_0^2} \log\left(\frac{g \tilde{\phi}_0}{k}\right).$$

For momenta k not very different from k_0 we obtain finally

$$\theta(k) = \theta(k=0) + \frac{k^2}{4k_0^2} \log\left(\frac{g^2}{\lambda}\right), \quad (15.58)$$

where we omitted a numerical factor in the argument of logarithm, which cannot be trusted in the leading logarithmic approximation anyway. Because of the large logarithm in (15.58) the dependence of θ on k is indeed quite strong. We see that in the leading logarithmic approximation, Eq. (15.57) reduces to

$$\frac{k^2}{4k_0^2} \log\left(\frac{g^2}{\lambda}\right) = \text{const} + \pi n, \quad (15.59)$$

where the constant depends on the ratio g^2/λ and belongs to the interval $[0, \pi]$.

For $k^2 \ll k_0^2$ we have $|\alpha| = \sqrt{1 + |\beta|^2} = \sqrt{2}$. So, the largest exponent ν is equal to $\log(1 + \sqrt{2}) = 0.88$, and the fastest growth of particle number proceeds according to

$$n_k \propto (1 + \sqrt{2})^{2 \cdot \frac{2n}{T_\phi}} = e^{3.5 \frac{n}{T_\phi}}.$$

In general,⁴ the minimum momentum k that solves Eq. (15.59) is of order

$$k \sim \frac{k_0}{\sqrt{\log\left(\frac{g^2}{\lambda}\right)}}.$$

The width of the first resonance zone, where the condition (15.56) is satisfied, is of the same order. Hence the term “wide resonance”. It follows from (15.59) that there are numerous solutions to Eq. (15.57), i.e., there are many resonance zones. The widths of zones $k \sim k_0$ and the distances between these zones are small in the parameter $\log^{-1}\left(\frac{g^2}{\lambda}\right)$.

⁴The solutions to Eq. (15.57) start from $k = 0$ for special values of g^2/λ .

Despite these peculiarities, the qualitative picture of the decay of the inflaton oscillations is similar to the picture we have encountered in Section 15.2.1. Since the particle creation is exponential, the energy transfer from the inflaton to χ -particles takes a few oscillation periods only, and the system quickly enters the non-linear regime. The analysis of the latter regime is beyond the scope of this book.

Problem 15.13. Consider the model of this Section with $m_\chi \neq 0$ and $g^2 \gg \lambda$. Estimate the range of the amplitudes of the inflaton oscillations where particle creation occurs in the wide resonance regime studied in this Section and where, on the contrary, it proceeds in the stochastic resonance regime of Section 15.2.1. Assume that the cosmological expansion is dominated by the oscillating inflaton field; neglect the back reaction of χ -particles on the inflaton oscillations.

Problem 15.14. Consider the large field inflationary model with the inflaton potential

$$V(\phi) = \frac{m^2}{2}\phi^2 + \frac{\lambda}{4}\phi^4,$$

and assume that the mass m is so small that the quartic term dominates in the beginning of the inflaton oscillations. In the model of this Section with $m_\chi = 0$ and $g^2 \gg \lambda$, estimate the amplitude of the oscillations when the wide resonance regime is replaced by stochastic resonance. Make the same assumptions as in the previous problem.

To end this Section, we note that if the condition (15.51) is not satisfied, particle creation occurs in the regime somewhat similar to that of Section 15.2.2. Unlike in Section 15.2.2, however, the conformal frequency of the inflaton oscillations is time-independent, so for $m_\chi = 0$, χ -particles of one and the same conformal momentum are created all the time (until the back reaction of these particles on the inflaton becomes sizeable). Hence, approximate conformal invariance leads to the more efficient Bose-enhancement. This comment also applies to the creation of the inflaton quanta themselves.

15.4 Creation of Heavy Fermions

A mechanism of heavy particle creation somewhat similar to that of Section 15.2.1 exists for fermions as well [167]. To see this, let us study a model with the Yukawa coupling of fermions to the inflaton. The fermion part of the action is

$$S_F = \int d^4x \sqrt{-g} (\bar{\psi} i\gamma^\alpha \mathcal{D}_\alpha \psi - M\bar{\psi}\psi - g\phi\bar{\psi}\psi),$$

where the covariant derivative \mathcal{D}_α is defined in Section D.2. Let us consider for definiteness the quadratic inflaton potential (15.8). A particularly interesting case is when the fermion mass in the vacuum $\phi = 0$ is much greater than that of inflaton,

$$M \gg m. \quad (15.60)$$

At the same time, we assume that the amplitude of the inflaton oscillations is large, at least in the beginning,

$$g\phi_0 > M. \quad (15.61)$$

This is not a particularly strong requirement in the large field inflationary models. Then the effective fermion mass is

$$M_{\text{eff}}(t) = M + g\phi(t).$$

It crosses zero twice in one oscillation period, the adiabaticity conditions for the fermion evolution get violated at that time, and sporadic fermion creation takes place.

To describe this process, we write the Dirac equation for the fermion field $\chi = a^{3/2}\psi$ in conformal coordinates (η, \mathbf{x}) , see Section D.2,

$$i\gamma^\mu \partial_\mu \chi - aM_{\text{eff}} \chi = 0.$$

In the chiral representation of the Dirac matrices (see Section I.B.1) it splits into two equations for two-component spinors χ_L and χ_R ,

$$i\chi'_L + i\boldsymbol{\sigma} \boldsymbol{\partial} \chi_L - aM_{\text{eff}} \chi_R = 0 \quad (15.62)$$

$$i\chi'_R - i\boldsymbol{\sigma} \boldsymbol{\partial} \chi_R - aM_{\text{eff}} \chi_L = 0, \quad (15.63)$$

where prime still denotes the derivative with respect to conformal time η and $\boldsymbol{\sigma} = (\sigma^1, \sigma^2, \sigma^3)$ is the set of the Pauli matrices. We work in the momentum space, and choose χ_L and χ_R as the positive helicity spinors, $\boldsymbol{\sigma} \mathbf{k} \chi_{L,R} = k \chi_{L,R}$ (negative helicity is considered in the same way with the same results). Then Eqs. (15.62), (15.63) reduce to two equations for complex functions of time, $u = \chi_L + \chi_R$, $v = \chi_L - \chi_R$, namely

$$\begin{aligned} iu' - kv - aM_{\text{eff}} u &= 0, \\ iv' - ku + aM_{\text{eff}} v &= 0. \end{aligned}$$

We get rid of the variable u by using the second equation and substituting the result into the first one. In this way we obtain the second order equation

$$v'' + [a^2 M_{\text{eff}}^2 + k^2 - i(aM_{\text{eff}})']v = 0. \quad (15.64)$$

Like in Section 15.2.1, the evolution is adiabatic at all times except for short intervals around η_* when $M_{\text{eff}}(\eta_*) = 0$. In these intervals, Eq. (15.64) becomes

$$v'' + [a_*^2 g^2 (\phi'_*)^2 (\eta - \eta_*)^2 + k^2 - i(aM_{\text{eff}})_*']v = 0, \quad (15.65)$$

where asterisk denotes the quantities taken at time η_* . The form of Eq. (15.65) coincides with (15.16) where $\varepsilon^2 = k^2 - i(aM_{\text{eff}})_*'$. Hence, we immediately use the result (15.20) and obtain that after M_{eff} crosses zero for the first time, the number of created fermions is determined by

$$|\beta|^2 = \exp\left(-\frac{\pi k^2}{a_* g \dot{\phi}_*'}\right) = \exp\left(-\frac{\pi q^2}{g \dot{\phi}_*}\right),$$

where the last expression is written in terms of the physical momentum $q = k/a$ and derivative $\dot{\phi}_*$ with respect to the cosmic time. We note that as long as the oscillation amplitude ϕ_0 is greater than M/g , the effective mass is equal to zero at $\phi_* \ll \phi_0$, so that $\dot{\phi}_* \sim m\phi_0$.

Hence, despite their large mass M , fermions are efficiently created by the oscillating inflaton field, provided the oscillation amplitude is large, $g\phi_0 > M$. Conformal momenta of the created fermions obey

$$k^2 \lesssim a^2 g \dot{\phi}_* \sim a^2 g m \phi_0. \quad (15.66)$$

Generally speaking, the number of fermions depends on parameters and time in an irregular way, in analogy to Section 15.2.1. The difference with respect to the bosonic case is that the fermion occupation number cannot exceed 1, so we can make estimates by setting $|\beta|^2 \simeq 1$ for all momenta obeying (15.66). Since $\phi_0 \propto a^{-3/2}$, we see from (15.66) that the range of conformal momenta of created fermions extends in time, and the total number of fermions increases. Fermion creation terminates when the inflaton oscillation amplitude drops to M/g . At that time, all modes with physical momenta

$$q^2 \lesssim q_{\max}^2 \sim Mm \quad (15.67)$$

are occupied, and the number of created fermions is of order

$$n_f \sim \frac{4\pi}{3} g_f \frac{q_{\max}^3}{(2\pi)^3} \sim \frac{g_f}{6\pi^2} (Mm)^{3/2}, \quad \text{when } g\phi_0 = M, \quad (15.68)$$

where g_f is the number of spin states; see the formula (E.22) in regard to the factor $(2\pi)^{-3}$.

The estimate (15.68) has limited range of validity. First, it has been obtained in the quasi-linear regime, when the back reaction of fermions on the inflaton oscillations is neglected. The necessary condition for that is

$$\rho_f \ll \rho_\phi = \frac{m^2}{2} \phi_0^2. \quad (15.69)$$

We see from (15.60) and (15.67) that the fermion momenta are small compared to the mass M , so the fermion energy density is given by

$$\rho_f = M n_f \sim \frac{g_f}{6\pi^2} M^{5/2} m^{3/2}.$$

Hence, one applicability condition for the result (15.68) is

$$g^4 M \ll \left(\frac{3\pi^2}{g_f} \right)^2 m, \quad (15.70)$$

where we recalled that $g\phi_0 \simeq M$ at the relevant time. Second, inflaton can decay into other particles, so its oscillations may get damped relatively fast. In that case, a conservative estimate for the number of created fermions is obtained by counting only those particles which are created when the effective mass $M_{\text{eff}} = M + g\phi$ crosses zero for the first time. Then the number of created fermions is given by the same formula (15.23) that works for bosons, and we obtain

$$n_f \gtrsim \frac{g_f}{8\pi^3} (g\phi_e m)^{3/2}, \quad (15.71)$$

where we recalled that the first inflaton oscillation has the amplitude $\phi_0 \sim \phi_e$. We again have to account for the bound (15.69), which now gives

$$g^3 M^2 \ll \left(\frac{4\pi^3}{g_f} \right)^2 m\phi_e. \quad (15.72)$$

If the condition (15.69) is not satisfied, the quasi-linear regime ends at a certain amplitude ϕ_0 . At that time $\rho_f \sim \rho_\phi$ and

$$n_f \sim \frac{m^2 \phi_0^2}{M}. \quad (15.73)$$

We will see in Section 15.5.1 that any of these estimates means that the number density of created fermions is large enough for physics applications.

15.5 Physics Applications

15.5.1 Generation of baryon asymmetry at reheating

We have seen in Sections 15.2.1 and 15.4 that the oscillating inflaton field can create very heavy particles. This suggests new mechanisms of the baryon asymmetry generation [168–171]. In particular, baryon number can be violated due to the Grand Unified Theory (GUT) interactions (see Section I.11.2.2) operating at the reheating epoch [168]. The relevant processes are decays of particles whose mass is comparable to the GUT scale $M_{\text{GUT}} \sim 10^{15} - 10^{16}$ GeV, see Section I.11.3. It follows from (15.22) and (15.61) that if these particles are sufficiently strongly coupled to the inflaton, they can be created at the preheating epoch in the large field inflationary models, where the initial amplitude of the inflaton oscillations is of order M_{Pl} . Indeed, the maximum mass of bosons that can be produced in the model of Section 15.2.1 is $m_\chi \sim (m M_{Pl})^{1/2} \sim 10^{16}$ GeV for $g \sim 1$, while the maximum mass of created fermions in the model of Section 15.4 is only slightly smaller than M_{Pl} . Since cosmic medium at preheating is far from thermal equilibrium, the third Sakharov condition (see Section I.11.1) can be easily satisfied. In particular, the inverse decays can be absent, so there is no need to impose the requirement of small width of the decaying particles, $\Gamma \lesssim H$, which is characteristic of thermal baryogenesis.

Let us discuss this possibility in the context of the models of Sections 15.2.1 and 15.4. In view of the fact that we have no good handle on particle physics at the GUT energy scale, especially on the interactions of the inflaton with other particles, our discussion is going to be rather vague. Our purpose is to show that the observed baryon asymmetry can indeed be generated at the reheating epoch without fine tuning of parameters.

We begin with the model of Section 15.2.1. Let the baryon asymmetry generated in the decay of one χ -particle be

$$\delta_\chi = \frac{\sum_f \Gamma(\chi \rightarrow f) B_f}{\Gamma_{\chi,tot}},$$

where the sum runs over all decay channels f , and B_f is the baryon number of the final state.⁵ Let us assume further that the decay into χ -particles is one of the dominant inflaton decay channels. According to Section 15.2.1, the quasi-linear regime of the decay of the inflaton oscillations lasts several oscillation periods only. During that time, the cosmological expansion does not considerably reduce the amplitude of oscillations, so this amplitude is of the order of the inflaton field at the end of inflation,

$$\phi_0 \sim \phi_e \sim \frac{M_{Pl}}{\sqrt{4\pi}}. \quad (15.74)$$

The number density of the created χ -particles is given by (15.34). Let us see that it is sufficient for generating the observed baryon asymmetry. Of course, χ -particles are created at the later stage of the non-linear evolution as well, so our estimates are in fact lower bounds.

The decays of χ -particles generate the following baryon number density,

$$n_B = \delta_\chi n_\chi.$$

The inverse decays do not occur in the situation we consider, since the system is very far from thermal equilibrium. At the preheating stage, the cosmological expansion is dominated by the energy density of the inflaton oscillations that decreases as $\rho \propto a^{-3}$, and n_B/ρ stays constant. After the transition to radiation domination, the time-independent parameter is $n_B/\rho^{3/4}$, and the same parameter estimates the baryon asymmetry n_B/s after thermal equilibrium is established (modulo a factor

⁵In fact, the appropriate quantum number is $(B - L)$ rather than baryon number B itself: electroweak processes do not conserve baryon and lepton numbers separately at fairly low temperatures, but conserve $(B - L)$, see Section I.11.2.1. Hence, the resulting baryon asymmetry is proportional to $(B - L)$ -asymmetry generated at the epoch we consider. Another qualification is that if χ -particles do not coincide with their antiparticles (which is often the case), and the inflaton decays into $\chi\bar{\chi}$ -pair, the asymmetry δ_χ is a sum of the contributions of a particle and antiparticle, see Section I.11.3.

of $g_*^{1/4}$ and numerical factor of order 1). Hence, the resulting baryon asymmetry is estimated as

$$\Delta_B \equiv \frac{n_B}{s} \sim \delta_\chi \rho^{1/4}(t_{reh}) \frac{n_\chi}{\rho},$$

where $\rho(t_{reh})$ is the energy density at the transition to radiation domination, n_χ/ρ is the ratio of number density of χ -particles to the energy density at the time of χ -particle creation. We use (15.34) and $\rho \sim m^2 \phi_0^2$ to obtain⁶ $n_\chi/\rho \sim (g\phi_0)^{-1}$. The lower bound on $\rho(t_{reh})$ is obtained by noticing that radiation domination begins not later than the time at which all non-relativistic particles, including inflaton and χ -particles, decay, i.e., not later than the time at which $H(t) \sim \Gamma$, where Γ is the relevant heavy particle width. Omitting a factor of order 1, we write

$$\rho^{1/4}(t_{reh}) \gtrsim (\Gamma M_{Pl})^{1/2}.$$

The heavy particle width is estimated as $\Gamma \sim h^2 M / (8\pi)$, where h is the relevant coupling and M is the mass scale of the theory. Hence,

$$\Delta_B \gtrsim \delta_\chi \frac{h}{g} \left(\frac{M M_{Pl}}{8\pi \phi_0^2} \right)^{1/2} \sim \delta_\chi \frac{h}{g} \left(\frac{M}{M_{Pl}} \right)^{1/2}, \quad (15.75)$$

where we recalled (15.74). The value of δ_χ can be of order $10^{-2} - 10^{-3}$, see Section I.11.3, while the second factor in the right hand side of (15.75) can be of order 1. If the relevant mass scale is of the order of the inflaton mass, $M \simeq 10^{-6} M_{Pl}$, then the third factor is roughly 10^{-3} . We see that despite all uncertainties, the mechanism we discuss is capable of producing the observed baryon asymmetry $\Delta_B \simeq 10^{-10}$ without fine tuning of parameters.

Let us turn to the baryon asymmetry generation in decays of fermions created as in Section 15.4. Let us make use of the estimate (15.68) and recall that fermion creation terminates when $\rho_\phi = m^2 \phi_0^2 / 2 = m^2 M^2 / (2g^2)$. Then we have

$$\frac{n_f}{\rho} \sim \frac{g_f g^2}{3\pi^2} \frac{1}{\sqrt{Mm}}.$$

Repeating the above analysis, we obtain the estimate for the resulting baryon asymmetry,

$$\Delta_B \sim \delta_f \frac{g_f g^2}{3\pi^2} \left(\frac{h^2 N_d M_{Pl}}{8\pi M} \right)^{1/2}, \quad (15.76)$$

where δ_f is the asymmetry in decays of a fermion and antifermion. We again see that the baryon asymmetry generation is efficient in a wide parameter range.

⁶This estimate is consistent with the fact that the energy density of χ -particles at the time of their creation, $\rho_\chi \sim \varepsilon n_\chi \sim g\phi_0 n_\chi$, is of the order of the total energy density, see the end of Section 15.2.1.

The result (15.76) is valid when the estimate (15.68) is correct. If we use the result (15.72) instead, then the fermion number density is conservatively estimated by (15.71). In that case, the ratio n_f/ρ before the transition to radiation domination is of order

$$\frac{n_f}{\rho} \gtrsim \frac{g_f g^{3/2}}{4\pi^3} \frac{1}{\sqrt{m\phi_e}}.$$

This gives for the resulting asymmetry

$$\Delta_B \gtrsim \delta_f \frac{g_f g^{3/2} h}{4\sqrt{2}\pi^3} N_d^{1/2}.$$

Similar estimate is valid in the bosonic case, if the inflaton oscillations last for only a few periods (say, because of strong inflaton interactions with other fields). The numerical value of the asymmetry is again comfortably large, unless the couplings are very small.

Finally, if the fermion creation itself leads to the termination of the quasi-linear stage of the inflaton oscillations, then at that time the fermion number density is estimated by (15.73). Hence, $n_f/\rho \gtrsim 1/(2M)$ and

$$\Delta_B \gtrsim \delta_f \left(\frac{\Gamma M_{Pl}}{4M^2} \right)^{1/2} \sim \delta_f \left(\frac{h^2 N_d m M_{Pl}}{32\pi M^2} \right)^{1/2}.$$

This value is again large enough without fine tuning of parameters.

It is worth noting that similar mechanism may lead to the dark matter generation by the inflaton oscillations. Dark matter can then consist of either the created particles themselves or their decay products. According to Sections 15.2, 15.4, the mass of the dark matter particles can be extremely high; it can exceed considerably the reheating temperature and even be of the order of the GUT scale 10^{16} GeV. Very heavy dark matter particle is sometimes called wimpzilla.

Problem 15.15. *Show that the latter mechanism of dark matter generation is capable of producing the right amount of dark matter, $\Omega_{CDM} \simeq 0.2$, in a wide range of parameters.*

15.5.2 Non-thermal phase transitions

As we have seen in Sections 15.2.1, 15.3, the decay of inflaton oscillations can produce large fluctuations of bosonic fields. These fluctuations can affect the symmetry properties of the cosmic medium. As a result, the preheating epoch can exhibit specific, non-thermal phase transitions [172–174]. This possibility is of interest, e.g., from the viewpoint of the production of stable or metastable topological defects via the Kibble mechanism, see Chapter I.12.

A simple model that illustrates the non-thermal phase transitions is obtained by adding to the model with the inflaton and the field χ yet another scalar field σ with the potential

$$V = \frac{\lambda_\sigma}{4} (\sigma^2 - v^2)^2 + \lambda_{\sigma\chi} \chi^2 \sigma^2.$$

The symmetry of this model is $\sigma \rightarrow -\sigma$. This symmetry is spontaneously broken in vacuo, where $\chi = 0$, $\sigma = \pm v$. If there are large fluctuations of the field χ , the expectation value of σ is at the minimum of the effective potential (we neglect the fluctuations of the field σ itself for the sake of argument)

$$V_{\text{eff}}(\sigma) = \frac{\lambda_\sigma}{4} (\sigma^2 - v^2)^2 + \lambda_{\sigma\chi} \langle \chi^2 \rangle \sigma^2.$$

The symmetry is restored, provided that

$$\lambda_{\sigma\chi} \langle \chi^2 \rangle > \frac{\lambda_\sigma}{2} v^2. \quad (15.77)$$

At finite temperature, $\langle \chi^2 \rangle$ is non-zero (although $\langle \chi \rangle = 0$) due to thermal fluctuations, and

$$\langle \chi^2 \rangle_T \sim T^2.$$

We note in passing that this observation can be used as the basis for an alternative derivation of the phase transition temperature as compared to Section I.10.2. If the maximum temperature in the Universe is smaller than $\sqrt{\lambda_\sigma/\lambda_{\sigma\chi}}v$, the symmetry is unbroken at the entire hot stage, $\langle \sigma \rangle \neq 0$, and the thermal phase transition never occurs. On the other hand, in the model of Section 15.2.1, the fluctuations of the field χ at the quasi-linear preheating epoch are estimated as in (15.33), and they are even stronger at the non-linear stage. Hence, the inequality (15.77) can be valid at preheating, the system is then in the unbroken phase, and the phase transition occurs towards the end of the reheating epoch. The study of the range of parameters in which the non-thermal phase transition occurs requires the analysis of the system in the non-linear regime. Such an analysis is performed within classical field theory, see the end of Section 15.2.1. It shows [174], that the non-thermal phase transitions indeed occur, and their energy scale can be comparable to the GUT scale, $v \sim M_{\text{GUT}}$. Hence, there is a possibility of the production of topological defects, including magnetic monopoles and cosmic strings, at preheating. The monopole (and domain wall) production is undesirable, see Sections I.12.2.3, I.12.4, while the production of cosmic strings is of considerable interest. Detailed discussion of these issues is beyond the scope of this book.

15.5.3 Small wavelength gravity waves

Yet another consequence of strong fluctuations of fields at preheating is the possibility of the generation of gravity waves. We see, e.g., from (15.21) that the fluctuations can have fairly large wavelengths; the only model-independent bound is that the wavelength must be smaller than the current horizon size. This is in the rough agreement with numerical simulations [175]. Hence, the minimum physical

momentum of gravity waves at their generation is of order H_{gen} ; hereafter the subscript “gen” refers to the epoch of the gravity wave generation. Let us estimate the present value of this momentum. We write

$$q_0 = q_{gen} \frac{a_{gen}}{a_0} \simeq H_{gen} \frac{a_{gen}}{a_0}.$$

Repeating the argument given at the end of Section 13.1.3, we find

$$\frac{a_0}{a_{gen}} = \left(\frac{90}{8\pi^3 g_{*,0}} \right)^{1/3} \frac{M_{Pl}}{T_0} \left(\frac{H_{gen}^2}{M_{Pl} T_{reh}} \right)^{1/3} = 2 \cdot 10^{31} \left(\frac{H_{gen}^2}{M_{Pl} T_{reh}} \right)^{1/3}, \quad (15.78)$$

where we assumed that the equation of state at the reheating epoch is $p = 0$ (this corresponds to $\beta = 3$ in the formulas of Section 13.1.3), and inserted numerical factors. Hence, the minimum present momentum is

$$q_0 \simeq \left(\frac{8\pi^3 g_{*,0}}{90} \right)^{1/3} T_0 \left(\frac{H_{gen} T_{reh}}{M_{Pl}^2} \right)^{1/3} \simeq 30 \left(\frac{H_{gen} T_{reh}}{M_{Pl}^2} \right)^{1/3} \text{ cm}^{-1}. \quad (15.79)$$

To make a numerical estimate, let us consider large field inflation with the quadratic inflaton potential. The lower bound is obtained by assuming that the inflaton couplings to other fields are small, and its oscillations decay via the mechanism discussed in Section 15.1. The minimum value of H_{gen} corresponds to the gravity wave generation just before the beginning of the hot stage, $H_{gen} = \sqrt{4g_*\pi^3/45} T_{reh}^2/M_{Pl}$. We now estimate T_{reh} according to (15.7) and obtain $q_0 \sim 7 \cdot 10^{-10} \text{ cm}^{-1}$. The minimum frequency and maximum wavelength are

$$\omega_0 \sim 20 \text{ Hz}, \quad \lambda_0 = \frac{2\pi}{q_0} \sim 10^5 \text{ km}. \quad (15.80)$$

We see that the gravity waves generated at preheating are of fairly high frequency. Gravity waves generated in models with faster inflaton decay have even higher frequencies.

Gravity waves in this frequency ballpark are searched for by terrestrial laser interferometers. In particular, detectors LIGO and Virgo are sensitive to gravity waves with $\omega_0 \approx 10^2 \text{ Hz}$, see Fig. 13.3. We see that the frequency range of the gravity waves produced at reheating can indeed be probed.

To estimate the present gravity wave amplitude h_0 , we recall that it decreases after the generation epoch as a^{-1} , see Section 3.2.2, so that $h_0 = (a_{gen}/a_0)h_{gen}$. Under the above assumptions we find from (15.78) that $h_0 \equiv \Delta_{T,0} \simeq 0.6 \cdot 10^{-21} h_{gen}$. The fractional contribution of a decimal momentum interval of gravity waves to the present energy density, $\Omega_{GW,log}$, is obtained from (13.71). We find

$$\Omega_{GW,log} = \frac{h_0^2}{12} \frac{q_0^2}{H_0^2} \simeq 2.5 \cdot 10^{-6} h_{gen}^2.$$

The most optimistic estimate is obtained by assuming that the perturbations in the energy density are large, $\delta\rho/\rho \sim 1$, at all length scales up to the current horizon, so

that $h_{gen} \sim 1$. If so, the gravity wave amplitudes are large enough to be detectable at the existing and planned interferometers, see Fig. 13.3. We note, however, that we have overestimated the amplitude; the actual gravity wave amplitudes [175,176] are much smaller in realistic models, and, furthermore, their frequencies are much higher than our estimate (15.80).

This page is intentionally left blank

Chapter 16

*Bouncing Universe

We have seen in previous Chapters that the inflationary theory successfully solves all problems related to the initial conditions for the hot Big Bang epoch. Still, it is legitimate to ask whether the Hot Big Bang epoch was preceded by some other, non-inflationary stage of the cosmological evolution. A logical possibility here is that the Universe underwent collapse before expansion, and the transition from one regime to the other, “bounce”, occurred at high enough energy density. Let us briefly discuss this alternative to inflation. Our purpose is to understand what properties of the theory are required in this scenario, and whether and how the problems of the Hot Big Bang theory are solved.

In the first place, the transition from collapse to expansion in the classical General Relativity requires very peculiar matter. To see this, we combine the evolution equations of the homogeneous and isotropic Universe, Eqs. (I.3.9) and (I.3.16), and obtain for the spatially flat Universe

$$\dot{H} \equiv \frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} = -4\pi G(\rho + p).$$

The Hubble parameter *grows* from negative to positive values at the bounce epoch, $\dot{H} > 0$. This is possible only if

$$p < -\rho, \tag{16.1}$$

and $\rho \geq 0$ in view of the Friedmann equation. Known forms of matter, including classical scalar and vector fields with standard Lagrangians, do not obey (16.1). Hypothetical matter with the property (16.1) is called phantom matter (another term is *phantom energy*). Constructing self-consistent field theory models with phantom properties is a non-trivial problem.

Another possibility is that the bounce is due to quantum phenomena that cannot be described within the classical General Relativity. Such a possibility is discussed, e.g., in the *pre-Big Bang scenario* [177, 178] and early version of the *ekpyrotic scenario* [179]. The problem here is that quantum gravity or theory that replaces it (e.g., superstring theory) is not developed to the extent that questions of whether

the bounce is possible at all in the quantum regime and what is the state of the Universe right after the bounce, cannot be unequivocally answered.

A simple possibility to satisfy the condition (16.1) is to introduce a scalar field ϕ with negative kinetic term in the Lagrangian,

$$\mathcal{L} = -\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi).$$

The energy density and pressure of the homogeneous field in this model are (cf. (12.6), (12.7))

$$\rho = -\frac{1}{2}\dot{\phi}^2 + V(\phi), \quad p = -\frac{1}{2}\dot{\phi}^2 - V(\phi).$$

Hence, the inequality (16.1) indeed holds. This model is unacceptable, however, since the quanta of the field ϕ have *negative energy*. Fields of this sort are called ghosts. Furthermore, the energy is unbounded from below: the energy of a ϕ -particle is equal to $E = -|\mathbf{q}|$ at high spatial momentum \mathbf{q} . This leads to the instability of vacuum against the spontaneous creation of the ϕ -particles together with conventional particles via, e.g., the graviton exchange, as shown in Fig. 16.1. Note that these processes are forbidden by energy conservation in theories with positive energies of all particles, but this argument does not work in the model at hand. This instability is catastrophic in a Lorentz-invariant theory, since all Lorentz-boosted processes are allowed, and the phase space is infinite. The only known possibility to get around the catastrophic instability in phantom theories is to introduce Lorentz-violation, see, e.g., Refs. [180, 181]; the cosmological models with bounce can indeed be constructed in this way [182–184].

The horizon and homogeneity problems discussed in Section 11.1.1 are two different problems in the bouncing Universe scenario. If the collapse took a long, let alone infinite, period of time, the observable Universe may well be causally connected; the horizon problem is then solved. However, if the Universe were inhomogeneous before the bounce, it would remain inhomogeneous, generally speaking, after that. Hence, the homogeneity problem is not solved automatically, it is merely moved to the earlier epoch of the collapse.

Problem 16.1. Let the Universe collapse according to $a(t) \propto |t - t_*|^\alpha$, $t_* > 0$, $\alpha > 0$ at $t < 0$, and start expansion at $t = 0$ in the same way as in the Hot Big Bang theory with $H(t = 0) \sim t_*^{-1}$. Consider a light cone emanating at time t . Find the

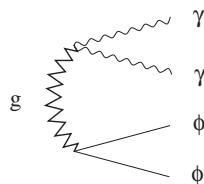


Fig. 16.1 The process “vacuum $\rightarrow \phi\phi\gamma\gamma$ ” in the model with ghost ϕ .

value of t for which this light cone presently covers the visible Universe (understood as the present interior of the photon last scattering sphere).

Although the flatness problem does not find its complete solution in the bouncing Universe scenario, it is alleviated substantially. The relative contribution of the spatial curvature to the Friedmann equation *decreases* as the Universe contracts, so the inequality of the type (11.10) is obeyed automatically, if the curvature was substantial, if at all, only very long before the bounce. Hence, the “geometric” problems of Sections 11.1.1 and 11.1.2 reduce, basically, to the assumption that the collapsing Universe passes through the state similar to the present one.

The entropy problem is solved if the phantom matter that dominates at the bounce epoch decays into conventional particles. Phantom matter most probably has very low or zero entropy, and its decay results in the appearance of the hot gas of particles. This solution of the entropy problem is similar to the inflationary one, at least in principle.

Even more attractive scenario is the pulsating, *cyclic* Universe. It assumes that the Universe undergoes many periods of expansion, halt, then contraction, bounce and expansion again. We are in one of these cycles. The homogeneity and spatial flatness of the visible Universe may be due to dark energy domination at the late stage of each cycle.

The generation of cosmological perturbations in bounce and cycling models is a difficult problem: unlike in inflationary theories, flat or nearly flat spectrum of scalar perturbations is not at all automatic there, see problem 16.7. Some progress in this direction has been achieved only recently, see discussion in Refs. [141, 184–186, 188–190].

There is a new problem specific to models with collapse and bounce. It has to do with the *anisotropy* in the collapsing Universe. It is known that if matter equation of state is $p < \rho$, then the collapse occurs in a locally anisotropic and chaotic way. This is the result of the Belinsky–Lifshits–Khalatnikov theory [191], see the details in the book [192]. Let us summarize its results. During sufficiently long time intervals, the evolution of a *local region* in the collapsing Universe is adequately described by the homogeneous *anisotropic* metric

$$ds^2 = dt^2 - a^2(t) \cdot \sum_{a=1}^3 e^{2\beta_a(t)} e_i^{(a)} e_j^{(a)} dx^i dx^j, \quad (16.2)$$

where $e_i^{(a)}$ are three linear independent vectors which are constant in time at each of the time intervals. These vectors, generally speaking, are not orthogonal to each other. The function $a(t)$ is chosen in such a way that

$$\sum_a \beta_a = 0; \quad (16.3)$$

in other words, $\det g_{ij} = a^6$. The fact that $e_i^{(a)}$ are not orthogonal to each other is irrelevant at each of the time intervals, and the Einstein equations give

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{1}{6} \sum_a \dot{\beta}_a^2 + \frac{8\pi}{3} G\rho, \quad (16.4)$$

$$\ddot{\beta}_a + 3\frac{\dot{a}}{a}\dot{\beta}_a = 0. \quad (16.5)$$

Problem 16.2. Show that the Einstein equations for the metric (16.2) with the orthonormalized vectors, $\sum_a e_i^{(a)} e_j^{(a)} = \delta_{ij}$, $\sum_i e_i^{(a)} e_i^{(b)} = \delta^{ab}$, are given by (16.4), (16.5).

Equation (16.5) gives

$$\dot{\beta}_a = \frac{d_a}{a^3}, \quad (16.6)$$

and in view of (16.3), the constants d_a obey $\sum_a d_a = 0$. Then Eq. (16.4) becomes

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{1}{6a^6} \sum_a d_a^2 + \frac{8\pi}{3} G\rho. \quad (16.7)$$

This equation shows that the overall contraction rate (the rate at which $\det g_{ij}$ decreases) is determined at small a by the local anisotropy rather than matter, provided that $p < \rho$. Therefore, one can set $\rho = 0$ late at the collapsing stage, and the system of equations (16.6), (16.7) has the Kasner solution,

$$\begin{aligned} a(t) &= |t|^{1/3}, \quad \beta_a = d_a \log |t| \\ \sum_a d_a &= 0, \quad \sum_a d_a^2 = \frac{2}{3}. \end{aligned} \quad (16.8)$$

In the general case when the vectors $e_i^{(a)}$ are not orthogonal to each other, this regime continues for finite time, and then the values of the parameters d_a change in a rather abrupt manner. The vectors $e_i^{(a)}$ change too. This change occurs infinitely many times in the limit $t \rightarrow 0$. This corresponds to the chaotic anisotropic collapse.

These results show that matter is irrelevant from the viewpoint of the cosmological evolution at the late stages of collapse. Hence, the transition from contraction to expansion is problematic. Even if it can be achieved, the Universe is very inhomogeneous before the bounce. The processes we described occur independently in Hubble-size regions and are very different in each of them because of their chaotic properties. This picture remains valid after the bounce, at least in the framework of the classical theory. Strong inhomogeneity of the Universe after the bounce is inconsistent with the smallness of the primordial cosmological perturbations, so the entire bounce scenario is up in the air.

To solve this problem, one invokes matter with super-stiff equation of state $p = w\rho$, $w > 1$. Its energy density behaves as $\rho \propto a^{-3(1+w)}$ (see (I.3.41)), so it

increases faster than a^{-6} for $w > 1$. The second term in the right hand side of Eq. (16.7) dominates, and one finds the following behavior of the scale factor,

$$a(t) \propto |t|^\alpha, \quad t < 0, \quad (16.9)$$

where

$$\alpha = \frac{2}{3(1+w)} < \frac{1}{3}. \quad (16.10)$$

It then follows from Eq. (16.6) that the parameters β_a tend to constants as $t \rightarrow 0$. If the Universe is nearly homogeneous at the early stages of collapse, and anisotropy is not strong, then the Universe becomes more and more homogeneous in the process of contraction, see details in Ref. [193].

Metric perturbations do not grow at the contraction epoch in the model with super-stiff matter. Hence, one can study the generation of cosmological perturbations in the collapsing Universe, and use the linearized theory for that purpose. Let us perform the analysis for tensor perturbations; see problem 16.7 in this Chapter regarding scalar perturbations. For the background given by (16.9), the equation for superhorizon tensor perturbations is (cf. Section 3.2)

$$\ddot{h} + \frac{3\alpha}{t}\dot{h} = 0,$$

where we omitted the polarization index. One of the solutions is constant in time, whereas the other behaves as $|t|^{1-3\alpha}$. If matter is not super-stiff, one has $p < \rho$ and $\alpha > 1/3$, so the second solution grows as the Universe contracts (this is the decreasing mode in the time reversed, expanding Universe). This is precisely the linearized theory counterpart of the growing anisotropy we discussed above. If matter is super-stiff, then $\alpha < 1/3$ and the second solution *decreases* in time, so the constant mode dominates in the superhorizon regime. This is a general property of the Universe filled with super-stiff matter, $p > \rho$; the assumption that the equation of state parameter w is independent of time is unnecessary.

Problem 16.3. *Prove the last statement above.*

The overall behavior of tensor perturbations in the model of collapse with super-stiff matter is in a sense similar to their behavior at inflation. A given mode is first subhorizon, then it becomes superhorizon and freezes in. A mode of momentum k exits the horizon at time t_k when

$$k = a(t_k)H(t_k). \quad (16.11)$$

Assuming that gravity waves are absent early at the collapsing stage, we repeat the analysis of Section 13.3 and find the tensor power spectrum after the horizon exit,

$$\mathcal{P}_T(k) = \frac{16c}{\pi} \frac{H^2(t_k)}{M_{Pl}^2}, \quad (16.12)$$

where the numerical constant c is of order 1 (it depends on the behavior of $a(t)$ at $t \approx t_k$). Short modes exit the horizon earlier, the Hubble parameter is larger at that time and the perturbation amplitude is larger. Hence, the tensor tilt is positive. For power-law contraction (16.9) we have $|H| \propto |t|^{-1}$, $a \propto |H|^{-\alpha}$, and the formula (16.11) gives $|H(t_k)| \propto k^{1/(1-\alpha)}$. This means that $\mathcal{P}_T \propto k^{2/(1-\alpha)}$, i.e.,

$$n_T = \frac{2}{1-\alpha} > 0.$$

Hence, the prediction of the class of models we consider is the blue tensor spectrum, in strong contrast to inflation. We note that the gravity wave amplitudes at frequencies suitable for their detection are typically so small that their experimental discovery is hopeless.

Problem 16.4. Estimate the maximum value of the Hubble parameter at the end of the contracting stage by requiring that the energy density of gravity waves at the Big Bang Nucleosynthesis epoch does not exceed the energy density of one neutrino species. Assume that the cosmological contraction proceeds according to the power law (16.9), the bounce and the transition to the hot stage occur instantaneously, and the Hubble parameters before and after the bounce are of the same order.

Problem 16.5. Using the result of the previous problem, estimate the maximum amplitudes of relic gravity waves of wavelengths 1 m , 10^3 km (accessible for terrestrial detectors, sensitivity to amplitudes are at the level 10^{-20} and 10^{-22} , respectively) and 10^8 km (projects in outer space, sensitivity at the level 10^{-23}). Compare with Fig. 13.3.

An example of super-stiff matter is a scalar field with the negative exponential potential,

$$V(\phi) = -V_0 e^{\phi/M}, \quad (16.13)$$

where V_0 and M are positive parameters. The equation for the homogeneous field $\phi(t)$ and the Friedmann equation have the following solution,

$$a(t) = |t|^\alpha, \quad \phi(t) = \text{const} - 2M \log |t|, \quad V[\phi(t)] = -\frac{2M^2(1-3\alpha)}{t^2}, \quad (16.14)$$

where

$$\alpha = 16\pi \frac{M^2}{M_{Pl}^2}. \quad (16.15)$$

This is the attractor solution in the case of collapse. According to (16.10) and (16.15), the effective equation of state is indeed super-stiff, $w \gg 1$, for $M \ll M_{Pl}$. Note that the energy density is positive and increases as the Universe collapses,

$$\rho = \frac{1}{2}\dot{\phi}^2 + V(\phi) = \frac{6M^2\alpha}{t^2}.$$

This leaves open the possibility that the field moves out of the negative potential well at the end of the contraction epoch, the equation of state changes and obeys (16.1), and the Universe undergoes the bounce. Ekpyrotic models are assumed to have precisely these properties [182–184].

Problem 16.6. *Show that (16.14) is indeed the attractor solution to the scalar field equation and the Friedmann equation for the collapsing Universe.*

Problem 16.7. *Assuming that the scalar field fluctuations are in the vacuum state early at the contraction epoch, find the power spectrum of the superhorizon scalar perturbations in the model (16.13). Show, in particular, that the scalar spectral index is*

$$n_s - 1 = \frac{2}{1 - \alpha},$$

i.e., the scalar and tensor tilts coincide. This result illustrates the difficulty of obtaining nearly flat scalar spectrum in bouncing models. Hint: Make use of Eq. (13.53).

This page is intentionally left blank

Chapter 17

Color Pages

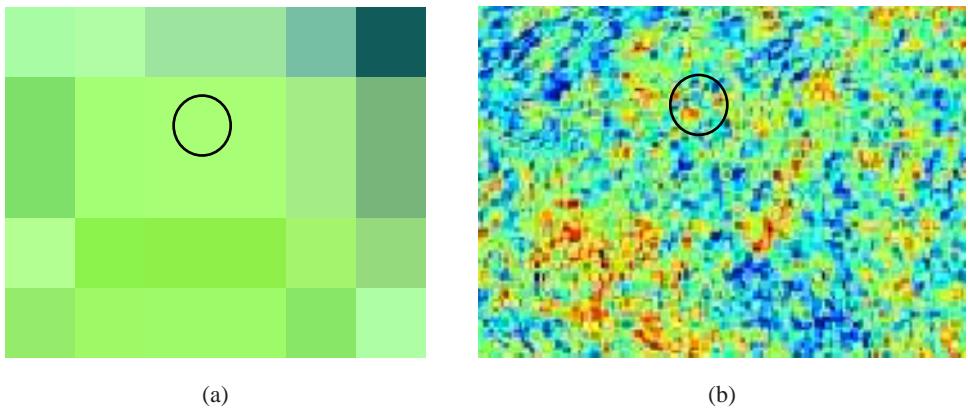


Fig. 17.1 Relationship between typical spatial scales of cosmological perturbations and the horizon size. Color encodes the value of a quantity characterizing the perturbation (e.g., $\delta\rho(\mathbf{x})$), the horizon is shown by circle. a) Superhorizon modes: perturbations are spatially homogeneous at the horizon scale; b) subhorizon modes: spatial inhomogeneities are sizeable inside the horizon.

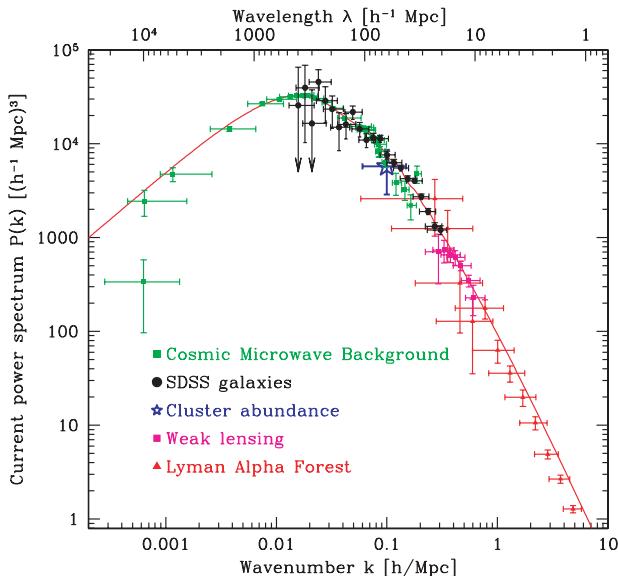


Fig. 17.2 Observational data on the present linear power spectrum of matter perturbations [15].

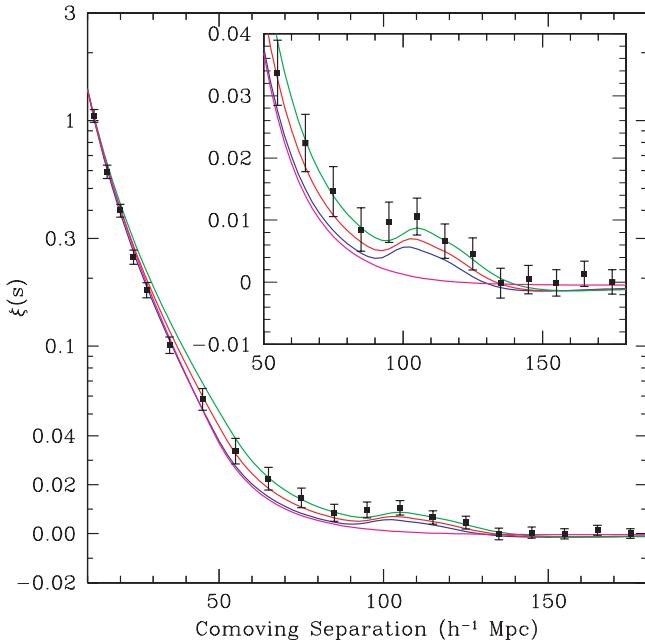


Fig. 17.3 The correlation function $\xi(s)$ determined by the analysis of the SDSS data on the distribution of distant galaxies [21]. Solid lines show the predictions of various cosmological models. Green, red and blue lines correspond to $\Omega_M h^2 = 0.12, 0.13, 0.14$, respectively, with $\Omega_B h^2 = 0.024$, $n_s = 0.98$ in all cases. Magenta line corresponds to unrealistic Universe without baryons.

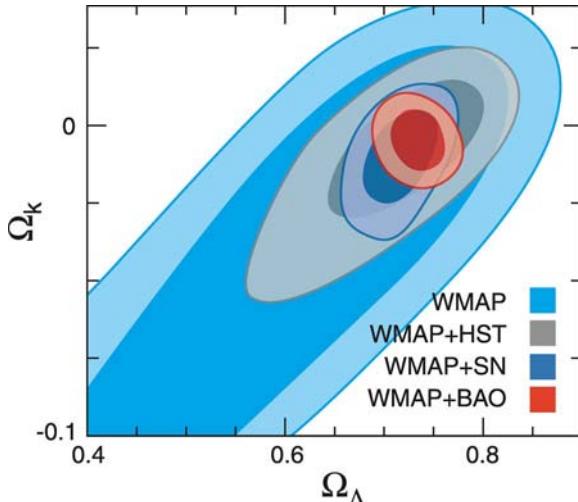


Fig. 17.4 The degeneracy in parameters (spatial curvature and dark energy density in this case) is removed by combining the data from different observations. The central (dark) and light colored regions are allowed at 68% and 95% confidence level, respectively.

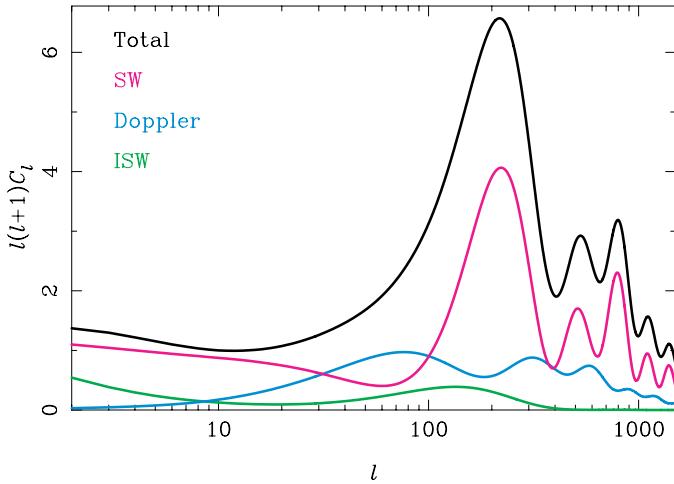


Fig. 17.5 The three contributions to CMB temperature anisotropy and the total angular spectrum [77] calculated for adiabatic perturbations in the spatially flat Λ CDM model with $\Omega_\Lambda = 0.75$, $\Omega_B h^2 = 0.023$, $\Omega_{CDM} h^2 = 0.111$, $h = 0.73$; primordial power spectrum is flat, $n_s = 1$.

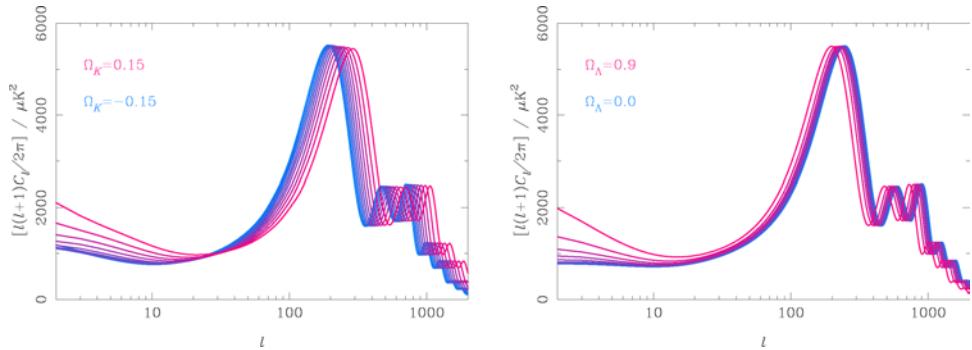


Fig. 17.6 The effect of the spatial curvature (left) and the cosmological constant (right) on the CMB temperature angular spectrum [77].

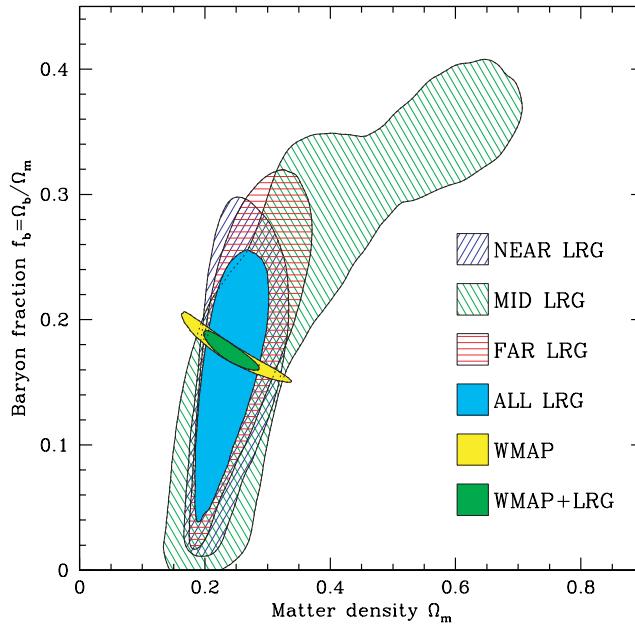


Fig. 17.7 Regions in the space of parameters $(\Omega_M, \Omega_B/\Omega_M)$, allowed at 95% C.L. by data on large scale structure (LRG) and on CMB temperature anisotropy (WMAP), Ref. [53].

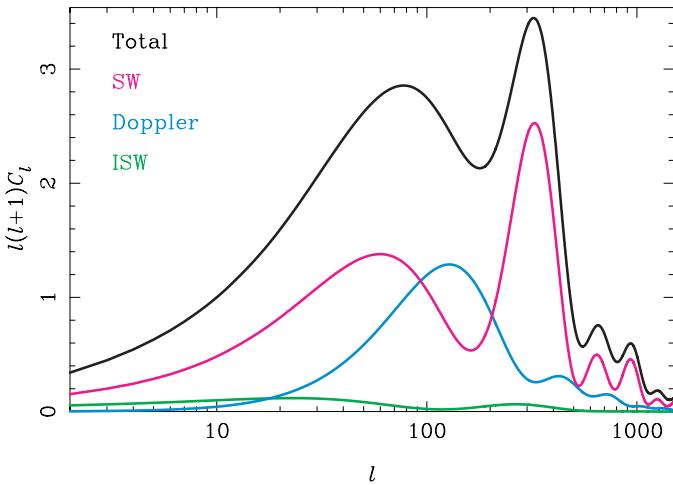


Fig. 17.8 CMB temperature anisotropy spectrum generated by isocurvature mode [77]. The primordial power spectrum is flat.

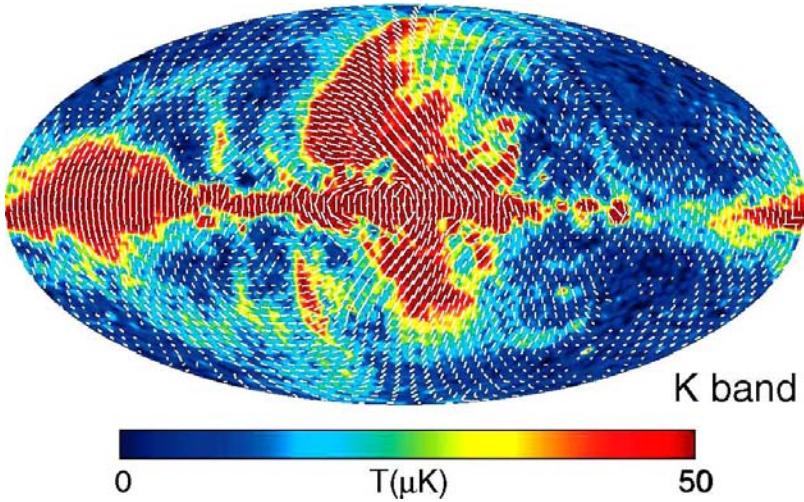


Fig. 17.9 CMB polarization map obtained by the WMAP experiment [68]. Color encodes the degree of polarization in K-band wavelength (23 GHz), $\mathcal{P} \cdot T_0$, white short dashes show the direction of polarization, their length is logarithmically dependent on the magnitude of \mathcal{P} .

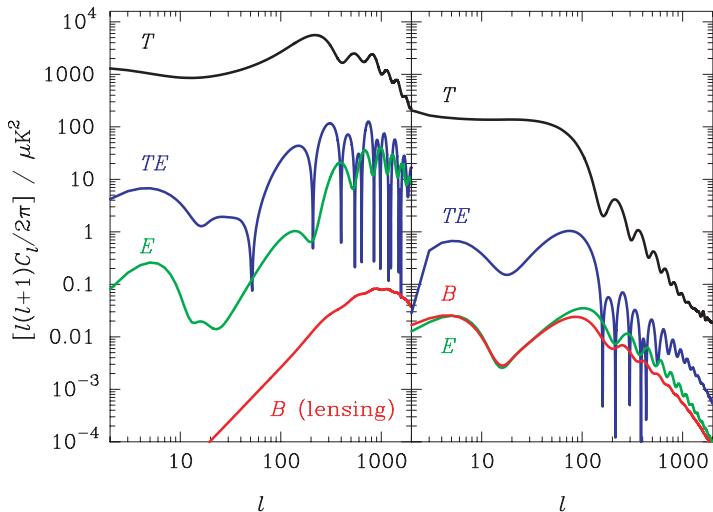


Fig. 17.10 An example of CMB temperature and polarization angular spectra [94]. Left and right panels show contributions of adiabatic scalar perturbations and tensor perturbations, respectively. The values of cosmological parameters are similar to those used in this book, except for unrealistically large tensor-to-scalar ratio, $r = 0.38$, and optical depth at reionization, $\tau_{rei} = 0.17$.

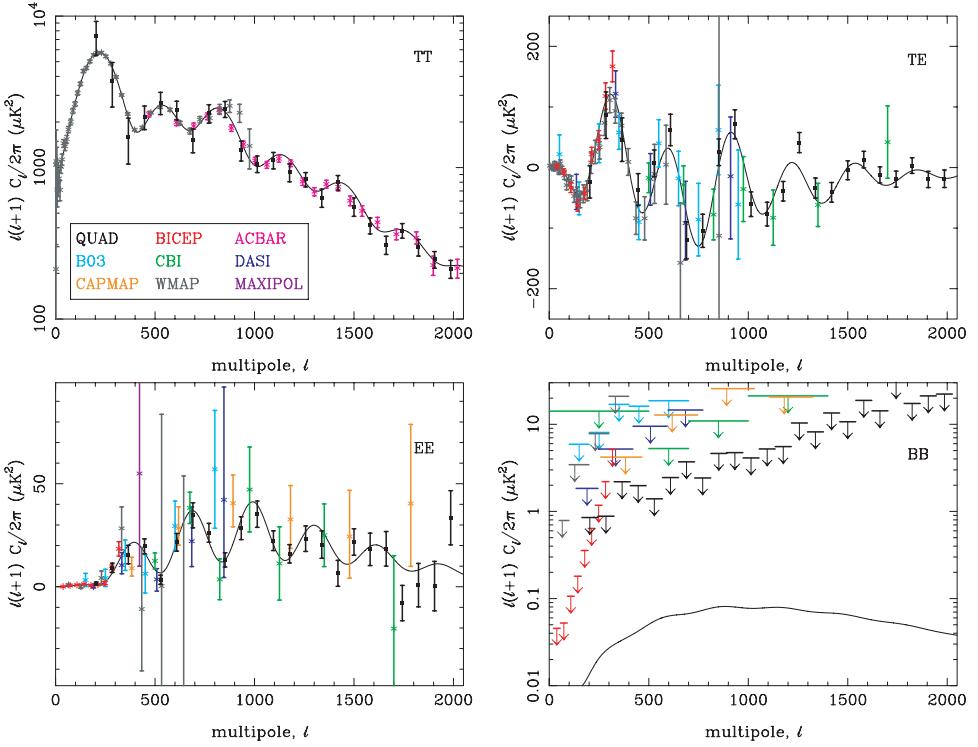


Fig. 17.11 Data on CMB temperature anisotropy and polarization [96]. a) Temperature anisotropy spectrum \mathcal{D}_l^{TT} ; b) Cross correlation spectrum \mathcal{D}_l^{TE} ; c) spectrum of E -mode \mathcal{D}_l^{EE} ; d) bounds on the spectrum of B -mode \mathcal{D}_l^{BB} at 95% C.L.

Appendix A

Exact Solutions for Gravitating Matter

A.1 Gravitational Collapse in Newtonian Theory

Let us study spherically symmetric gravitational collapse of pressureless ideal fluid (“dust”) in the Newtonian theory. We assume for definiteness that the medium is at rest at the initial moment of time, and has spherically symmetric initial mass distribution,

$$t = 0: \quad \mathbf{v} = 0, \quad \rho(t = 0, \mathbf{x}) = \rho_i(r). \quad (\text{A.1})$$

Then the subsequent motion of particles is radial and the spherical symmetry persists during the whole process. The spherically symmetric version of the system of Eqs. (1.1)–(1.3) is

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \phi}{\partial r} \right) = 4\pi G \rho, \quad (\text{A.2a})$$

$$\frac{\partial \rho}{\partial t} + \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 v \rho) = 0, \quad (\text{A.2b})$$

$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial r} + \frac{\partial \phi}{\partial r} = 0, \quad (\text{A.2c})$$

where v is the radial velocity of the fluid.

Let us introduce the variable

$$m(t, r) = r^2 \frac{\partial \phi}{\partial r}. \quad (\text{A.3})$$

In view of Eq. (A.2a) it can be written as

$$m(t, r) = 4\pi G \int_0^r \rho(t, r') r'^2 dr'. \quad (\text{A.4})$$

We see that $m(t, r)$ is proportional to the mass in a ball of radius r at time t . It follows from Eq. (A.2b) that m obeys

$$\frac{\partial}{\partial r} \left(\frac{\partial m}{\partial t} + v \frac{\partial m}{\partial r} \right) = 0. \quad (\text{A.5})$$

This shows that the expression in parenthesis is independent of r . In fact, it is equal to zero. Indeed, for the dust ball of finite size we have $\partial m/\partial r = 0$ and $\partial m/\partial t = 0$ as $r \rightarrow \infty$, since $m(r = \infty)$ is the total mass of the ball which is time-independent.

Let us express the gradient of the gravitational potential in terms of $m(t, r)$ by making use of Eq. (A.2a). Then we have two equations for $v(t, r)$ and $m(t, r)$,

$$\begin{aligned}\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial r} + \frac{m}{r^2} &= 0, \\ \frac{\partial m}{\partial t} + v \frac{\partial m}{\partial r} &= 0.\end{aligned}$$

It is convenient to solve them by treating v and m as independent variables and t and r as unknown functions of these variables. This yields the following system of equations,

$$\begin{aligned}\left(\frac{\partial r}{\partial m} - v \frac{\partial t}{\partial m} \right) \left(1 + \frac{m}{r^2} \frac{\partial t}{\partial v} \right) &= 0, \\ \frac{\partial r}{\partial v} - v \frac{\partial t}{\partial v} &= 0.\end{aligned}\tag{A.6}$$

The first term in Eq. (A.6) cannot vanish for our initial conditions (A.1), and we finally arrive at the system

$$\frac{\partial t}{\partial v} = -\frac{r^2}{m},\tag{A.7a}$$

$$\frac{\partial}{\partial(v^2)} \left(\frac{1}{r} \right) = \frac{1}{2m}.\tag{A.7b}$$

The solution to Eq. (A.7b), consistent with the initial data (A.1), is

$$\frac{1}{r} = \frac{v^2}{2m} + C(m),\tag{A.8}$$

where the function $C(m)$ is determined by the initial mass distribution. Namely, since the initial velocity is assumed to vanish, one has $r = [C(m)]^{-1}$, so (A.4) yields the algebraic equation for $C(m)$,

$$m = 4\pi G \int_0^{[C(m)]^{-1}} \rho_i(r) r^2 dr.\tag{A.9}$$

The solution to Eq. (A.7a) is conveniently written in terms of the combination

$$\xi = -\frac{v}{\sqrt{2mC(m)}},\tag{A.10}$$

where the sign corresponds to collapse: in that case $v < 0$ and $\xi > 0$. The solution to Eq. (A.7a) has the form

$$t = \sqrt{\frac{2}{m}} \frac{1}{[C(m)]^{3/2}} \int_0^\xi \frac{d\xi'}{(\xi'^2 + 1)^2} = \frac{1}{(2m)^{1/2}[C(m)]^{3/2}} \left(\arctan \xi + \frac{\xi}{\xi^2 + 1} \right), \quad (\text{A.11})$$

and the expression for r is simplified,

$$r = \frac{1}{C(m)(\xi^2 + 1)}. \quad (\text{A.12})$$

The last two formulas determine $m(t, r)$ and $\xi(t, r)$ for given $C(m)$, and then one can use (A.4) and (A.10) to evaluate the physical quantities, the mass density and velocity, as functions of radius and time.

The formula (A.4) shows that $m \rightarrow 0$ as $r \rightarrow 0$; we see from (A.9) that $C(m) \rightarrow \infty$ in this limit. Let $\rho_0 = \rho_i(r=0)$ be the initial central density, and let the initial density decrease with the radius. Then one has near the center

$$C(m) = \left(\frac{4\pi G \rho_0}{3m} \right)^{1/3} \cdot [1 - \epsilon(m)], \quad (\text{A.13})$$

where $\epsilon(m)$ is a non-negative function that grows with m starting from zero at $m = 0$. As an example, consider the initial profile whose form near the center is

$$\rho_i(r) = \rho_0 \cdot \left(1 - \frac{r^2}{r_0^2} \right), \quad r \ll r_0. \quad (\text{A.14})$$

We find from (A.9) that near the center, i.e., at small m ,

$$\epsilon(m) = \frac{1}{5r_0^2} \left(\frac{3m}{4\pi G \rho_0} \right)^{2/3}, \quad m \ll \frac{4}{3}\pi r_0^3 G \rho_0.$$

The solution (A.11), (A.12) is valid in a finite time interval only. Indeed, ξ becomes infinite at finite time $t(m)$. We see from (A.11) and (A.13) that the smallest value of $t(m)$ corresponds to $m = 0$ and is given by

$$t_s = \min t(m) = \sqrt{\frac{3\pi}{32G\rho_0}}. \quad (\text{A.15})$$

The solution we obtained cannot be trusted at $t > t_s$. Note that t_s is of the order of the free fall time for a ball of mass density ρ_0 , see Section 1.1, and, in particular, problem 1.2 there.

From the physics viewpoint, the fact that the solution is valid at $t < t_s$ only means that the configuration becomes singular at $t = t_s$. Since the minimum in (A.15) corresponds to $m = 0$, the singularity develops in the center of the ball. The behavior of the solution near the singularity can be found explicitly. Indeed, the time t_s defines the line in the plane (m, ξ) where

$$\left[1 + \frac{3}{2}\epsilon(m) \right] \cdot \int_0^\xi \frac{d\xi'}{(\xi'^2 + 1)^2} = \frac{\pi}{4}.$$

Here we consider small values of m , and hence large values of ξ near the singularity. The explicit relation between m and ξ is

$$\xi = \left[\frac{9\pi}{8} \epsilon(m) \right]^{-1/3}.$$

Equation (A.12) is now an algebraic equation for $m(r)$ which at small m has the following form,

$$r = \left(\frac{3m}{4\pi G\rho_0} \right)^{1/3} \left[\frac{9\pi}{8} \epsilon(m) \right]^{2/3}.$$

We see that the behavior $m(r)$ near the singularity depends on the initial density distribution which determines the function $\epsilon(m)$. For the initial distribution (A.14), we find

$$m(r) = \frac{4\pi}{3} B^2 G \rho_0 r_0^3 \left(\frac{r}{r_0} \right)^{9/7},$$

where B is a numerical constant, $B = (40/9\pi)^{3/7}$. The distributions of the density and velocity are obtained from (A.4) and (A.10),

$$\rho(r) = \frac{3}{7} B^2 \rho_0 \cdot \left(\frac{r}{r_0} \right)^{-12/7}, \quad v(r) = -2Br_0 \sqrt{\frac{2\pi G\rho_0}{3}} \left(\frac{r}{r_0} \right)^{1/7}.$$

Note that the singularity is integrable; the mass is finite in the central region, $M(R) = 4\pi \int_0^R \rho(r)r^2 dr \propto R^{9/7}$. For small R we consider, this mass is smaller than the mass of a black hole of the Schwarzschild radius R , $M_{BH}(R) = M_{Pl}^2 R / 2$ (see, e.g, the book [192]). Hence, for the chosen initial distribution (A.14), a black hole is not formed at the moment when the Newtonian singularity appears. A black hole is obtained for a flatter initial mass distribution, e.g.,

$$\tilde{\rho}(r) = \rho_0 \cdot (1 - (r^2/r_0^2)^n), \quad n \geq 2. \quad (\text{A.16})$$

In that case one has

$$M(R) \propto R^{9/(3+4n)}. \quad (\text{A.17})$$

Problem A.1. Examine the singularity for the initial distribution (A.16) and obtain (A.17), thus showing that the singularity in the Newtonian theory can indicate the formation of a black hole.

A.2 Spherical Collapse of Pressureless Fluid in General Relativity

The analytical solution describing spherically symmetric collapse of pressureless ideal fluid exists in General Relativity as well (see the book [192] for some of the details). It is convenient to work in the comoving reference frame associated with the dust particles, and use the time variable coinciding with the proper time of the

particle (this frame is both synchronous and comoving). The metric in this frame is given by

$$ds^2 = dt^2 - e^{\lambda(t,r)} dr^2 - R^2(t,r) \cdot (d\theta^2 + \sin^2 \theta d\phi^2). \quad (\text{A.18})$$

The particle world line is $r = \text{const}$, whereas $R(t,r)$ is the radius of a sphere through which the particle passes at time t (in the sense that $4\pi R^2$ is the area of that sphere). A somewhat non-trivial fact here is that both particle velocity $u^i = \dot{x}^i$ and mixed metric component g_{0r} can be set equal to zero. It is related to the property that there is no force acting on the dust particles, except for gravity.

Problem A.2. Show that metric in the situation we discuss can be always chosen in the form (A.18). Hint: Make use of the fact that the coordinate transformation $t \rightarrow t + \tau(r)$, $r \rightarrow r$, where $\tau(r)$ is an arbitrary function, keeps the reference frame in the class of comoving frames.

In this Section, unlike in other Chapters of this book, prime denotes the derivative with respect to r . Dot, as usual, denotes the derivative with respect to time t .

The only non-zero component of the energy-momentum tensor in the comoving frame is $T_{00} = \rho$. The independent components of the Einstein equations are

$$(00) \times R^2: R\dot{R}\dot{\lambda} + \dot{R}^2 + 1 - e^{-\lambda}(2RR'' + (R')^2 - RR'\lambda') = 8\pi G\rho R^2 \quad (\text{A.19})$$

$$(0r) \times R: \dot{\lambda}R' - 2\dot{R}' = 0 \quad (\text{A.20})$$

$$(rr) \times R^2: (R')^2 - e^{\lambda}(2R\ddot{R} + \dot{R}^2 + 1) = 0 \quad (\text{A.21})$$

Problem A.3. Show that other Einstein equations are automatically satisfied by solutions to the above equations.

One can integrate Eq. (A.20) once over time; the result is conveniently represented as follows,

$$e^{\lambda}(1 + f(r)) = (R')^2, \quad (\text{A.22})$$

where $f(r)$ is an arbitrary function of the radial coordinate that must obey $f(r) > -1$. One inserts (A.22) into Eq. (A.21) and obtains

$$\dot{R}^2 + 2R\ddot{R} = f(r). \quad (\text{A.23})$$

One then multiplies Eq. (A.23) by \dot{R} and obtains its first integral

$$R\dot{R}^2 = Rf(r) + g(r), \quad (\text{A.24})$$

where we introduced another arbitrary function $g(r)$. Since $R(r=0) = 0$, we impose the requirement $g(0) = 0$. We assume in what follows that $g(r) \geq 0$; this property

will be justified later on, see (A.30). Equation (A.24) can be integrated, and the result is conveniently written in the parametric form,

$$\begin{cases} R = \frac{g}{2f} (\cosh \eta - 1) \\ t_0(r) - t = \frac{g}{2f^{3/2}} (\sinh \eta - \eta) \\ f > 0 \end{cases} \quad \text{(A.25)}$$

or

$$\begin{cases} R = \frac{g}{2|f|} (1 - \cos \eta) \\ t_0(r) - t = \frac{g}{2|f|^{3/2}} (\eta - \sin \eta) \\ f < 0 \end{cases} \quad \text{(A.26)}$$

or

$$\begin{cases} R = \left(\frac{9g}{4}\right)^{1/3} (t_0(r) - t)^{2/3} \\ f = 0 \end{cases} \quad \text{(A.27)}$$

where $t_0(r)$ is an arbitrary function of the radial coordinate. Of course, there are also solutions to Eq. (A.24) with the opposite sign of time. In view of (A.22), the preceding formulas fully determine the space-time metric. Note that our choice of metric (A.18) admits the reparameterization of the radial coordinate, $r = r(\tilde{r})$. Hence, only two of the three arbitrary functions $f(r)$, $g(r)$, $t_0(r)$ are physical. This agrees with the fact that the most general spherically symmetric initial data are determined by two functions of the radial variable, the initial mass density and velocity. It is also important that the evolution inside the sphere of comoving radius r is independent of the initial data outside that sphere, unless there is a caustic singularity.

Equation (A.19) determines the energy density in terms of the functions $f(r)$, $g(r)$, $t_0(r)$. To obtain a simple formula, we take the radial derivative of (A.22), express R'' and insert it into Eq. (A.19). We obtain

$$-f - R \frac{f'}{R'} + R \dot{R} \dot{\lambda} + \dot{R}^2 = 8\pi G \rho R^2. \quad \text{(A.28)}$$

We then take the radial derivative of (A.24) and use Eq. (A.20) to get

$$f' \cdot R = R' \left(R \dot{R} \dot{\lambda} + \frac{g}{R} \right) - g'.$$

Finally, we insert the latter expression into Eq. (A.28), express \dot{R} by using Eq. (A.24) and find

$$8\pi G \rho = \frac{g'}{R^2 R'}. \quad \text{(A.29)}$$

The total mass is

$$M = 4\pi \int_0^{R(t,r_0)} \rho R^2 dR = \frac{g(r_0)}{2G}. \quad (\text{A.30})$$

The solutions (A.25), (A.26) and (A.27) describe both collapse and expansion of a spherically symmetric dust cloud. Let us first study a cloud of finite size r_0 in asymptotically flat space-time. Then the sign chosen in (A.25) is appropriate. The parameter t increases and tends to $t_0(r)$; at $t = t_0(r)$, matter particles of comoving coordinate r reach the center $R = 0$. Collapse occurs for arbitrarily small mass M of the cloud; this property is due to the fact that we completely neglect pressure. To the leading order, the singular behavior at $t \rightarrow t_0(r)$ is

$$R = \left(\frac{9g}{4}\right)^{1/3} (t_0 - t)^{2/3}, \quad e^{\lambda/2} = \left(\frac{2g}{3}\right)^{1/3} \frac{t'_0}{\sqrt{1+f}} (t_0 - t)^{-1/3}.$$

In the comoving frame, radial and angular distances tend to infinity and zero, respectively. The volume also tends to zero, $\sqrt{g} \propto (t_0 - t)$, and the energy density grows,

$$\rho = \frac{1}{12\pi G} \frac{g'}{gt'_0(t_0 - t)}.$$

The accretion of dust on the center gives rise to the formation and growth of a black hole.

Problem A.4. *Find the time of the formation of the black hole horizon. Show that the complete collapse occurs when $R(t, r_0) = g(r_0)$.*

Problem A.5. *Find the solution for the collapse of a homogeneous dust cloud of size r_0 whose particles are initially at rest.*

We now consider a situation which is more interesting from the cosmological viewpoint. Namely, let us assume that the dust fills the whole space, its density is homogeneous at spatial infinity, and there is a small initial overdensity near the origin.

In the first place, let us discuss the homogeneous background solution. To this end, we introduce, instead of $R(r, t)$, $f(r)$ and $g(r)$ the variable $A(r, t)$ and the arbitrary functions $q(r)$ and $\tilde{g}(r)$,

$$R(r, t) = rA(r, t), \quad f(r) = -r^2 q(r), \quad g(r) = r^3 \tilde{g}(r).$$

Then Eq. (A.24) takes the form reminiscent of the Friedmann equation with non-vanishing spatial curvature,

$$\frac{\dot{A}^2}{A^2} = \frac{\tilde{g}(r)}{A^3} - \frac{q(r)}{A^2}. \quad (\text{A.31})$$

The background solution $A = a(t)$ depends on time only, so we set f and \tilde{g} independent of r . Furthermore, spatially flat background is obtained with $q = 0$, and

we can set $\tilde{g} = 1$ without loss of generality. Then the solution to Eq. (A.31) is $A(t) = a(t) \propto t^{2/3}$, and (A.22), (A.29) give

$$e^\lambda = a^2(t), \quad \frac{8\pi}{3}G\rho = \frac{4}{9t^2} = \left(\frac{\dot{a}}{a}\right)^2,$$

as should be the case for the Universe filled with dust. Note that this solution is obtained from (A.27) with $t_0 = 0$ by the replacement $t \rightarrow -t$.

Consider now the perturbed medium at early times, when the deviation from the background solution is small, and hence can be studied within the linearized theory. To see what choice of the functions $q(r)$, $\tilde{g}(r)$ and $t_0(r)$ is appropriate, we again use Eq. (A.31) and set

$$A(r, t) = a(t)[1 + h(t, r)],$$

where h is the small perturbation. We also write $\tilde{g}(r) = 1 + \zeta(r)$ and treat $\zeta(r)$ and $q(r)$ as small quantities. Then the linearized Eq. (A.31) reads

$$\frac{\partial h}{\partial t} + \frac{1}{t}h = -\left(\frac{2}{3}\right)^{1/3} \frac{1}{2t^{1/3}}q(r) + \frac{1}{3t}\zeta(r).$$

This equation has a time-independent solution, $h = \zeta/3$. It corresponds to a transformation of the coordinate r , and hence is irrelevant. In what follows we use the gauge choice $\zeta = 0$. The solution to the homogeneous equation decreases in time, $h = c(r)/t$. It corresponds to non-vanishing $t_0(r)$. Since this solution is singular at early times, it describes strongly inhomogeneous early Universe, and hence it is also irrelevant. Using the terminology of Section 3.2.1, this solution is the decaying mode. The relevant solution is proportional to $q(r)$,

$$h = -\frac{3}{10}\left(\frac{2}{3}\right)^{1/3} \cdot q(r) \cdot t^{2/3}. \quad (\text{A.32})$$

We obtain from (A.29) that the relative density perturbation is

$$\frac{\delta\rho}{\rho} = -\frac{1}{r^2}(r^3h)' \quad (\text{A.33a})$$

$$= \text{const} \cdot \frac{1}{r^2} \frac{\partial}{\partial r}[r^3q(r)] \cdot t^{2/3}. \quad (\text{A.33b})$$

In accordance with the results of Sections 1.2 and 4.3, the density perturbation grows as $t^{2/3} \propto a(t)$. The overdensity situation corresponds to $q > 0$, i.e., $f < 0$; the density contrast is finite as $r \rightarrow 0$, provided that $q(r)$ is positive and has finite limiting value $q(0)$. The quantity $q(0)$ characterizes the primordial perturbation amplitude. The function $q(r)$ vanishes as $r \rightarrow \infty$ according to the assumption of the asymptotic homogeneity.

So, the solution we discuss here is given by (A.26) with $g = r^3$, $t_0 = 0$ and reversed sign of time,

$$\begin{aligned} R &= \frac{r}{2q(r)}(1 - \cos \eta) \\ t &= \frac{1}{2q^{3/2}(r)}(\eta - \sin \eta) \end{aligned} \quad (\text{A.34})$$

This solution reduces to the homogeneous one as $\eta \rightarrow 0$. At later times, a particle with comoving coordinate r lags behind the Hubble flow (in our case this is the flow of particles at spatial infinity), and starts moving back to the center of the cloud at $\eta = \pi$. The cosmic time at that moment is

$$t = \frac{\pi}{2q^{3/2}(r)}. \quad (\text{A.35})$$

This shows that the central, densest region of the cloud collapses first. The density contrast of matter that begins to collapse is of the order of the average density,

$$[\rho(r) - \rho] \sim \rho, \quad (\text{A.36})$$

where ρ is the average density, $\rho = \rho(r = \infty)$. Like in the linear theory, the size of the region that collapses at time t is smaller than the Hubble size, $R \lesssim t$. To see this, we recall that $f(r) > -1$, i.e., $r^2 q(r) < 1$. Hence, at the beginning of collapse one has

$$\frac{R}{t} = \frac{2}{\pi} r \sqrt{q(r)} < \frac{2}{\pi},$$

which proves the above statement.

In the matter dominated Universe we discuss, the entire overdense cloud collapses to the center in the end. Different regions do that at different moments of time, however. A particle with comoving coordinate r reaches the origin at $\eta = 2\pi$, so the time interval between the moment the particle stops and the moment it reaches the center is given precisely by (A.35). The entire collapse takes about one Hubble time.

Note that the time-independent part of the density contrast $\delta\rho/\rho$ is absent in the comoving synchronous gauge we use here, while this part exists in the conformal Newtonian gauge and dominates in the superhorizon regime, see (4.18). In the comoving synchronous gauge, there is a time-independent part of metric perturbation, $\delta g_{rr}/g_{rr} \propto f(r)$, see (A.22). It dominates until the perturbation becomes subhorizon.

Problem A.6. Show that the time-independent part of metric perturbation dominates over the time-dependent part at early times, and that these parts become comparable when the physical size of the perturbation becomes of the order of the Hubble size, i.e., when $a(t)r_0 \sim t$, where r_0 is the coordinate size of the perturbation. Hint: Perform the analysis at the linear level.

The case of the homogeneous overdense ball of finite coordinate size r_0 is described by setting $q(r) = q_0 = \text{const}$ at $r < r_0$ and $q(r) = q_0(r_0/r)^3$ at $r > r_0$.

Indeed, the density contrast (A.33b) is independent of r inside the ball and vanishes outside it. It follows from (A.34) that the interior of the ball collapses simultaneously at time $t = t_c = \pi/q_0^{3/2}$. The density contrast at that time *calculated in the linear theory* by using (A.33a) and (A.32) is

$$\delta_c \equiv \left(\frac{\delta\rho}{\rho} \right)_{lin}(t_c) \quad (\text{A.37a})$$

$$= \frac{3}{20}(12\pi)^{2/3} = 1.686 \quad (\text{A.37b})$$

Obviously, this quantity has no direct physical meaning, but sometimes it is useful, see Section 7.2. Note that the value of δ_c is independent of the size of the overdense ball and of the time the collapse ends up. It is worth noting that this property is not valid for the Universe with dark energy.

Problem A.7. Consider the overdense ball with the density contrast corresponding to the function $q(r)$ of the following form,

$$q(r) = \text{const} \cdot e^{-r/r_0}.$$

Define, in analogy to (A.37a), the quantity δ_c as the density contrast at the point $r = r_0$ calculated in the linear theory at the time when the dust particle of the comoving coordinate $r = r_0$ reaches the center in the full theory. Show that $\delta_c = \frac{1}{10}(12\pi)^{2/3} = 1.124$.

To end this Section, let us point out that the above results get modified in the presence of the cosmological constant. In that case, the analog of Eq. (A.21) reads

$$(R')^2 - e^\lambda(2R\ddot{R} + \dot{R}^2 + 1) = -8\pi G\rho_\Lambda e^\lambda, \quad (\text{A.38})$$

where $\rho_\Lambda = \text{const}$. Equation (A.20) remains intact. The modified Eq. (A.19) is irrelevant for our purposes.

The formula (A.22) remains valid, and repeating the argument we obtain instead of Eq. (A.31) the following equation,

$$\frac{\dot{A}^2}{A^2} = \frac{1}{A^3} - \frac{q(r)}{A^2} + \frac{8\pi G}{3}\rho_\Lambda, \quad (\text{A.39})$$

where we have set $\tilde{g} = 1$. The solution describing the termination of expansion and subsequent collapse may now be absent. The right hand side of Eq. (A.39) has a minimum at $A = 3/(2q)$ and it can vanish only for $q^3 \geq 18\pi G\rho_\Lambda$. Assuming that $q(r)$ decreases with r at interesting r , we obtain the maximum physical size of a region that stops its expansion,

$$R_{max} = \frac{3r}{2q} = \frac{3r\sqrt{q}}{2q^{3/2}} = \frac{r\sqrt{q}}{\sqrt{3}H_\Lambda} < \frac{1}{\sqrt{3}H_\Lambda},$$

where $H_\Lambda^2 = (8\pi/3)G\rho_\Lambda$ and we used the inequality $r^2q < 1$. Hence, the size of a collapsing region cannot exceed the de Sitter horizon size characteristic of Λ -domination.

Appendix B

Derivation of Linearized Einstein Equations

The purpose of this Appendix is to present the main elements of the derivation of the linearized Einstein and covariant conservation equations for cosmological perturbations of various helicities.

To obtain the general expression for the linearized Einstein tensor, we make use of the fact that the metric $g_{\mu\nu} = a^2 \gamma_{\mu\nu} \equiv a^2(\eta_{\mu\nu} + h_{\mu\nu})$ and metric $\gamma_{\mu\nu} \equiv (\eta_{\mu\nu} + h_{\mu\nu})$ are related by conformal transformation of the form (I.A.61) with $\varphi = \log a$. Hence, the relation between the Einstein tensors of the two metrics is given by (I.A.64), i.e.,

$$G_{\mu\nu}(g) = G_{\mu\nu}(\gamma) - 2 \frac{\nabla_\mu \nabla_\nu a}{a} + 4 \frac{\partial_\mu a \partial_\nu a}{a^2} + \gamma_{\mu\nu} \gamma^{\lambda\rho} \left(2 \frac{\nabla_\lambda \nabla_\rho a}{a} - \frac{\partial_\lambda a \partial_\rho a}{a^2} \right), \quad (\text{B.1})$$

where covariant derivatives are evaluated with the metric $\gamma_{\mu\nu}$. We raise one index with the metric $g^{\mu\nu} = a^{-2} \gamma^{\mu\nu}$ and obtain

$$\begin{aligned} G_\nu^\mu(g) &= \frac{1}{a^2} \left[G_\nu^\mu(\gamma) - 2\gamma^{\mu\lambda} \frac{\nabla_\lambda \nabla_\nu a}{a} + 4\gamma^{\mu\lambda} \frac{\partial_\lambda a \partial_\nu a}{a^2} \right. \\ &\quad \left. + \delta_\nu^\mu \gamma^{\lambda\rho} \left(2 \frac{\nabla_\lambda \nabla_\rho a}{a} - \frac{\partial_\lambda a \partial_\rho a}{a^2} \right) \right]. \end{aligned}$$

We now recall the explicit expression for the covariant derivative and linearize the latter expression. The result for the linearized Einstein tensor is

$$\begin{aligned} \delta G_\nu^\mu &= \frac{1}{a^2} \left[R_\nu^\mu(\gamma) + 2h^{\mu\lambda} \frac{\partial_\nu \partial_\lambda a}{a} - 4h^{\mu\lambda} \frac{\partial_\nu a \partial_\lambda a}{a^2} + 2\eta^{\mu\lambda} \gamma_{\lambda\nu}^\sigma \frac{\partial_\sigma a}{a} \right. \\ &\quad \left. - \delta_\nu^\mu \left(\frac{1}{2}R(\gamma) + 2h^{\lambda\rho} \frac{\partial_\lambda \partial_\rho a}{a} - h^{\lambda\rho} \frac{\partial_\lambda a \partial_\rho a}{a^2} + 2\eta^{\lambda\rho} \gamma_{\lambda\rho}^\sigma \frac{\partial_\sigma a}{a} \right) \right], \quad (\text{B.2}) \end{aligned}$$

where $R_\nu^\mu(\gamma)$, $R(\gamma)$ and $\gamma_{\mu\nu}^\lambda$ are the linearized Ricci tensor, curvature scalar and Christoffel symbols calculated with the metric $\gamma_{\mu\nu}$.

The linearized Christoffel symbols are obtained directly from the general definition (I.A.23),

$$\gamma_{\mu\nu}^\lambda = \frac{1}{2}(\partial_\mu h_\nu^\lambda + \partial_\nu h_\mu^\lambda - \partial^\lambda h_{\mu\nu}). \quad (\text{B.3})$$

Hereafter indices are raised and lowered with Minkowski metric. Now, we recall the definition (I.A.41) of the Ricci tensor and see that only the first two terms survive at the linear level. They give to the linear order

$$R_\nu^\mu(\gamma) = \frac{1}{2}(\partial_\nu \partial_\lambda h^{\mu\lambda} + \partial^\mu \partial_\lambda h_\nu^\lambda - \partial_\lambda \partial^\lambda h_\nu^\mu - \partial^\mu \partial_\nu h_\lambda^\lambda) \quad (\text{B.4})$$

and

$$R(\gamma) = \partial_\mu \partial_\nu h^{\mu\nu} - \partial_\mu \partial^\mu h_\nu^\nu. \quad (\text{B.5})$$

Hence, all objects entering (B.2) are explicitly described. Note that

$$\partial_\mu a = \delta_{\mu 0} a', \quad \partial_\mu \partial_\nu a = \delta_{\mu 0} \delta_{\nu 0} a''.$$

Problem B.1. *The Einstein tensor $(\bar{G}_\nu^\mu + \delta G_\nu^\mu)$ must be covariant under gauge transformations (2.31) to the linear order in ξ^μ . Show explicitly that this property indeed holds for the right hand side of (B.2).*

Let us impose the gauge $h_{0i} = 0$ and write the Christoffel symbols entering (B.2),

$$\gamma_{00}^0 = \frac{1}{2}h'_{00}, \quad \gamma_{0i}^0 = \frac{1}{2}\partial_i h_{00}, \quad \gamma_{ij}^0 = -\frac{1}{2}h'_{ij}. \quad (\text{B.6})$$

The components of the linearized Ricci tensor are

$$\begin{aligned} R_0^0(\gamma) &= \frac{1}{2}(h'' + \Delta h_{00}), \\ R_i^0(\gamma) &= \frac{1}{2}(\partial_i h' - \partial_j h'_{ij}), \\ R_j^i(\gamma) &= \frac{1}{2}(\partial_i \partial_k h_{jk} + \partial_j \partial_k h_{ik} + h''_{ij} - \Delta h_{ij} + \partial_i \partial_j h_{00} - \partial_i \partial_j h), \end{aligned} \quad (\text{B.7})$$

and the curvature scalar is given by

$$R(\gamma) = h'' + \Delta h_{00} + \partial_i \partial_j h_{ij} - \Delta h. \quad (\text{B.8})$$

Here

$$h = h_{ii},$$

summation over spatial (Latin) indices is performed with the 3-dimensional Euclidean metric, and $\Delta \equiv \partial_i \partial_i$. We insert these expressions into (B.2) and obtain explicit formulas for the components of the linearized Einstein tensor,

$$a^2 \delta G_0^0 = -3h_{00} \left(\frac{a'}{a} \right)^2 - \frac{1}{2}\partial_i \partial_j h_{ij} + \frac{1}{2}\Delta h - \frac{a'}{a}h', \quad (\text{B.9a})$$

$$a^2 \delta G_i^0 = \frac{1}{2}\partial_i h' - \frac{1}{2}\partial_j h'_{ij} + \frac{a'}{a}\partial_i h_{00}, \quad (\text{B.9b})$$

$$a^2 \delta G_j^i = \frac{1}{2}\partial_i \partial_k h_{jk} + \frac{1}{2}\partial_j \partial_k h_{ik} + \frac{1}{2}h''_{ij} - \frac{1}{2}\Delta h_{ij} + \frac{1}{2}\partial_i \partial_j h_{00} - \frac{1}{2}\partial_i \partial_j h$$

$$\begin{aligned}
& + \frac{a'}{a} h'_{ij} - \delta_j^i \left[\frac{1}{2} h'' + \frac{1}{2} \Delta h_{00} + \frac{1}{2} \partial_l \partial_k h_{lk} - \frac{1}{2} \Delta h + 2 \frac{a''}{a} h_{00} \right. \\
& \left. - \left(\frac{a'}{a} \right)^2 h_{00} + \frac{a'}{a} (h'_{00} + h') \right]
\end{aligned} \tag{B.9c}$$

We do not impose the conformal Newtonian gauge in this Appendix, but work in the gauge $h_{0i} = 0$ for the time being. The perturbation of the energy-momentum tensor has the general form

$$\delta T_0^0 = \delta \rho, \tag{B.10a}$$

$$\delta T_i^0 = -(\rho + p)v_i, \tag{B.10b}$$

$$T_j^i = -\delta_j^i \cdot p - \Pi_j^i, \tag{B.10c}$$

where Π_{ij} is the symmetric traceless tensor of the anisotropic stress, $\Pi_{ii} = 0$, and by definition $\Pi_j^i = \Pi_{ij}$. For ideal fluid one has $\Pi_{ij} = 0$. In the latter case, the linearized covariant conservation equations for the energy-momentum tensor are particularly simple. We write by definition

$$\nabla_\mu T_\nu^\mu = \partial_\mu T_\nu^\mu + \Gamma_{\mu\lambda}^\mu T_\nu^\lambda - \Gamma_{\mu\nu}^\lambda T_\lambda^\mu = 0, \tag{B.11}$$

where the Christoffel symbols are evaluated with the metric $a^2(\eta_{\mu\nu} + h_{\mu\nu})$. Their explicit forms in the gauge $h_{0i} = 0$ are

$$\Gamma_{00}^0 = \frac{a'}{a} + \frac{1}{2} h'_{00}, \quad \Gamma_{0i}^0 = \Gamma_{00}^i = \frac{1}{2} \partial_i h_{00}, \tag{B.12a}$$

$$\Gamma_{0j}^i = \frac{a'}{a} \delta_{ij} - \frac{1}{2} h'_{ij}, \tag{B.12b}$$

$$\Gamma_{ij}^0 = \frac{a'}{a} (1 - h_{00}) \delta_{ij} - \frac{a'}{a} h_{ij} - \frac{1}{2} h'_{ij}, \tag{B.12c}$$

$$\Gamma_{jk}^i = -\frac{1}{2} (\partial_j h_{ik} + \partial_k h_{ij} - \partial_i h_{jk}). \tag{B.12d}$$

Here we keep the zeroth and first order terms. Equations (2.40), (2.41) are obtained by inserting the above expressions into (B.11).

In the general case, the linearized Einstein and covariant conservation equations are conveniently written in the sectors of definite helicity. The helicity decompositions for the velocity and anisotropic stress tensor are

$$v_i = V_i^T + \partial_i v, \tag{B.13a}$$

$$\Pi_{ij} = \Pi_{ij}^{TT} + \partial_i \Pi_j^T + \partial_j \Pi_i^T + (p + \rho) \left(-\frac{1}{2} \delta_{ij} + \frac{3}{2} \frac{\partial_i \partial_j}{\Delta} \right) \pi. \tag{B.13b}$$

Here v and π are velocity and anisotropic stress potentials, respectively, and the factor $(\rho + p)$ is extracted for convenience. The helicity decomposition of the metric in the gauge $h_{0i} = 0$ is, see (2.48),

$$h_{00} = 2\Phi, \quad h_{0i} = 0, \quad (\text{B.14a})$$

$$h_{ij} = -2\Psi\delta_{ij} + 2\partial_i\partial_j E + (\partial_i W_j^T + \partial_j W_i^T) + h_{ij}^{TT}. \quad (\text{B.14b})$$

In the *helicity-2* (tensor) sector we have $h_{00} = 0$, $h_{ij} = h_{ij}^{TT}$, $T_0^0 = T_i^0 = 0$, $T_{ij} = -\Pi_{ij}^{TT}$. We insert these expressions into the linearized Einstein equations and arrive at single equation

$$\partial_\eta^2 h_{ij}^{TT} + 2\frac{a'}{a}\partial_\eta h_{ij}^{TT} - \Delta h_{ij}^{TT} = -16\pi G a^2 \cdot \Pi_{ij}^{TT}. \quad (\text{B.15})$$

Covariant conservation equations are obeyed in the tensor sector automatically. Hereafter we have in mind without indicating explicitly that the right hand side of the Einstein equations is the sum over all non-interacting components of the cosmic medium.

The 00-component of the Einstein equations is identically zero in the *helicity-1* (vector) sector. Other Einstein equations in the vector sector are

$$0i: \quad \partial_\eta \Delta W_i^T = 16\pi G a^2 (\rho + p) V_i^T, \quad (\text{B.16a})$$

$$ij: \quad \partial_\eta^2 W_i^T + 2\frac{a'}{a}\partial_\eta W_i^T = -16\pi G a^2 (\rho + p) \Pi_i^T, \quad (\text{B.16b})$$

while only one of the covariant conservation equations is non-trivial,

$$\partial_\eta [(\rho + p) V_i^T] + 4\frac{a'}{a}(\rho + p) V_i^T + \Delta \Pi_i^T = 0. \quad (\text{B.17})$$

In the ideal fluid case, we have $\Pi_i^T = 0$ and arrive at the equations written in Section 2.3.2.

Problem B.2. Show that Eq. (B.16b) is a consequence of Eqs. (B.16a) and (B.17).

Let us now turn to the *helicity-0* (scalar) sector. Inserting the scalar part of the expressions (B.14) into (B.9c) yields the following form of the ij -components of the Einstein tensor,

$$\begin{aligned} \delta G_j^i = \frac{1}{a^2} & \left\{ \partial_i \partial_j \left[\Psi + \Phi + E'' + 2\frac{a'}{a}E' \right] + \delta_{ij} \left[2\Psi'' - \Delta\Psi - \Delta\Phi - \Delta E'' \right. \right. \\ & \left. \left. + \frac{a'}{a}(4\Psi' - 2\Phi' - 2\Delta E') - 4\frac{a''}{a}\Phi + 2\frac{a'^2}{a^2}\Phi \right] \right\}. \end{aligned}$$

We make use of (B.10) and extract the trace and longitudinal parts of the ij -components of the Einstein equations. Then the ij -components give two scalar

equations. These equations together with the 00- and $0i$ -components combine into the system of four equations:

$$\begin{aligned} \frac{a^2}{2}\delta G_0^0: \quad & -\Delta\Psi + \frac{a'}{a}(3\Psi' - \Delta E') - 3\frac{a'^2}{a^2}\Phi = 4\pi Ga^2\delta\rho, \\ \frac{a^2}{2}\delta G_i^0: \quad & -\Psi' + \frac{a'}{a}\Phi = -4\pi G(\rho + p)v, \\ \frac{a^2}{2}\delta G_j^i: \quad & \Psi'' - \frac{1}{3}(\Delta\Phi + \Delta\Psi + \Delta E'') + \frac{a'}{a}\left(2\Psi' - \Phi' - \frac{2}{3}\Delta E'\right) \\ & - 2\frac{a''}{a}\Phi + \frac{a'^2}{a^2}\Phi = -4\pi Ga^2\delta p, \\ & \Delta\left(\Phi + \Psi + E'' + 2\frac{a'}{a}E'\right) = -12\pi Ga^2(\rho + p)\pi. \end{aligned}$$

The covariant conservation equations are

$$\begin{aligned} \delta\rho' + 3\frac{a'}{a}(\delta\rho + \delta p) + (\rho + p)(\Delta v + 3\Psi' - \Delta E') &= 0, \\ [(\rho + p)v]' + \delta p + (\rho + p)\left(4\frac{a'}{a}v + \pi + \Phi\right) &= 0. \end{aligned}$$

In the ideal fluid case, when $\pi = 0$, these equations in the conformal Newtonian gauge $E = 0$ reduce to equations of Section 2.3.3. In the general case, they have the form (8.58) in the conformal Newtonian gauge.

Problem B.3. Consider gauge transformations (2.31) with gauge functions obeying (2.33); recall that these gauge transformations are consistent with the gauge $h_{0i} = 0$. The relation (2.33) shows that ξ_i and ξ_0 are expressed in terms of one arbitrary function $\sigma(x)$,

$$\xi_i = -\partial_i\sigma, \quad \xi_0 = \sigma'.$$

Show that the transformation law under these gauge transformations is

$$\begin{aligned} \Phi &\rightarrow \Phi + \sigma'' + \frac{a'}{a}\sigma', \quad \Psi \rightarrow \Psi + \frac{a'}{a}\sigma', \quad E \rightarrow E - \sigma, \\ \rho &\rightarrow \rho + \rho'\sigma', \quad \delta p \rightarrow \delta p + p'\sigma', \quad v \rightarrow v - \sigma', \quad \pi \rightarrow \pi. \end{aligned}$$

Show that the linearized Einstein and covariant conservation equations transform in a covariant way under these gauge transformations.

For completeness, let us give the expressions for the perturbation of the Einstein tensor in arbitrary gauge. We use the notations introduced in (2.46), (2.47), (2.48).

The formulas of Section 2.3.1 are valid in the tensor sector irrespective of the gauge, and perturbations h_{ij}^{TT} are gauge-invariant.

The general formula (B.2) gives for the vector sector

$$\begin{aligned}\delta G_0^0 &= 0, \\ \delta G_i^0 &= \frac{1}{2a^2} \left(\Delta Z_i^T - \partial_\eta \Delta W_i^T \right), \\ \delta G_i^0 &= \frac{1}{a^2} \left[\frac{1}{2} \left(\partial_\eta \Delta W^T - \Delta Z^T \right) - 2 \left(\frac{a''}{a} - 2 \frac{a'^2}{a^2} \right) Z_i^T \right], \\ \delta G_j^i &= -\frac{1}{a^2} \left\{ \frac{1}{2} \left[\partial_\eta \left(\partial_i Z_j^T + \partial_j Z_i^T \right) - \partial_\eta^2 \left(\partial_i W_j^T + \partial_j W_i^T \right) \right] \right. \\ &\quad \left. + \frac{a'}{a} \left[\partial_i Z_j^T + \partial_j Z_i^T - \partial_\eta \left(\partial_i W_j^T + \partial_j W_i^T \right) \right] \right\}.\end{aligned}$$

The components of the linearized Einstein tensor in the scalar sector are

$$\delta G_0^0 = -\frac{2}{a^2} \left[\Delta \Psi + \frac{a'}{a} (-\Delta Z + \Delta E' - 3\Psi') + 3 \frac{a'^2}{a^2} \Phi \right], \quad (\text{B.18a})$$

$$\delta G_i^0 = \frac{2}{a^2} \left[-\partial_i \Psi' + \frac{a'}{a} \partial_i \Phi \right], \quad (\text{B.18b})$$

$$\delta G_0^i = -\frac{2}{a^2} \left[-\partial_i \Psi' + \frac{a'}{a} \partial_i \Phi - \left(2 \frac{a'^2}{a^2} - \frac{a''}{a} \right) \partial_i Z \right], \quad (\text{B.18c})$$

$$\begin{aligned}\delta G_j^i &= \frac{1}{a^2} \left\{ \partial_i \partial_j \left[\Psi + \Phi - Z' + E'' - 2 \frac{a'}{a} (Z - E') \right] \right. \\ &\quad \left. + \delta_{ij} \left[2\Psi'' - \Delta \Psi - \Delta \Phi + \Delta Z' - \Delta E'' + \frac{a'}{a} (4\Psi' - 2\Phi' \right. \right. \\ &\quad \left. \left. + 2\Delta Z - 2\Delta E') - 4 \frac{a''}{a} \Phi + 2 \frac{a'^2}{a^2} \Phi \right] \right\}. \quad (\text{B.18d})\end{aligned}$$

We note that in the scalar sector, the gauge transformations (2.31) with $\xi_i = \partial_i \xi$ read

$$\delta Z = \xi' + \xi_0, \quad \delta E = \xi, \quad \delta \Phi = \xi'_0 + \frac{a'}{a} \xi_0, \quad \delta \Psi = \frac{a'}{a} \xi_0. \quad (\text{B.19})$$

According to the general transformation rule for tensors, the perturbation of the Einstein tensor transforms as follows,

$$\begin{aligned}\delta_\xi(\delta G_0^0) &= \partial_0 \bar{G}_0^0 \cdot \xi_0, \\ \delta_\xi(\delta G_i^0) &= \partial_i \xi_0 \cdot \bar{G}_0^0 - \partial_k \xi_0 \cdot \bar{G}_i^k, \\ \delta_\xi(\delta G_0^i) &= \partial_i \xi'_0 \cdot \bar{G}_0^0 - \partial_k \xi'_0 \cdot \bar{G}_k^i, \\ \delta_\xi(\delta G_j^i) &= \partial_0 \bar{G}_j^i \cdot \xi_0,\end{aligned} \quad (\text{B.20})$$

where $\bar{G}_{\mu\nu}$ is the unperturbed Einstein tensor, see (B.21).

Problem B.4. Show that transformation (B.20) is indeed the linearized version of the general transformation rule for tensors.

Problem B.5. Show that the right hand side of (B.18) indeed has the property (B.20).

Problem B.6. Show that the expressions for δG_ν^μ in the vector sector are covariant under gauge transformations analogous to (B.19).

It is useful for some applications to have the explicit expressions for the quadratic action for perturbations. The quadratic action in the tensor sector is given by (2.56). The calculation in the scalar and vector sectors, similar to that performed in Section 2.3.1, would involve the analysis of the contribution coming from δT_ν^μ , see (2.54). Hence, we follow somewhat different route. Namely, we find the quadratic part of the gravitational action alone, without assuming that the background metric obeys the Einstein equations.¹ We perform the calculation for the gravitational action including both the Einstein–Hilbert action (I.A.47) and the cosmological constant term,

$$-\Lambda \int d^4x \sqrt{-g}.$$

We proceed along the lines of Section 2.3.1, but do not use the equations of motion for the background metric. As an example, we write the integrand in the variation of the Einstein–Hilbert action to the quadratic order in $h_{\mu\nu}$ and $\delta h_{\mu\nu}$,

$$\sqrt{-g}G^{\mu\nu}\delta g_{\mu\nu} \implies a^4 \left(\frac{1}{2}h_\lambda^\lambda \bar{G}^{\mu\nu} + a^{-2}\eta^{\mu\lambda}\delta G_\lambda^\nu - a^{-2}h^{\mu\lambda}\bar{G}_\lambda^\nu \right) a^2\delta h_{\mu\nu}.$$

The expressions for δG_ν^μ are given above, while the background values are well-known,

$$\bar{G}_0^0 = \frac{3a'^2}{a^4}, \quad \bar{G}_j^i = \delta_j^i \left(2\frac{a''}{a^3} - \frac{a'^2}{a^4} \right). \quad (\text{B.21})$$

We proceed in a similar way with the cosmological constant term. Then the procedure analogous to the one employed in Section 2.3.1 gives the quadratic action in the vector sector,

$$S_V = \frac{1}{32\pi G} \int d^3x d\eta a^2 \left[(\partial_k Z_i^T)^2 + 6 \left(\frac{a'^2}{a^2} - H_\Lambda^2 a^2 \right) (Z_i^T)^2 + 2\partial_\eta W_i^T \Delta Z_i^T + (\partial_\eta \partial_k W_i^T)^2 - 2 \left(2\frac{a''}{a} - \frac{a'^2}{a^2} - 3H_\Lambda^2 a^2 \right) (\partial_k W_i^T)^2 \right], \quad (\text{B.22})$$

where $(\partial_k Z_i)^2 \equiv \partial_k Z_i \partial_k Z_i$, etc., and

$$H_\Lambda^2 = \frac{8\pi}{3}G\Lambda.$$

Note that the expressions in parentheses in (B.22) vanish, if there is no matter in the Universe and it expands due to the cosmological constant. The quadratic action in the scalar sector is

$$S_S = \frac{1}{8\pi G} \int d^4x a^2 \left[-2\Phi\Delta\Psi - 2\Psi'\Delta Z + 2\Psi'\Delta E' + 3\Psi\Psi'' - \Psi\Delta\Psi + \frac{a'}{a}(2\Phi\Delta Z - 2\Phi\Delta E' + 6\Phi\Psi') + \left(-\frac{9}{2}\frac{a'^2}{a^2} + \frac{3}{2}H_\Lambda^2 a^2 \right) \Phi^2 \right]$$

¹The linear in perturbations part of the gravitational action does not vanish in this situation. This part is straightforward to obtain, but it is irrelevant as long as the Einstein equations for the background are taken into account in the end.

$$\begin{aligned}
& + \left(-\frac{9}{2} \frac{a'^2}{a^2} - \frac{9}{2} H_\Lambda^2 a^2 \right) \Psi^2 + \left(\frac{a'^2}{a^2} - H_\Lambda^2 a^2 \right) \left(9\Phi\Psi - 3\Phi\Delta E + \frac{3}{2}(\partial_i Z)^2 \right) \\
& + \left(2\frac{a''}{a} - \frac{a'^2}{a^2} - 3H_\Lambda^2 a^2 \right) \left(-\Psi\Delta E - \frac{1}{2}(\Delta E)^2 \right)
\end{aligned}$$

Let us discuss briefly what happens in the absence of matter. In the vector sector, the variable Z_i^T is non-dynamical, as it enters the action without time-derivatives. It can be integrated out: one finds the equation for it by varying the action with respect to Z_i^T , finds Z_i^T from that equation and substitutes it back into the action. The result is the action for a single variable W_i^T . If the background metric obeys the Einstein equations with the cosmological constant, the latter action for W_i^T vanishes identically. Hence, W_i^T is not subject to any equation, and it can be an arbitrary function of coordinates and time. The entire configuration (Z_i^T, W_i^T) is then pure gauge. This is the way the gauge invariance of General Relativity in the de Sitter background shows up. Similar properties are inherent in the scalar sector, where the non-dynamical variables are Φ and Z .

Problem B.7. Prove the above statements.

Appendix C

Gaussian Random Variables and Gaussian Random Fields

C.1 Properties of Gaussian Random Variables

We recall in this Appendix the definitions and basic properties of the Gaussian random variables and Gaussian random fields.

The notion of a random variable applies when there is an ensemble of systems. Some variable q takes a certain value in each of them, but this value randomly varies from one system to another.¹ The ensemble can then be characterized by the distribution function² $F(q)$ of the random variable q . Let the variable q be continuous (the discussion is straightforwardly generalized to discrete random variables). The distribution function determines the probability that the random variable takes value in the interval from q to $q + dq$,

$$P(q, q + dq) = F(q) dq.$$

Clearly $F(q)$ is real and positive. Since the probability that the random variable takes *some* value is equal to 1, the distribution function obeys the normalization condition

$$\int F(q) dq = 1,$$

where the integration is performed over the entire range of possible values of q . In what follows we assume for definiteness that this range is the line $(-\infty, +\infty)$.

The Gaussian random variable is defined by the property that its distribution is Gaussian,

$$F(q) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(q - q_0)^2}{2\sigma^2}\right], \quad (\text{C.1})$$

¹A simple example is the position, at a given moment of time, of a particle that undergoes the Brownian motion.

²We use the same notation q for the random variable and its value in a given system from the ensemble. This does not lead to a confusion.

where q_0 and σ are real parameters, and $\sigma > 0$. The parameter q_0 is the average value of the random variable and σ is its variance,

$$\langle q \rangle = \int_{-\infty}^{+\infty} dq \, q \, F(q) = q_0, \quad \langle (q - q_0)^2 \rangle = \int_{-\infty}^{+\infty} dq \, (q - q_0)^2 \, F(q) = \sigma^2.$$

Any linear function $\tilde{q} = aq + b$ is the Gaussian random variable with the average value $\tilde{q}_0 = aq_0 + b$ and variance $a\sigma$. We often set the average values of the Gaussian random variables equal to zero in what follows, since this can be achieved by the redefinition $q \rightarrow q - q_0$.

An interesting example of the Gaussian random variable is the coordinate of a quantum mechanical oscillator in its ground state. The ground state wave function of the oscillator Hamiltonian $\hat{H} = \frac{1}{2}(\hat{p}^2 + \omega^2\hat{q}^2)$ is

$$\psi(q) = \left(\frac{\omega}{\pi}\right)^{1/4} e^{-\frac{\omega}{2}q^2},$$

so the probability density in the coordinate representation reads

$$F(q) = |\psi(q)|^2 = \sqrt{\frac{\omega}{\pi}} e^{-\omega q^2}.$$

By comparing this expression to (C.1), we find that the coordinate q is indeed the Gaussian random variable with zero average value and variance $(2\omega)^{-1/2}$.

The momentum of a quantum mechanical oscillator is also the Gaussian random variable. Furthermore, any linear combination of the coordinate and momentum with real coefficients is the Gaussian random variable.

Problem C.1. *Prove the last statement above.*

We note that the operators \hat{q} and \hat{p} are linear combinations of the creation and annihilation operators. Hence, any Hermitean linear combination of the creation and annihilation operators is the Gaussian random variable, provided that the oscillator is in its ground state.

Consider now a set of several random variables q_1, \dots, q_N . The distribution function $F(q_1, \dots, q_N)$ now depends on all these variables. The set is called Gaussian, if this function has the form

$$F(q_1, \dots, q_N) \equiv F(q) = \mathcal{N} \exp\left(-\frac{1}{2} q_m M_{mn} q_n + L_n q_n\right), \quad (\text{C.2})$$

where the summation from 1 to N over repeated indices m, n is assumed, the coefficients M_{mn} and L_n are independent of q and M_{mn} is a symmetric non-degenerate real matrix. The positive factor \mathcal{N} normalizes the distribution function,

$$\int F(q) d^N q = 1. \quad (\text{C.3})$$

This integral is convergent if and only if the quadratic combination $q_m M_{mn} q_n$ is positive-definite, i.e., all eigenvectors of the matrix M are positive.

The coefficients L_n determine the average values of the random variables q_1, \dots, q_N . Indeed, let us make the linear transformation

$$q_m = \tilde{q}_m + D_{mn}L_n, \quad (\text{C.4})$$

where

$$D_{mn} = (M^{-1})_{mn}$$

is the matrix inverse to M . The distribution function for the set \tilde{q}_n is obtained then by inserting (C.4) into (C.2),

$$F(\tilde{q}) = \mathcal{N} \exp\left(-\frac{1}{2}\tilde{q}_m M_{mn} \tilde{q}_n\right). \quad (\text{C.5})$$

Due to the symmetry $\tilde{q}_n \rightarrow -\tilde{q}_n$, we have $\langle \tilde{q}_n \rangle = 0$, so that

$$\langle q_m \rangle = D_{mn}L_n.$$

Problem C.2. Show that the normalization factor is (C.2) and (C.5) equals $\mathcal{N} = (\det M)^{-1/2}(2\pi)^{-N/2}$.

It is useful to note that any N independent linear combinations of q_1, \dots, q_N with real coefficients again make a set of the Gaussian random variables.

Let us study the averages of products of the Gaussian random variables $\langle q_{n_1} \dots q_{n_k} \rangle$, called correlators. To calculate them, one introduces the generating function

$$Z(j) \equiv Z(j_1, \dots, j_N) = \int F(q_1, \dots, q_N) e^{j_n q_n} d^N q. \quad (\text{C.6})$$

The averages we are after are expressed in terms of the derivatives of the generating function at zero arguments,

$$\langle q_{n_1} \dots q_{n_k} \rangle = \left[\frac{\partial^k Z}{\partial j_{n_1} \dots \partial j_{n_k}} \right]_{j_m=0}. \quad (\text{C.7})$$

The generating function is straightforwardly calculated. Let us set L_m in (C.2) equal to zero, without loss of generality. Hence, the average values $\langle q_n \rangle$ vanish. To calculate the integral in (C.6), we change the integration variables in a way similar to (C.4),

$$q_m = \tilde{q}_m + D_{mn}j_n.$$

We obtain

$$Z(j) = e^{\frac{1}{2}j_m D_{mn} j_n} \int F(\tilde{q}) d^N \tilde{q} = e^{\frac{1}{2}j_m D_{mn} j_n}. \quad (\text{C.8})$$

We see that the simplest non-vanishing correlator is

$$\langle q_m q_n \rangle = D_{mn}.$$

Furthermore, the result (C.8) shows that the averages of products of odd number of the Gaussian random variables vanish, while the averages of products of even number of variables are expressed in terms of the matrix $D = M^{-1}$. The latter obey what is called the *Wick theorem* in quantum field theory: to calculate the average $\langle q_{n_1} \dots q_{n_{2k}} \rangle$ one divides $q_{n_1}, \dots, q_{n_{2k}}$ into pairs, replaces each pair (q_{n_r}, q_{n_s}) with $D_{n_r n_s}$, takes the product over pairs, and finally sums up over all partitions. As an example, the average of the product of four Gaussian random variables is

$$\langle q_p q_r q_s q_t \rangle = D_{pr} D_{st} + D_{ps} D_{rt} + D_{pt} D_{rs}. \quad (\text{C.9})$$

The inverse is valid too: if all correlators of random variables obey the Wick theorem with the matrix D_{mn} , and this matrix is positive-definite, then the set of random variables is Gaussian with the distribution function (C.2) and $L_m = 0$, $M = D^{-1}$.

Problem C.3. Derive the formula (C.9) by direct calculation.

Problem C.4. Prove the Wick theorem for Gaussian random variables with zero average values.

Problem C.5. Prove the inverse statement.

We have not discussed yet the possible dependence of random variables on time. Let us briefly consider this possibility. Let the variables $q_n(t)$ obey some equations of motion³ which uniquely determine $q_1(t), \dots, q_N(t)$ for given initial values $q'_1 \equiv q_1(t_i), \dots, q'_N \equiv q_N(t_i)$,

$$q_n(t) = f_n(q'_1, \dots, q'_N; t). \quad (\text{C.10})$$

Let us assume that Eq. (C.10) has the inverse, i.e., that one can uniquely reconstruct the initial data once $q_1(t), \dots, q_N(t)$ are known,

$$q'_n \equiv q_n(t_i) = g_n(q_1(t), \dots, q_N(t)). \quad (\text{C.11})$$

For a given initial distribution function $F(q'_1, \dots, q'_N; t_i)$, the correlators at time t are calculated as follows,⁴

$$\langle q_{n_1}(t) \dots q_{n_k}(t) \rangle = \int f_{n_1}(q'; t) \dots f_{n_k}(q'; t) \cdot F(q'_1, \dots, q'_N; t_i) d^N q'. \quad (\text{C.12})$$

Upon the obvious change of integration variables,

$$q'_n = g_n(q), \quad (\text{C.13})$$

³We do not consider here the possibility that the equations of motion themselves involve random forces or random parameters, so that the evolution of $q_n(t)$ is not deterministic. The latter case is in fact of relevance for numerous physical systems. However, it is sufficient for our purposes to consider the simpler case of the deterministic evolution.

⁴It is sometimes useful to study correlators of random variables taken at different times. The corresponding generalization of the formula (C.12) is straightforward.

this expression becomes

$$\langle q_{n_1}(t) \dots q_{n_k}(t) \rangle = \int q_{n_1} \dots q_{n_k} \cdot F(q_1, \dots, q_N; t) d^N q,$$

where

$$F(q_1, \dots, q_N; t) = J(t) \cdot F(g_1(q), \dots, g_N(q); t_i)$$

is the distribution function at time t . Here

$$J(t) = \frac{\partial(g_1, \dots, g_N)}{\partial(q_1, \dots, q_N)}$$

is the Jacobian of the change of variables (C.13).

If the initial distribution $F(q'_1, \dots, q'_N; t_i)$ is Gaussian, and the evolution is linear and non-degenerate, then the set of variables $q_1(t), \dots, q_N(t)$ is Gaussian at any time t . Indeed, the variables $q_1(t), \dots, q_N(t)$ are linear functions of the initial values q'_1, \dots, q'_N by definition of the linear evolution; since the evolution is non-degenerate, the linear functions $f_1(q'), \dots, f_N(q')$ are linearly independent. In view of the above remark, this immediately implies that the set $q_1(t), \dots, q_N(t)$ is Gaussian.

Problem C.6. Consider a classical oscillator with the Hamiltonian

$$H = \frac{1}{2}(p^2 + \omega^2 q^2).$$

Let the coordinate and momentum be independent Gaussian random variables at initial time which average values q_0, p_0 and variances σ_q and σ_p (the independence of random variables means that the distribution function is factorized, $F(q, p) = F_q(q) \cdot F_p(p)$). Find the distribution function of $q(t), p(t)$ at time t .

To end the discussion of the Gaussian random variables, let us make the following comment. In some cases, instead of the complete set of random variables q_1, \dots, q_N , one is interested in a smaller set of N' of their linear combinations

$$r_a = \sum_{n=1}^N c_{an} q_n, \quad a = 1, \dots, N'; \quad N' < N. \quad (\text{C.14})$$

Here c_{an} are real coefficients, and the variables r_a are linear independent in the sense that none of their linear combinations vanishes identically. The variables r_a are random, and their distribution function is

$$\bar{F}(r_1, \dots, r_{N'}) \equiv \bar{F}(r) = \int F(q) \delta^{(N')} \left(r_a - \sum_{n=1}^N c_{an} q_n \right) d^N q. \quad (\text{C.15})$$

If q_1, \dots, q_N are Gaussian random variables, then $(r_1, \dots, r_{N'})$ is also a set of Gaussian random variables. To see this, let us find the generating function for the correlators of r_a ,

$$\bar{Z}(j_1, \dots, j_{N'}) = \int \bar{F}(r) e^{j_a r_a} d^{N'} r.$$

We insert the expression (C.15) here, integrate first over r_a and then over q_n . Assuming for definiteness that $L_n = 0$ in (C.2), we then obtain

$$\bar{Z}(j_a) = e^{\frac{1}{2} j_a \bar{D}_{ab} j_b}, \quad (\text{C.16})$$

where $\bar{D}_{ab} = (\bar{M}^{-1})_{ab}$ and

$$\bar{M}_{ab} = c_{am} M_{mn} c_{bn}.$$

The generating function (C.16) has the same form as (C.8). Therefore, the correlators of the random variables r_a obey the Wick theorem, and hence $r_1, \dots, r_{N'}$ are indeed Gaussian. Note that since M_{mn} is positive-definite, \bar{M}_{ab} is positive-definite as well,

$$r_a \bar{M}_{ab} r_b = (r_a c_{am}) M_{mn} (r_b c_{bn}) > 0,$$

as should be the case for the Gaussian random variables.

C.2 Gaussian Random Fields

Let us now turn to random fields and consider for definiteness a real scalar field at a given moment of time. It is convenient for our purposes to take the following viewpoint on the field $\phi(\mathbf{x})$. Let us first consider a theory in large but finite volume, say, in a cube of size L . Let us also discretize the space, so that it contains a large but finite number of points \mathbf{x}_i . The simplest possibility is the cubic lattice; the index i is then a set of three integers (i_1, i_2, i_3) , where i_1, i_2, i_3 run from 1 to $N = L/a$, and a is the lattice spacing. We then define the variable $\phi_i \equiv \phi(\mathbf{x}_i)$ at each lattice site. Hence, we are dealing with the system of finite, albeit large number of degrees of freedom. Let ϕ_i be a set of random variables. Then in the limit $a \rightarrow 0$ we arrive at the random field $\phi(\mathbf{x})$. If, furthermore, the set $\{\phi_i\}$ is Gaussian, then $\phi(\mathbf{x})$ is the Gaussian random field.

For finite lattice spacing and finite spatial volume, the distribution function $F[\{\phi_i\}]$ has the standard meaning. However, the expressions for correlators like

$$\langle \phi(\mathbf{x}) \phi(\mathbf{y}) \rangle = \int F[\phi] \phi(\mathbf{x}) \phi(\mathbf{y}) D\phi$$

become functional integrals in the limit $a \rightarrow 0$. Hence, the distribution function is an inconvenient notion in the theory of random fields, and the primary objects are correlation functions $\langle \phi(\mathbf{x}_1) \dots \phi(\mathbf{x}_k) \rangle$. Generally speaking, these are generalized functions. Of direct physical significance are the quantities obtained by some sort of averaging in space, say the correlators of variables $\int f(\mathbf{x}, \mathbf{y}) \phi(\mathbf{y}) d^3y$, where $f(\mathbf{x}, \mathbf{y})$ is a smooth function considerably different from zero when \mathbf{y} is in the vicinity of \mathbf{x} . These correlators are conventional functions rather than generalized functions (distributions).

Many results valid for Gaussian random variables remain valid for Gaussian random fields. This is clear from the above discretized construction. In particular, a finite set of linear independent integrals (cf. (C.14))

$$r_a = \int c_a(\mathbf{x})\phi(\mathbf{x}) d^3x$$

is a set of Gaussian random variables (cf. (C.14)). Also, linear evolution of a Gaussian field keeps it Gaussian. Finally, a Gaussian and only Gaussian field obeys the Wick theorem: all many-point correlation functions either vanish (if the number of fields is odd) or are expressed in terms of the two-point correlation function according to the Wick theorem formulated before Eq. (C.9). As an example,

$$\begin{aligned} & \langle \phi(\mathbf{x}_1)\phi(\mathbf{x}_2)\phi(\mathbf{x}_3)\phi(\mathbf{x}_4) \rangle \\ &= D(\mathbf{x}_1, \mathbf{x}_2)D(\mathbf{x}_3, \mathbf{x}_4) + D(\mathbf{x}_1, \mathbf{x}_3)D(\mathbf{x}_2, \mathbf{x}_4) + D(\mathbf{x}_1, \mathbf{x}_4)D(\mathbf{x}_2, \mathbf{x}_3), \end{aligned}$$

where

$$D(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x})\phi(\mathbf{y}) \rangle \quad (\text{C.17})$$

is the two-point correlation function. Hereafter we assume, without loss of generality, that the average value of the field vanishes, $\langle \phi(\mathbf{x}) \rangle = 0$. We note that correlators of fields at different times, $\langle \phi(\mathbf{x}_1, t_1) \cdots \phi(\mathbf{x}_k, t_k) \rangle$ also obey the Wick theorem in the case of linear evolution. We do not use correlators at unequal times.

Hence, the central object in the theory of Gaussian random fields is the two-point correlation function (C.17). If the ensemble of systems is invariant under spatial translations, this correlator depends only on $(\mathbf{x} - \mathbf{y})$. If, furthermore, the ensemble is invariant under spatial rotations, then the correlator is a function of the scalar $|\mathbf{x} - \mathbf{y}|$,

$$\langle \phi(\mathbf{x})\phi(\mathbf{y}) \rangle = D(|\mathbf{x} - \mathbf{y}|), \quad (\text{C.18})$$

The latter properties are valid in the Euclidean space; they can be generalized to 3-sphere and 3-hyperboloid.

Continuing with the correlation function (C.18), it is convenient to perform the Fourier transformation,

$$\phi(\mathbf{x}) = \int d^3k e^{i\mathbf{k}\mathbf{x}}\phi(\mathbf{k}),$$

where real-valuedness of $\phi(\mathbf{x})$ implies

$$\phi^*(\mathbf{k}) = \phi(-\mathbf{k}). \quad (\text{C.19})$$

According to the above discussion, $\phi(\mathbf{k})$ is the Gaussian random field. Its correlation function is

$$\begin{aligned}\langle \phi(\mathbf{k})\phi(\mathbf{k}') \rangle &= \int \frac{d^3x d^3y}{(2\pi)^6} e^{-i\mathbf{k}\mathbf{x}-i\mathbf{k}'\mathbf{y}} D(|\mathbf{x}-\mathbf{y}|) \\ &= \int \frac{d^3z d^3y}{(2\pi)^6} e^{-i(\mathbf{k}+\mathbf{k}')\mathbf{y}} e^{-i\mathbf{k}\mathbf{z}} D(|\mathbf{z}|).\end{aligned}$$

Therefore,

$$\langle \phi(\mathbf{k})\phi(\mathbf{k}') \rangle = \frac{P(k)}{(2\pi)^3} \delta(\mathbf{k} + \mathbf{k}'), \quad (\text{C.20})$$

where

$$P(k) = \int d^3z e^{-i\mathbf{k}\mathbf{z}} D(|\mathbf{z}|).$$

The function $P(k)$ depends only on $k \equiv |\mathbf{k}|$ in the case of invariance under spatial rotations. It is non-negative. To see this, consider the variance of the variable

$$\phi_f = \int f(\mathbf{k})\phi(\mathbf{k}) d^3k,$$

where $f(\mathbf{k})$ is an arbitrary smooth function that obeys the only condition $f^*(\mathbf{k}) = f(-\mathbf{k})$, so that ϕ_f is real. Clearly, $\langle \phi_f^2 \rangle \geq 0$. On the other hand, we have

$$\langle \phi_f^2 \rangle = \int d^3k |f(\mathbf{k})|^2 \frac{P(k)}{(2\pi)^3}.$$

Since $f(\mathbf{k})$ is arbitrary, the right hand side is non-negative for non-negative $P(k)$ only.

The function $P(k)$ is called the *power spectrum* of the Gaussian random field $\phi(\mathbf{x})$. The same term, power spectrum, is also used for

$$\mathcal{P}(k) = \frac{k^3 P(k)}{2\pi^2}.$$

To see the meaning of the latter, consider the fluctuation of the field $\phi(\mathbf{x})$, namely

$$\langle \phi^2(\mathbf{x}) \rangle = \int d^3k d^3k' e^{i(\mathbf{k}+\mathbf{k}')\mathbf{x}} \langle \phi(\mathbf{k})\phi(\mathbf{k}') \rangle = \int d^3k \frac{P(k)}{(2\pi)^3}.$$

After angular integration we arrive at the integral over the absolute value of momentum,

$$\langle \phi^2(\mathbf{x}) \rangle = \int_0^\infty \frac{dk}{k} \mathcal{P}(k).$$

We see that $\mathcal{P}(k)$ determines the contribution of a decimal interval of momenta around a given value k into the fluctuation of the field $\phi(\mathbf{x})$.

To summarize, a homogeneous and isotropic Gaussian random field with zero average value is completely characterized by its power spectrum $P(k)$, or, equivalently, $\mathcal{P}(k)$. Its two-point correlation function in the momentum representation is given by (C.20).

In complete analogy to the quantum mechanical oscillator considered in Section C.1, a linear (free) bosonic quantum field (e.g., scalar field) in its vacuum state is the Gaussian random field. Indeed, the vacuum correlators of free quantum fields obey the Wick theorem. In fact, much more general statement is valid. Namely, let $A_{\mathbf{k}}^\dagger$ and $A_{\mathbf{k}}$ be the creation and annihilation operators of free quantum scalar field that obey the standard commutational relations,

$$[A_{\mathbf{k}}, A_{\mathbf{k}'}^\dagger] = \delta(\mathbf{k} - \mathbf{k}').$$

Let the system be in the vacuum state $|0\rangle$, such that $A(\mathbf{k})|0\rangle = 0$ for all \mathbf{k} . Then the field

$$\hat{\phi}(\mathbf{x}) = \int \frac{d^3 k}{(2\pi)^{3/2}} (f(\mathbf{k}) e^{-i\mathbf{k}\mathbf{x}} A_{\mathbf{k}}^\dagger + f^*(\mathbf{k}) e^{i\mathbf{k}\mathbf{x}} A_{\mathbf{k}})$$

is the Gaussian random field for any complex function $f(\mathbf{k})$.

Problem C.7. Show that the vacuum correlators of the field $\hat{\phi}(\mathbf{x})$ obey the Wick theorem, i.e., this field is Gaussian.

Problem C.8. Find the power spectrum of the field $\hat{\phi}(\mathbf{x})$.

This page is intentionally left blank

Appendix D

Fermions in Gravitational Fields

D.1 Lorentz Group as Gauge Group

General Relativity in a certain sense is a gauge theory whose gauge group is the Lorentz group. It is this aspect that we consider in this Section.

We begin with the construction of gauge theory with the Lorentz group serving as the gauge group, and pretend for the time being that the Lorentz group does not have anything to do with space-time. Then the construction proceeds in the same way as for any non-Abelian theory. Recall (see, e.g., the book [32]), that the infinitesimal Lorentz transformation is defined in terms of small parameters $\omega^{\alpha\beta}$, which are antisymmetric in α, β :

$$\omega^{\alpha\beta} = -\omega^{\beta\alpha} \ll 1.$$

The indices take values $\alpha, \beta = 0, 1, 2, 3$. We use the first letters of the Greek alphabet on purpose here, and reserve the letters μ, ν, \dots for space-time coordinates. The Lorentz transformation for an arbitrary tensor or spinor A has the general form

$$A \rightarrow \Lambda A = \left(1 + \frac{1}{2} \omega^{\alpha\beta} \Sigma_{\alpha\beta} \right) A, \quad (\text{D.1})$$

where $\Sigma_{\alpha\beta}$ are the generators of the Lorentz group in the representation of A . They are also antisymmetric, $\Sigma_{\alpha\beta} = -\Sigma_{\beta\alpha}$, and obey the commutational relations of the Lorentz algebra,

$$[\Sigma_{\alpha\beta}, \Sigma_{\gamma\delta}] = \eta_{\beta\gamma} \Sigma_{\alpha\delta} + \eta_{\alpha\delta} \Sigma_{\beta\gamma} - \eta_{\alpha\gamma} \Sigma_{\beta\delta} - \eta_{\beta\delta} \Sigma_{\alpha\gamma}.$$

As an example, the infinitesimal Lorentz transformation of a vector A^α reads

$$A^\alpha \rightarrow A^\alpha + \omega_\beta^\alpha A^\beta,$$

where indices are raised and lowered with the Minkowski tensor $\eta_{\alpha\beta}$. This gives the expression for the generators in the vector representation,

$$\text{vector: } [\Sigma_{\alpha\beta}]^\gamma_\delta = \delta_\alpha^\gamma \eta_{\beta\delta} - \delta_\beta^\gamma \eta_{\alpha\delta}.$$

The generators in the Dirac spinor representation are

$$\text{Dirac spinor: } \Sigma_{\alpha\beta} = \frac{1}{4}[\gamma_\alpha, \gamma_\beta],$$

where γ_α are the Dirac matrices.

Consider now some field theory invariant under global (independent of space-time point) Lorentz transformations (D.1), still discarding the relevance of the Lorentz group for space-time. Let us extend this theory by requiring the invariance under local transformations

$$A(x) \rightarrow \Lambda(x)A(x) = \left(1 + \frac{1}{2}\omega^{\alpha\beta}(x)\Sigma_{\alpha\beta}\right) A(x), \quad (\text{D.2})$$

where the parameters of the transformation depend on the space-time point. Then the theory has the Lorentz *gauge* symmetry. Like in other gauge theories, we have to introduce the gauge field $\Gamma_\mu(x)$ (in gauge theories of particle physics, gauge fields are usually denoted by A_μ rather than Γ_μ). This field is called the *spin connection* in the context of General Relativity. It takes values in the Lorentz algebra, and enters the covariant derivative in the standard way,¹

$$\mathcal{D}_\mu A = \partial_\mu A + \Gamma_\mu A, \quad (\text{D.3})$$

where Γ_μ is the spin connection in the representation of the field A (we simplify the notations here as much as we can). This construction in a detailed form is as follows. The field $\Gamma_\mu(x)$ is determined by its real components² $\Gamma_\mu^{\alpha\beta}(x)$, which are antisymmetric in α, β , and the spin connection (gauge field) is

$$\Gamma_\mu(x) = \frac{1}{2}\Gamma_\mu^{\alpha\beta}(x)\Sigma_{\alpha\beta},$$

where $\Sigma_{\alpha\beta}$ are the same objects as in (D.1). Say, in the vector representation

$$[\Gamma_\mu^{\alpha\beta}\Sigma_{\alpha\beta}]^\gamma_\delta = \Gamma_\mu^\gamma{}_\delta - \Gamma_\mu^\gamma{}_\delta,$$

so that

$$(\Gamma_\mu A)^\alpha = \Gamma_\mu^{\alpha\beta} A_\beta. \quad (\text{D.4})$$

For the Dirac spinor

$$\Gamma_\mu = \frac{1}{4}\Gamma_\mu^{\alpha\beta}\gamma_\alpha\gamma_\beta. \quad (\text{D.5})$$

¹For the time being we use the terminology of the gauge field theories, see, e.g., the book [194]. The covariant derivative we introduce here does not coincide with the expression involving the Christoffel symbols for the covariant derivative in the coordinate basis. This is the reason for using the special notation \mathcal{D}_μ .

²Another commonly used notation is $\omega_\mu^{\alpha\beta} \equiv \Gamma_\mu^{\alpha\beta}$.

As always in gauge theory, the main requirement imposed on the covariant derivative is that it must transform homogeneously under the gauge transformation (D.2) accompanied by the transformation of the gauge field,

$$(\mathcal{D}_\mu A)(x) \rightarrow \Lambda(x)(\mathcal{D}_\mu A)(x).$$

This yields the following transformation law for the gauge field

$$\Gamma_\mu \rightarrow \Lambda \Gamma_\mu \Lambda^{-1} + \Lambda \partial_\mu \Lambda^{-1} = \Gamma_\mu - \frac{1}{2} (\partial_\mu \omega + [\Gamma_\mu, \omega]), \quad (\text{D.6})$$

where $\omega(x) = \omega^{\alpha\beta}(x)\Sigma_{\alpha\beta}$.

Until now we have followed the standard procedure of constructing a gauge field theory, discarding the relationship of the Lorentz group to space-time. We now deviate from this logic. The point is that the local (gauge) Lorentz symmetry is automatically present in the Riemannian geometry. To see this, let us introduce four vectors³ \mathbf{e}_α at every space-time point x whose components are e_α^μ . The subscript α merely enumerates these vectors; this set of vectors is called vierbein (or tetrade). Vierbein can be viewed as the basis in the tangent space at a given point of the space-time manifold. Let us choose the orthonormalized vierbein, in the sense that

$$\mathbf{e}_\alpha \cdot \mathbf{e}_\beta \equiv e_{\alpha\mu} e_\beta^\mu = \eta_{\alpha\beta}, \quad (\text{D.7})$$

where the space-time index is raised and lowered with the space-time metric $g_{\mu\nu}$, i.e., $e_{\alpha\mu} = g_{\mu\nu} e_\alpha^\nu$. The vierbein index α is raised and lowered with the Minkowski metric, and the condition (D.7) is written as

$$e_\mu^\alpha e_\beta^\mu = \delta_\beta^\alpha. \quad (\text{D.8})$$

The latter relation shows that the matrix e_μ^α is inverse to the matrix e_α^μ . Hence,

$$e_\mu^\alpha e_\alpha^\nu = \delta_\mu^\nu,$$

which gives

$$e_\mu^\alpha e_{\alpha\nu} = g_{\mu\nu}. \quad (\text{D.9})$$

The latter property is instrumental for choosing convenient vierbeins in concrete cases.

There are various vierbeins $\{\mathbf{e}_\alpha\}$ at a given space-time point. All of them obey the orthonormality condition (D.7). Therefore, the vierbeins are related by the Lorentz transformations

$$\mathbf{e}^\alpha \rightarrow \Lambda^\alpha_\beta \mathbf{e}^\beta. \quad (\text{D.10})$$

We emphasize that these transformations are unrelated to the transformations of space-time coordinates, and hence do not touch the index μ . In the component form, the transformation is $e_\mu^\alpha \rightarrow \Lambda^\alpha_\beta e_\mu^\beta$.

³In this Appendix, unlike in other Chapters, we use bold face notation for 4-vectors.

We see that the Lorentz group acts independently at each space-time point. Space-time tensors and spinors can be viewed as objects that transform in an appropriate way under these Lorentz transformations. As an example, a 4-vector \mathbf{A} , whose components in the coordinate basis are A^μ , has the following components in the Lorentz basis,

$$A^\alpha = e_\mu^\alpha A^\mu.$$

Geometrically, this construction means that the vector $\mathbf{A}(x)$ belongs to the tangent space at the point x , and A^α are components of this vector in the basis \mathbf{e}_α . The Lorentz components of the vector \mathbf{A} transform under the Lorentz rotation of the vierbein (D.10) as $A^\alpha \rightarrow \Lambda^\alpha{}_\beta A^\beta$. In general, the transformation of a tensor or spinor has the form (D.1). Note that the Lorentz components of tensors and spinors are scalars under the space-time coordinate transformations; as an example, under the coordinate transformation $x'^\mu = x^\mu(x^\nu)$ one has $A'^\alpha(x') = A^\alpha(x)$.

Since the Lorentz transformation of vierbein (D.10) can be performed independently at any space-time point, the theory is automatically invariant under gauge transformations (D.2). Clearly, it must be possible to express the gauge field, i.e. the spin connection Γ_μ , in terms of the fundamental objects of the Riemannian geometry, the metric $g_{\mu\nu}(x)$ and vierbein $e_\mu^\alpha(x)$.

To find this expression, consider a vector field $\mathbf{A}(x)$. We recall (see Section I.A.2), that the vector $\tilde{\mathbf{A}}(\tilde{x})$ obtained by the parallel transport of the vector $\mathbf{A}(x)$ from the point x to the point \tilde{x} with coordinates $\tilde{x}^\mu = x^\mu + dx^\mu$ has the following components in the coordinate basis,

$$\tilde{A}^\mu(\tilde{x}) = A^\mu(x) - \Gamma_{\nu\lambda}^\mu(x) A^\nu(x) dx^\lambda,$$

and the covariant differential of the vector field equals the difference between the vector $\mathbf{A}(\tilde{x})$, that exists at the point \tilde{x} , and parallel transported vector $\tilde{\mathbf{A}}(\tilde{x})$,

$$\nabla_\nu A^\mu dx^\nu = A^\mu(\tilde{x}) - \tilde{A}^\mu(\tilde{x}). \quad (\text{D.11})$$

The covariant derivative (D.3) has the same meaning, but in the Lorentz basis,

$$(\mathcal{D}_\nu A)^\alpha dx^\nu = A^\alpha(\tilde{x}) - \tilde{A}^\alpha(\tilde{x}). \quad (\text{D.12})$$

The right hand side of (D.12) is $e_\mu^\alpha [A^\mu(\tilde{x}) - \tilde{A}^\mu(\tilde{x})]$ (the difference between $e_\mu^\alpha(\tilde{x})$ and $e_\mu^\alpha(x)$ is irrelevant to the linear order). By equating the covariant differentials (D.11) and (D.12), we find

$$\mathcal{D}_\nu A^\alpha = e_\mu^\alpha \nabla_\nu A^\mu.$$

In view of (D.4) and the definition of the covariant derivative in the coordinate basis (I.A.17), the latter equality is written as

$$\partial_\nu A^\alpha + \Gamma_\nu^{\alpha\beta} A_\beta = e_\mu^\alpha \left[\partial_\nu (e_\beta^\mu A^\beta) + \Gamma_{\nu\lambda}^\mu e^{\beta\lambda} A_\beta \right].$$

Making use of (D.8), we obtain finally

$$\Gamma_{\nu}^{\alpha\beta} = e_{\mu}^{\alpha} \nabla_{\nu} e^{\beta\mu}. \quad (\text{D.13})$$

We emphasize that ∇_{ν} always denotes the covariant derivative “acting” on space-time indices only; in particular,

$$\nabla_{\nu} e^{\beta\mu} = \partial_{\nu} e^{\beta\mu} + \Gamma_{\nu\lambda}^{\mu} e^{\beta\lambda}.$$

Two comments regarding the result (D.13) are in order. First, it follows from (D.8) that

$$e_{\mu}^{\alpha} \nabla_{\nu} e^{\beta\mu} + (\nabla_{\nu} e_{\mu}^{\alpha}) e^{\beta\mu} = 0,$$

so the right hand side of (D.13) is antisymmetric in α, β , as it should. Second, we see from (D.13) that the “total covariant derivative” of vierbein vanishes,

$$\mathbf{D}_{\nu} e^{\alpha\mu} \equiv \partial_{\nu} e^{\alpha\mu} + \Gamma_{\nu\lambda}^{\mu} e^{\alpha\lambda} + \Gamma_{\nu}^{\alpha\beta} e_{\beta}^{\mu} = 0.$$

By definition, the total covariant derivative \mathbf{D}_{ν} “acts” on both space-time and Lorentz indices. This definition is sometimes convenient.

Problem D.1. *Show that the spin connection (D.13) transforms according to (D.6) under the local Lorentz transformations.*

Problem D.2. *Using the covariant derivative (D.3), one can define the curvature $\mathcal{R}_{\mu\nu}$ by the relation*

$$[\mathcal{D}_{\mu}, \mathcal{D}_{\nu}] A = \mathcal{R}_{\mu\nu} A \equiv \frac{1}{2} \mathcal{R}_{\mu\nu}^{\alpha\beta} \Sigma_{\alpha\beta} A,$$

which is valid for any tensor or spinor A (note that in the gauge field theory, the object analogous to $\mathcal{R}_{\mu\nu}$, is called the field strength and denoted by $F_{\mu\nu}$). Find the relationship between the components $\mathcal{R}_{\mu\nu}^{\alpha\beta}$ and the components of the Riemann tensor.

D.2 Fermion Action and Dirac Equation

A Lagrangian written in terms of Lorentz tensors and spinors is invariant under coordinate transformations if it involves the covariant derivatives (D.3) in the combination

$$\mathcal{D}_{\alpha} \equiv e_{\alpha}^{\mu} \mathcal{D}_{\mu}.$$

Indeed, we have already noted that Lorentz tensors and spinors are scalars under coordinate transformations; the same holds for the covariant derivative \mathcal{D}_{α} . If, furthermore, the Lagrangian is invariant under the local Lorentz rotations of vierbein, i.e., gauge Lorentz transformations, the theory is completely covariant, i.e., independent of the choice of both the coordinate frame and the vierbein field. This observation enables one to construct covariant theories for any tensor and spinor

fields. An important example is the Dirac fermion field. The above discussion shows that the simplest covariant generalization of the Dirac action is

$$S_F = \int d^4x \sqrt{-g} (i\bar{\psi}\gamma^\alpha \mathcal{D}_\alpha \psi - m\bar{\psi}\psi), \quad (\text{D.14})$$

where, according to (D.5),

$$\mathcal{D}_\alpha \psi = e_\alpha^\mu \left(\partial_\mu + \frac{1}{4} \Gamma_\mu^{\beta\gamma} \gamma_\beta \gamma_\gamma \right) \psi. \quad (\text{D.15})$$

The variation of this action with respect to $\bar{\psi}$ gives the Dirac equation in curved space-time,

$$i\gamma^\alpha \mathcal{D}_\alpha \psi - m\psi = 0. \quad (\text{D.16})$$

It is worth noting that the Dirac matrices in four dimensions obey

$$\frac{1}{2} \gamma^\alpha [\gamma^\beta, \gamma^\gamma] = (\eta^{\alpha\beta} \gamma^\gamma - \eta^{\alpha\gamma} \gamma^\beta) + i\epsilon^{\alpha\beta\gamma\delta} \gamma_\delta \gamma^5,$$

where, as usual, $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. This relation can be used for obtaining the convenient form of the Dirac operator,

$$i\gamma^\alpha \mathcal{D}_\alpha = i \left[\gamma^\alpha e_\alpha^\mu \partial_\mu + \frac{1}{2} \gamma^\alpha (\nabla_\mu e_\alpha^\mu) \right] + \frac{1}{4} \gamma^\alpha \gamma^5 \epsilon_{\alpha\beta\gamma\delta} e^{\beta\mu} e^{\gamma\nu} \nabla_\mu e_\nu^\delta. \quad (\text{D.17})$$

This form is ready for concrete calculations.

Let us find the form of the Dirac equation in the spatially flat cosmological metric $ds^2 = a^2(\eta) \eta_{\mu\nu} dx^\mu dx^\nu$. The natural choice of vierbein is

$$e_\mu^\alpha = a(\eta) \delta_\mu^\alpha,$$

where the factor a ensures the validity of the normalization condition (D.9). The non-vanishing Christoffel symbols are $\Gamma_{0\nu}^\mu = (a'/a)\delta_\nu^\mu$, $\Gamma_{ij}^0 = (a'/a)\delta_{ij}$. The non-vanishing covariant derivatives of the vierbein vectors are found by direct calculation,

$$\nabla_i e_\alpha^0 = \frac{a'}{a^2} \delta_{i\alpha}, \quad \nabla_i e_\alpha^j = \frac{a'}{a^2} \delta_i^j \delta_\alpha^0.$$

The term with γ^5 in (D.17) equals zero, while other terms give

$$i\gamma^\alpha \mathcal{D}_\alpha = i \left(\frac{1}{a} \gamma^\mu \partial_\mu + \frac{3}{2} \frac{a'}{a^2} \gamma^0 \right).$$

Hence, the Dirac equation in the spatially flat Universe reads

$$i \left(\frac{1}{a} \gamma^\mu \partial_\mu + \frac{3a'}{2a^2} \gamma^0 \right) \psi - m\psi = 0.$$

Finally, the change of variables $\psi = a^{-3/2}\chi$ simplifies this equation,

$$i\gamma^\mu \partial_\mu \chi - am\chi = 0.$$

For $m = 0$, this equation coincides with the massless Dirac equation in the Minkowski space-time. The latter property reflects the fact that the massless Dirac field is conformal. Namely, let $g_{\mu\nu}$ and

$$\hat{g}_{\mu\nu}(x) = e^{2\varphi(x)} g_{\mu\nu}(x)$$

be two conformally related metrics, see Section I.A.5. Then the following identity holds,

$$i\gamma^\alpha \hat{\mathcal{D}}_\alpha \hat{\psi} = e^{-5\varphi/2} \cdot i\gamma^\alpha \mathcal{D}_\alpha \psi, \quad (\text{D.18})$$

where $\hat{\mathcal{D}}_\alpha$ and \mathcal{D}_α are the derivatives (D.15) for the metrics $\hat{g}_{\mu\nu}$ and $g_{\mu\nu}$, and the fermion fields are related by

$$\hat{\psi} = e^{-3\varphi/2} \psi.$$

The relationship between the action functionals is particularly simple,

$$S_F(\hat{g}_{\mu\nu}, \hat{\psi}) = S_F(g_{\mu\nu}, \psi).$$

This means that the massless fermion field is conformal indeed.

Problem D.3. *Prove the relation (D.18) by making use of (D.17) and the formulas of Section I.A.5. Hint: Notice that according to (D.9), the vierbeins are related by*

$$\hat{e}_\mu^\alpha = e^\varphi e_\mu^\alpha, \quad \hat{e}^{\alpha\mu} = e^{-\varphi} e^{\alpha\mu}.$$

This page is intentionally left blank

Appendix E

Particle Creation in Background Fields. Method of Bogoliubov Transformations

The quantum process of particle creation in time-dependent external fields is of importance in various circumstances. It is often possible to neglect the interactions between the created particles, and treat the external field as classical. Then the equation for the quantum field describing created particles is linear. Our purpose in this Appendix is to describe the method that reduces this quantum problem to the problem of solving linear classical field equations. This method is based on the Bogoliubov transformations known from statistical mechanics.

Although the method is applicable to broad class of background fields, we assume that the background is homogeneous in space and invariant under spatial rotations. This situation is of major interest for cosmological applications. We assume in this Appendix that the space-time is flat; the generalization to the expanding Universe is given in appropriate places of this book.

E.1 Bosons

We begin with the creation of bosons. In the case of spatially homogeneous and isotropic background, bosons of arbitrary spin behave as scalars after the helicity decomposition, so it suffices to study the creation of scalar particles. Furthermore, it is sufficient to consider Hermitean scalar field $\hat{\varphi}$. We temporarily assume that the space is a cube of large but finite size L , and the quantum field $\hat{\varphi}(\mathbf{x}, t)$ obeys periodic boundary conditions. Then the Fourier transformation of this field reads

$$\hat{\varphi}(\mathbf{x}, t) = \sum_{\mathbf{q}} \frac{1}{L^{3/2}} e^{-i\mathbf{qx}} \hat{\varphi}_{\mathbf{q}}(t), \quad (\text{E.1})$$

where \mathbf{q} takes discrete values,

$$\mathbf{q} = \left(\frac{2\pi n_1}{L}, \frac{2\pi n_2}{L}, \frac{2\pi n_3}{L} \right). \quad (\text{E.2})$$

Here n_1, n_2, n_3 are integers, and the summation in (E.1) is performed over them. The normalization factor in (E.1) is chosen in such a way that the functions $L^{-3/2} e^{i\mathbf{qx}}$ are orthonormalized.

Let the Heisenberg field equation for the Fourier transform $\hat{\varphi}_{\mathbf{q}}(t)$, in the time-dependent background, read

$$\ddot{\hat{\varphi}}_{\mathbf{q}} + \omega_q^2(t)\hat{\varphi}_{\mathbf{q}} = 0, \quad (\text{E.3})$$

where $\omega_q^2(t)$ is a real function which is positive¹ as $t \rightarrow \pm\infty$; here $q \equiv |\mathbf{q}|$. Due to the assumed spatial isotropy of the background, the frequency ω depends on the absolute value of momentum q ; this property is in fact unimportant for what follows. Equation (E.3) is fairly generic. As an example, consider massive field interacting with an external scalar field $\phi(t)$. Let the Lagrangian be

$$L_{\varphi} = \frac{1}{2}(\partial_{\mu}\dot{\varphi})^2 + \frac{m_{\varphi}^2}{2}\dot{\varphi}^2 + \frac{1}{2}h\phi(t)\dot{\varphi}^2. \quad (\text{E.4})$$

Hence, the field equation has the form (E.3) with

$$\omega_q^2(t) = q^2 + m_{\varphi}^2 + h\phi(t).$$

More complicated cases, including the case of the expanding Universe, also reduce to Eq. (E.3) after the change of variables. The Hamiltonian corresponding to Eq. (E.3) is

$$\hat{H} = \sum_{\mathbf{q}} \left(\frac{1}{2}\dot{\hat{\varphi}}_{\mathbf{q}}\dot{\hat{\varphi}}_{-\mathbf{q}} + \frac{\omega_q^2(t)}{2}\hat{\varphi}_{\mathbf{q}}\hat{\varphi}_{-\mathbf{q}} \right). \quad (\text{E.5})$$

Problem E.1. Show that the Hamiltonian derived from the Lagrangian (E.4) has the form (E.5).

The general solution to Eq. (E.3) is written as follows,

$$\hat{\varphi}_{\mathbf{q}}(t) = \varphi_q^+(t)A_{\mathbf{q}}^\dagger + \varphi_q^-(t)A_{-\mathbf{q}} \quad (\text{E.6})$$

where the operators $A_{\mathbf{q}}$ and $A_{\mathbf{q}}^\dagger$ are Hermitean conjugate, the complex functions $\varphi_q^\pm(t)$ are complex conjugate to each other,

$$\varphi_q^+(t) = (\varphi_q^-(t))^*, \quad (\text{E.7})$$

and obey the equation

$$\ddot{\varphi}_q^\pm + \omega_q^2(t)\varphi_q^\pm = 0. \quad (\text{E.8})$$

Note that the notation in (E.6) is consistent with the standard expansion of the quantum field in the Minkowski space-time without external fields (see, e.g., the book [32]),

$$\hat{\varphi}(\mathbf{x}, t) = L^{-3/2} \sum_{\mathbf{q}} \frac{1}{\sqrt{2\omega_q}} (e^{i\omega_q t - i\mathbf{qx}} A_{\mathbf{q}}^\dagger + e^{-i\omega_q t + i\mathbf{qx}} A_{\mathbf{q}}).$$

¹The case $\omega^2 < 0$ corresponds to the tachyonic instability; hence, we assume that the tachyonic instability, if any, occurs in a finite time interval.

The condition (E.7) ensures that the field $\hat{\varphi}(\mathbf{x}, t)$ is Hermitean. We often omit the subscripts \mathbf{q} and q in what follows, keeping in mind that the formulas involve $A_{\mathbf{q}}^\dagger$ and $A_{-\mathbf{q}}$, if the opposite is not stated explicitly.

Any two solutions φ_1, φ_2 to Eq. (E.8) have the time-independent Wronskian

$$W(\varphi_1, \varphi_2) = -i(\dot{\varphi}_1 \varphi_2 - \varphi_1 \dot{\varphi}_2). \quad (\text{E.9})$$

The functions φ^\pm can be chosen in such a way that

$$W(\varphi^+, \varphi^-) = 1. \quad (\text{E.10})$$

With this choice, the operators $A_{\mathbf{q}}^\dagger$ and $A_{\mathbf{q}}$ obey the standard commutational relations of the creation and annihilation operators,

$$[A_{\mathbf{q}}, A_{\mathbf{q}'}^\dagger] = \delta_{\mathbf{q}, \mathbf{q}'}, \quad (\text{E.11})$$

where the symbol δ is understood in the sense of the equality between the integers entering (E.2). These commutational relations follow from the standard equal-time commutational relation between the field operator and its conjugate momentum,

$$[\dot{\varphi}(\mathbf{x}, t), \hat{\varphi}(\mathbf{y}, t)] = -i\delta(\mathbf{x} - \mathbf{y}).$$

Problem E.2. Prove the last statement above.

The important issue for the problem we consider is the choice of the state of the quantum field. We work in the Heisenberg picture, so the question is about the initial state of the system. In most applications, particles are created from *vacuum*, so the initial state should not contain particles. Such a state cannot be defined for arbitrary background, however: particles, generally speaking, are created (and annihilate) in the arbitrarily distant past. The vacuum of the field $\hat{\varphi}$ is well-defined at early times only if particle creation is absent in the asymptotic past. This is the case, e.g., if the background field is switched off as $t \rightarrow -\infty$. This is also the case in more general situation, when the external field varies with time adiabatically slowly in the distant past. It is the latter case that we are going to study.

If the background, and hence the frequency $\omega(t)$ evolve adiabatically slowly in the asymptotic past, then Eq. (E.8) can be solved in the WKB approximation at that time. The two solutions can be chosen in such a way that they approach positive- and negative-frequency functions as $t \rightarrow -\infty$:

$$\varphi^\pm(t) = \frac{1}{\sqrt{2\omega(t)}} e^{\pm i \int \omega(t) dt}, \quad t \rightarrow -\infty. \quad (\text{E.12})$$

In the limit $t \rightarrow -\infty$, and hence at any time, they obey (E.7) and (E.10). Furthermore, with this choice of φ^\pm , the expression for the Hamiltonian (E.5) in the asymptotic past is

$$\hat{H} = \sum_{\mathbf{q}} \omega_q \left(A_{\mathbf{q}}^\dagger A_{\mathbf{q}} + \frac{1}{2} \right), \quad t \rightarrow -\infty.$$

Hence, the state $|0_{in}\rangle$, defined by

$$A_{\mathbf{q}}|0_{in}\rangle = 0,$$

has the smallest energy in the distant past, and serves as the initial vacuum state. This state does not evolve in the Heisenberg picture.

Once one knows the classical solutions to the field equations that asymptote (E.12), one can calculate the correlation functions $\langle 0_{in}|\hat{\varphi}(\mathbf{x}, t)\hat{\varphi}(\mathbf{x}', t')|0_{in}\rangle$, etc., at arbitrary time. This is done by using (E.1) and (E.6) directly. In particular, one can calculate some physical quantities like the average value of the energy-momentum tensor and the field fluctuation. The latter is given by (cf. (13.6), (13.26))

$$\langle 0_{in}|\hat{\varphi}^2(\mathbf{x}, t)|0_{in}\rangle = L^{-3} \sum_{\mathbf{q}} \varphi_q^+ \varphi_q^-.$$

In the large volume limit, summation is replaced with integration,

$$\sum_{\mathbf{k}} \rightarrow \int d^3n = \int \frac{L^3}{(2\pi)^3} d^3q, \quad (\text{E.13})$$

and one obtains

$$\langle 0_{in}|\hat{\varphi}^2(\mathbf{x}, t)|0_{in}\rangle = \int \frac{d^3q}{(2\pi)^3} |\varphi_q^+|^2. \quad (\text{E.14})$$

We use the latter formula in what follows.

We are primarily interested here in the number of created particles. It is well-defined if the background field adiabatically slowly varies with time in the distant future as well, so that particle creation terminates as $t \rightarrow +\infty$. In that case one defines two other solutions $\tilde{\varphi}^\pm$ to Eq. (E.8), such that they approach positive- and negative-frequency functions in the future,

$$\tilde{\varphi}^\pm(t) = \frac{1}{\sqrt{2\omega(t)}} e^{\pm i \int \omega(t) dt}, \quad t \rightarrow +\infty. \quad (\text{E.15})$$

The field $\hat{\varphi}$ can be expanded in this set of solutions as well,

$$\hat{\varphi}_{\mathbf{q}}(t) = \tilde{\varphi}_q^+(t) \tilde{A}_{\mathbf{q}}^\dagger + \tilde{\varphi}_q^-(t) \tilde{A}_{-\mathbf{q}}, \quad (\text{E.16})$$

where the operators $\tilde{A}_{\mathbf{q}}$ and $\tilde{A}_{\mathbf{q}}^\dagger$ also obey the standard commutational relations

$$[\tilde{A}_{\mathbf{q}}, \tilde{A}_{\mathbf{q}'}^\dagger] = \delta_{\mathbf{q}, \mathbf{q}'}. \quad (\text{E.17})$$

The vacuum $|0_{out}\rangle$, such that $\tilde{A}_{\mathbf{q}}|0_{out}\rangle = 0$, is the lowest energy state in the distant future, and the operators \tilde{A}^\dagger , \tilde{A} create and annihilate particles at that time.

The state vector of the system is $|0_{in}\rangle$ at all times. Generically, it does not coincide with $|0_{out}\rangle$, i.e., the system contains particles as $t \rightarrow +\infty$. This is precisely the effect of particle creation in time-dependent background. The particle number operator in the mode \mathbf{q} has the following form in the asymptotic future,

$$\hat{N}_{\mathbf{q}} = \tilde{A}_{\mathbf{q}}^\dagger \tilde{A}_{\mathbf{q}},$$

hence the average number of particles of momentum \mathbf{q} is

$$N_{\mathbf{q}} = \langle 0_{in} | \tilde{A}_{\mathbf{q}}^\dagger \tilde{A}_{\mathbf{q}} | 0_{in} \rangle. \quad (\text{E.18})$$

Our next purpose is to find this number, assuming that the classical solutions φ_q^\pm are known.

The expansions (E.6) and (E.16) are the two expressions for one and the same operator $\hat{\varphi}$. Hence, the pairs of operators A , A^\dagger and \tilde{A} , \tilde{A}^\dagger are linearly related,²

$$\tilde{A}_{\mathbf{q}}^\dagger = \alpha A_{\mathbf{q}}^\dagger + \beta A_{-\mathbf{q}}, \quad (\text{E.19})$$

$$\tilde{A}_{\mathbf{q}} = \alpha^* A_{\mathbf{q}} + \beta^* A_{-\mathbf{q}}^\dagger. \quad (\text{E.20})$$

The complex parameters α and β depend on the absolute value of momentum.

The linear relations between the creation and annihilation operators (E.19), (E.20) are called the *Bogoliubov transformations*, and the coefficients α and β are the *Bogoliubov coefficients*. The consistency of the computational relations (E.11) and (E.17) requires that the Bogoliubov coefficients obey

$$|\alpha|^2 - |\beta|^2 = 1. \quad (\text{E.21})$$

Inserting (E.19), (E.20) into (E.18), we find the average number of created particles with momentum \mathbf{q} ,

$$N_{\mathbf{q}} = \langle 0_{in} | (\alpha_q A_{\mathbf{q}}^\dagger + \beta_q A_{-\mathbf{q}})(\alpha_q^* A_{\mathbf{q}} + \beta_q^* A_{-\mathbf{q}}^\dagger) | 0_{in} \rangle = |\beta_q|^2.$$

Hence, the calculation of the number of created particles reduces to the calculation of the Bogoliubov coefficient β_q . Note that the relation (E.21) does not impose any *a priori* restrictions on $|\beta_q|^2$, so the number of created particles can be arbitrarily large. We will see in Section E.2 that this is not the case for fermions: the number of fermions created in a given mode never exceeds 1. This property reflects the Pauli principle.

The number of created particles in the interval of momenta from \mathbf{q} to $\mathbf{q} + \Delta\mathbf{q}$ equals $\Delta N = \sum_{\mathbf{n}} N_{\mathbf{q}}$, where the summation runs over the values of (n_1, n_2, n_3) that correspond to momenta in this interval. In the large volume limit, we replace the sum by the integral with the use of Eq. (E.13) and obtain for the spatial number density of created particles in the element of momentum space $d^3 q$:

$$n(q)d^3 q = \frac{\Delta N}{L^3} = |\beta_q|^2 \frac{d^3 q}{(2\pi)^3}. \quad (\text{E.22})$$

To find the expressions for the Bogoliubov coefficients, we make use of the equality (E.10) written for the solutions $\tilde{\varphi}^\pm$. The expansion (E.16) implies

$$\tilde{A}^\dagger = W(\hat{\varphi}, \tilde{\varphi}^-).$$

²Note that Eqs. (E.19), (E.20) involve operators with one and the same momentum \mathbf{q} . This is the peculiarity of the spatially homogeneous background. In inhomogeneous backgrounds, the right hand sides of (E.19) and (E.20) involve the sums over momenta of the type $\sum_{\mathbf{q}'} \alpha_{\mathbf{q}, \mathbf{q}'} A_{\mathbf{q}'}^\dagger$; formulas get more cumbersome, but the formalism remains the same in principle.

We insert here $\hat{\varphi}$ in the form (E.6) and obtain that \tilde{A}^\dagger is indeed expressed through A^\dagger and A in accordance with (E.19), where the Bogoliubov coefficients are given by

$$\alpha = W(\varphi^+, \tilde{\varphi}^-), \quad (\text{E.23})$$

$$\beta = W(\varphi^-, \tilde{\varphi}^+). \quad (\text{E.24})$$

Since the functions φ^\pm are complex conjugate to each other, and the same holds for $\tilde{\varphi}^\pm$, this means that the solution φ^+ to Eq. (E.8), which is positive-frequency as $t \rightarrow -\infty$, is decomposed in terms of the solutions of definite frequencies in the distant future as follows,

$$\varphi^+ = \alpha \tilde{\varphi}^+ + \beta^* \tilde{\varphi}^-, \quad (\text{E.25})$$

where α and β are the Bogoliubov coefficients. This gives the prescription for the calculation of the latter: one should find the solution to the field equation which is positive-frequency in the distant past, evaluate the asymptotics of this solution in the distant future and expand this asymptotics in the positive- and negative-frequency solutions defined for $t \rightarrow +\infty$. The coefficients of this expansion are precisely the Bogoliubov coefficients. Equivalently, as we see from (E.23), (E.24), one can solve the field equation backwards in time with the “initial” condition that the solution is negative-frequency as $t \rightarrow +\infty$, and expand this solution in positive- and negative-frequency solutions defined in the distant past. In either case, particle creation in this language corresponds to mixing between positive and negative frequencies.

The formula (E.22) for the number of created particles expressed through the Bogoliubov coefficient, and the method of calculating β_q are central to this Section. It is also useful for some applications to have the expression for the field fluctuation $\hat{\varphi}(\mathbf{x}, t)$ as $t \rightarrow +\infty$, relative to its vacuum value (i.e., the value in the vacuum $|0_{out}\rangle$). We find from (E.14)

$$\langle \hat{\varphi}^2(\mathbf{x}, t) \rangle \equiv \langle 0_{in} | \hat{\varphi}^2(\mathbf{x}, t) | 0_{in} \rangle - \langle 0_{out} | \hat{\varphi}^2(\mathbf{x}, t) | 0_{out} \rangle = \int \frac{d^3 q}{(2\pi)^3} \left(|\varphi_q^+|^2 - \frac{1}{2\omega} \right).$$

We insert here the expression (E.25), omit rapidly oscillating terms and obtain

$$\langle \hat{\varphi}^2(\mathbf{x}, t) \rangle = \int \frac{d^3 q}{(2\pi)^3 \cdot 2\omega} (|\alpha_q|^2 + |\beta_q|^2 - 1) = \int \frac{d^3 q}{(2\pi)^3 \omega} |\beta_q|^2.$$

Hence, the field fluctuation is also expressed through the Bogoliubov coefficient β_q . Typically, the order-of-magnitude estimate for the fluctuation is

$$\langle \hat{\varphi}^2(\mathbf{x}, t) \rangle \sim \frac{n}{\omega}, \quad (\text{E.26})$$

where n is the total number density of created particles and ω is their typical energy.

Problem E.3. Find the relationship between the number density of created particles, energy density and power spectrum of the corresponding field in the end of the particle creation process.

Problem E.4. Generalize the analysis of this Section to spatially inhomogeneous background, assuming that the background field is switched off as $t \rightarrow \pm\infty$.

Let us show explicitly that the Bogoliubov coefficients (E.23), (E.24) obey (E.21). We write

$$\beta^* = -W(\varphi^+, \tilde{\varphi}^+).$$

Making use of the explicit form of the Wronskian (E.9), we find

$$|\alpha|^2 - |\beta|^2 = (\dot{\varphi}^+ \tilde{\varphi}^- - \varphi^+ \dot{\tilde{\varphi}}^-)(\varphi^- \tilde{\varphi}^+ - \varphi^- \dot{\tilde{\varphi}}^+) - (\varphi^+ \dot{\tilde{\varphi}}^+ - \varphi^+ \dot{\tilde{\varphi}}^+)(\dot{\varphi}^- \tilde{\varphi}^- - \varphi^- \dot{\tilde{\varphi}}^-),$$

where we have used (E.7). A straightforward algebraic manipulation then gives

$$|\alpha|^2 - |\beta|^2 = \dot{\varphi}^+ \varphi^- \cdot [-iW(\tilde{\varphi}^+, \tilde{\varphi}^-)] + \dot{\varphi}^- \varphi^+ \cdot [iW(\tilde{\varphi}^+, \tilde{\varphi}^-)] = W(\varphi^+, \varphi^-) = 1,$$

which is the desired result.

To end this Section, we make the following remark. The unambiguous formulation of the problem of particle creation in background field is possible only if the notion of particle is well-defined in the beginning and at the end of the process. This is the case we have considered. At the time when the evolution of the background is far from adiabatic, the notion of particle is not well-defined. At formal level this is due to the fact that there is no unique definition of positive- and negative-frequency functions entering (E.6). Still, one can sometimes get an idea of the state of the quantum field by employing the basis of the instantaneous Hamiltonian. To construct this basis, we write at given time t

$$\begin{aligned}\hat{\varphi}_{\mathbf{q}} &= \frac{1}{\sqrt{2\omega_q(t)}}(\mathcal{A}_{\mathbf{q}}^\dagger + \mathcal{A}_{-\mathbf{q}}), \\ \dot{\hat{\varphi}}_{\mathbf{q}} &= i\sqrt{\frac{\omega_q(t)}{2}}(\mathcal{A}_{\mathbf{q}}^\dagger - \mathcal{A}_{-\mathbf{q}}).\end{aligned}$$

Here the operators $\mathcal{A}^\dagger, \mathcal{A}$ depend on time, and obey the standard commutational relations for the creation and annihilation operators. The instantaneous Hamiltonian (E.5) has the following form in terms of these operators,

$$\hat{H} = \sum_{\mathbf{q}} \omega_{\mathbf{q}} \left(\mathcal{A}_{\mathbf{q}}^\dagger \mathcal{A}_{\mathbf{q}} + \frac{1}{2} \right).$$

Thus, one tends to interpret the operators $\mathcal{A}^\dagger, \mathcal{A}$ as the creation and annihilation operators of particles at time t . In a non-adiabatic situation this interpretation is rather vague, and the outlined approach should be used with caution.

Problem E.5. Find the dependence of the operators $\mathcal{A}^\dagger, \mathcal{A}$ on time for adiabatically evolving background.

Problem E.6. Assuming that the background evolves adiabatically as $t \rightarrow -\infty$ and the initial state is vacuum, find the coefficients of the Bogoliubov transformation from the operators A^\dagger, A to operators $\mathcal{A}^\dagger, \mathcal{A}$ at arbitrary moment of time in terms of the classical solutions $\varphi^\pm(t)$; the latter are assumed to be known at all times.

E.2 Fermions

We now consider creation of fermions in time-dependent background fields, which are again assumed to be spatially homogeneous and isotropic. We consider for

definiteness the theory in flat space-time, and assume that the quantum fermionic field $\hat{\psi}(\mathbf{x}, t)$ obeys the non-stationary Dirac equation

$$i\gamma^\mu \partial_\mu \hat{\psi} - m(t) \hat{\psi} = 0$$

with real and positive $m(t)$. The effect of non-stationary background is reflected in the time-dependence of the parameter $m(t)$. As an example, the Yukawa interaction with the background scalar field $\phi(t)$ yields $m(t) = m_0 + h\phi(t)$, where m_0 is the fermion mass and h is the Yukawa coupling.

The Dirac equation can be written in the form reminiscent of the Schrödinger equation,

$$i \frac{\partial}{\partial t} \hat{\psi} = [-i\alpha^i \partial_i + \beta \cdot m(t)] \hat{\psi},$$

where $\alpha^i = \gamma^0 \gamma^i$, $\beta = \gamma^0$. We perform the Fourier transformation analogous to (E.1) and obtain in the momentum representation

$$i \frac{\partial}{\partial t} \hat{\psi}_{\mathbf{q}} = [-\alpha^i q_i + \beta \cdot m(t)] \hat{\psi}_{\mathbf{q}}. \quad (\text{E.27})$$

The matrices α^i and β are Hermitean, so the scalar product of any two complex solutions $\psi_{\mathbf{q}}, \psi'_{\mathbf{q}}$ to Eq. (E.27) is conserved,

$$\frac{\partial}{\partial t} (\psi_{\mathbf{q}}^\dagger \psi'_{\mathbf{q}}) = 0.$$

The expansion analogous to (E.6) is³

$$\hat{\psi}_{\mathbf{q}} = u_{\mathbf{q}}^{(s)}(t) A_{\mathbf{q}}^{(s)\dagger} + v_{-\mathbf{q}}^{(s)}(t) B_{-\mathbf{q}}^{(s)}. \quad (\text{E.28})$$

Here the index $s = 1, 2$ labels the spin states, summation over this index is implied. We often omit this index in what follows. The four-component columns $u_{\mathbf{q}}^{(s)}, v_{-\mathbf{q}}^{(s)}$ make the orthonormalized set at each moment of time, and obey the Dirac equation (E.27). The operators $A_{\mathbf{q}}$ and $B_{\mathbf{q}}$ obey the standard anticommutational relations for the fermion creation and annihilation operators:

$$\{A_{\mathbf{q}}^{(s)}, A_{\mathbf{q}'}^{(s')\dagger}\} = \delta_{\mathbf{q}, \mathbf{q}'} \delta_{s, s'} \quad (\text{E.29})$$

$$\{B_{\mathbf{q}}^{(s)}, B_{\mathbf{q}'}^{(s')\dagger}\} = \delta_{\mathbf{q}, \mathbf{q}'} \delta_{s, s'}, \quad (\text{E.30})$$

other anticommutators vanish. These anticommutational relations ensure the standard equal-time anticommutational relations for the Dirac quantum field (see, e.g., the book [32]),

$$\begin{aligned} \{\hat{\psi}(\mathbf{x}, t), \hat{\psi}^\dagger(\mathbf{x}', t)\} &= \delta(\mathbf{x} - \mathbf{x}'), \\ \{\hat{\psi}(\mathbf{x}, t), \hat{\psi}(\mathbf{x}', t)\} &= 0. \end{aligned}$$

³To simplify notations, we do not use the notation involving superscripts \pm for the c -number solutions to the Dirac equation.

Problem E.7. Prove the last statement. Hint: Make use of the fact that the columns $u_{\mathbf{q}}^{(s)}, v_{-\mathbf{q}}^{(s)}$ make complete orthonormal set.

Due to C -invariance of the Dirac equation (E.27), the solutions $v_{-\mathbf{q}}$ can be chosen as follows,

$$v_{\mathbf{q}}(t) = U_C u_{\mathbf{q}}^*(t), \quad (\text{E.31})$$

where U_C is the time-independent C -conjugation matrix (see Appendix I.B) whose form is irrelevant for us, and $u_{\mathbf{q}}^*$ is the complex conjugate spinor.

Consider again the situation where the background field evolves in time adiabatically slowly in the distant past and distant future. Then the solutions to the Dirac equation $u(t)$ and $v(t)$ can be chosen in such a way that they are positive- and negative-frequency as $t \rightarrow -\infty$, respectively (the subscript \mathbf{q} is omitted wherever possible),

$$\begin{aligned} u(t) &= C_u(t) e^{i \int \omega dt}, \\ v(t) &= U_C u^*(t) = C_v(t) e^{-i \int \omega dt}, \end{aligned}$$

where C_u and C_v are slowly varying amplitudes (four-columns). Then the operators A^\dagger and A are the creation and annihilation operators of fermions in the distant past, while B^\dagger and B are the creation and annihilation operators of antifermions. We assume again that there are no particles and antiparticles in the distant past, so the initial state is vacuum $|0_{in}\rangle$ with respect to these operators (recall that we work in the Heisenberg picture, so the state vector is independent of time),

$$A|0_{in}\rangle = B|0_{in}\rangle = 0.$$

In the asymptotic future, $t \rightarrow +\infty$, the expansion analogous to (E.28) is

$$\hat{\psi}_{\mathbf{q}} = \tilde{u}_{\mathbf{q}}^{(s)}(t) \tilde{A}_{\mathbf{q}}^{(s)\dagger} + \tilde{v}_{-\mathbf{q}}^{(s)}(t) \tilde{B}_{-\mathbf{q}}^{(s)}, \quad (\text{E.32})$$

where the functions \tilde{u} and \tilde{v} are positive- and negative-frequency, respectively, as $t \rightarrow +\infty$, and the operators \tilde{A}^\dagger , \tilde{A} , \tilde{B}^\dagger and \tilde{B} create and annihilate fermions and antifermions in the distant future. The two sets of operators are again linearly related, and this relation is also called the Bogoliubov transformation,

$$\tilde{A}_{\mathbf{q}}^\dagger = \alpha_q A_{\mathbf{q}}^\dagger + \beta_q B_{-\mathbf{q}}, \quad (\text{E.33})$$

$$\tilde{B}_{\mathbf{q}}^\dagger = \alpha_q B_{\mathbf{q}}^\dagger + \beta_q A_{-\mathbf{q}}. \quad (\text{E.34})$$

We have anticipated that (E.33) and (E.34) involve the same Bogoliubov coefficients; we will soon see that this is a consequence of C -invariance. Like in the bosonic theory, the relations (E.33) and (E.34) yield for the average numbers of fermions and antifermions of momentum \mathbf{q} :

$$N_{\mathbf{q}} = \langle 0_{in} | \tilde{A}_{\mathbf{q}}^\dagger A_{\mathbf{q}} | 0_{in} \rangle = |\beta_q|^2, \quad (\text{E.35})$$

$$\bar{N}_{\mathbf{q}} = \langle 0_{in} | \tilde{B}_{\mathbf{q}}^\dagger B_{\mathbf{q}} | 0_{in} \rangle = |\beta_q|^2. \quad (\text{E.36})$$

The numbers of created particles and antiparticles are equal to each other for a given momentum \mathbf{q} , again due to C -invariance. Note that the result (E.35), (E.36) is valid for each spin state separately; this is a consequence of the homogeneity and isotropy of the background. To find the total number of created particles one sums over the spin states (in our case multiplies the result by the spin factor $g = 2$). In the infinite volume limit, the spatial number density in the element $d^3 q$ of momentum space is given by (cf. (E.22))

$$n(q)d^3 q = \bar{n}(q)d^3 q = g|\beta_q|^2 \frac{d^3 q}{(2\pi)^3}.$$

The operators referring to both distant past and distant future obey the anticommutational relations (E.29), (E.30). This gives the relation between the Bogoliubov coefficients (cf. (E.21)),

$$|\alpha|^2 + |\beta|^2 = 1.$$

We see that the number of created particles in each mode obeys the constraint

$$N_{\mathbf{q}} \leq 1.$$

This is the manifestation of the Pauli principle.

To find the Bogoliubov coefficients, we follow the same logic that has lead to (E.23), (E.24). Namely, since the columns $\tilde{u}_{\mathbf{q}}^{(s)}$, $\tilde{v}_{-\mathbf{q}}^{(s)}$ form the orthonormalized set, we find from the expansion (E.32) that

$$\begin{aligned}\tilde{A}_{\mathbf{q}}^\dagger &= \tilde{u}_{\mathbf{q}}^\dagger \hat{\psi} = (\tilde{u}_{\mathbf{q}}^\dagger u_{\mathbf{q}}) \cdot A_{\mathbf{q}}^\dagger + (\tilde{u}_{\mathbf{q}}^\dagger v_{-\mathbf{q}}) \cdot B_{-\mathbf{q}}, \\ \tilde{B}_{\mathbf{q}} &= \tilde{v}_{\mathbf{q}}^\dagger \hat{\psi} = (\tilde{v}_{\mathbf{q}}^\dagger u_{-\mathbf{q}}) \cdot A_{-\mathbf{q}} + (\tilde{v}_{\mathbf{q}}^\dagger v_{\mathbf{q}}) \cdot B_{\mathbf{q}}.\end{aligned}$$

The first of these relations is used to obtain the Bogoliubov coefficients entering (E.33),

$$\alpha_q = \tilde{u}_{\mathbf{q}}^\dagger u_{\mathbf{q}}, \tag{E.37}$$

$$\beta_q = \tilde{u}_{\mathbf{q}}^\dagger v_{-\mathbf{q}}. \tag{E.38}$$

The second relation gives the Bogoliubov coefficients in (E.34),

$$\alpha_q = (\tilde{v}_{\mathbf{q}}^\dagger v_{\mathbf{q}})^*, \tag{E.39}$$

$$\beta_q = (\tilde{v}_{\mathbf{q}}^\dagger u_{-\mathbf{q}})^*. \tag{E.40}$$

The two expressions for α_q are in fact equal. This can be seen by using the relation (E.31) for both u , v , and \tilde{u} , \tilde{v} . The two expressions for β_q are also equal in view of (E.31).

It follows from (E.37), (E.40) that the expansion of the function $u(t)$ in positive- and negative-frequency functions defined with respect to the distant future is

$$u(t) = \alpha \tilde{u}(t) + \beta^* \tilde{v}(t).$$

This formula is completely analogous to (E.25) and yields the same prescription for the calculation of the Bogoliubov coefficients as in the bosonic case.

To end this Section, we note that we have considered for simplicity the C -invariant theory. This is not always the case. Our analysis can be generalized to the theory in which C and/or CP are violated. The formulas are more cumbersome, and the number of created particles may be different from that of antiparticles. The basic mechanism is, however, the same.

Problem E.8. *Construct a model of the Yukawa type that describes C -violating interaction of fermions with non-stationary background scalar fields. Generalize the results of this Section. The same for the case of CP -violation. Hint: Do not forget that the Lagrangian must be Hermitean.*

This page is intentionally left blank

Appendix F

Some Special Functions and Their Properties

We present in this Appendix some special functions and their properties used in the main text. Our presentation here is by no means complete or mathematically rigorous, it is rather a summary of the facts about special functions, sometimes with comments.

F.1 Spherical Bessel Functions $j_l(x)$ of Integer Order

The spherical Bessel function of order $l = 0, 1, 2, \dots$ is defined as the non-singular at the origin solution¹ to the equation

$$\frac{d^2 j_l}{dx^2} + \frac{2}{x} \frac{d j_l}{dx} + \left(1 - \frac{(l+1)l}{x^2}\right) j_l = 0. \quad (\text{F.1})$$

There is a simple relation between the spherical and conventional Bessel functions,

$$j_l(x) = \sqrt{\frac{\pi}{2x}} J_{l+\frac{1}{2}}(x). \quad (\text{F.2})$$

The relative normalization of the solutions to Eq. (F.1) is such that the following recurrence relations hold,

$$j_{l+1} + j_{l-1} = \frac{2l+1}{x} j_l, \quad l = 1, 2, \dots, \quad (\text{F.3})$$

$$lj_{l-1} - (l+1) j_{l+1} = (2l+1) \frac{d j_l}{dx}, \quad l = 0, 1, 2, \dots. \quad (\text{F.4})$$

Note that the recurrence relations (F.3), (F.4) are also valid for the spherical Bessel functions of non-integer order, which are defined as the solutions to Eq. (F.1) with arbitrary l . The spherical Bessel functions of non-integer order are irrelevant in the context of this book, so we do not consider them here.

¹There is also the solution $n_l(x)$ to Eq. (F.1) which is singular at the origin. This solution is called the spherical Bessel function of the second kind, while j_l is the spherical Bessel function of the first kind. The functions n_l are not relevant for us. When talking about the spherical Bessel functions throughout this book, we always mean the functions $j_l(x)$.

Problem F.1. Find the recurrence relations for the coefficients in the Taylor series for $j_l(x)$. Show that these relations are consistent with (F.3), (F.4).

Problem F.2. The recurrence relations (F.3), (F.4) enable one to express j_{l+1} through j_l and dj_l/dx . Show that j_{l+1} obtained in this way solves Eq. (F.1).

It follows directly from (F.1) that the spherical Bessel function of the zeroth order is

$$j_0(x) = \frac{\sin x}{x}. \quad (\text{F.5})$$

This formula, combined with the recurrence relations (F.3), (F.4), enables one to obtain all spherical Bessel functions explicitly. Quite conveniently, they can be written in terms of the Rayleigh representation,

$$j_l(x) = (-1)^l x^l \left(\frac{1}{x} \frac{d}{dx} \right)^l j_0(x). \quad (\text{F.6})$$

Problem F.3. Prove the Rayleigh representation by induction.

Problem F.4. Show that the Rayleigh representation is equivalent to the following representation,

$$j_l(x) = \frac{x^l}{l!2^l} \left(1 + \frac{d^2}{dx^2} \right)^l j_0(x).$$

The representation (F.6) shows that every spherical Bessel function is a series in the trigonometric functions $\sin x$, $\cos x$ with coefficients, polynomial in $1/x$. It gives for the first three spherical Bessel functions,

$$j_0(x) = \frac{\sin x}{x}, \quad j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x}, \quad j_2(x) = \left(\frac{3}{x^3} - \frac{1}{x} \right) \sin x - \frac{3}{x^2} \cos x.$$

The representation (F.6) can also be used to find the behavior of the spherical Bessel functions at small argument. To this end, one writes the expression (F.5) as a series in x ,

$$j_0(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n+1)!}.$$

Since

$$\frac{1}{x} \frac{d}{dx} \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n+1)!} = \sum_{n=1}^{\infty} \frac{2n(-1)^n x^{2n-2}}{(2n+1)!}, \quad (\text{F.7})$$

the spherical Bessel function has the following series representation:

$$\begin{aligned} j_l(x) &= (-1)^l x^l \left(\frac{1}{x} \frac{d}{dx} \right)^{l-1} \cdot \frac{1}{x} \frac{d}{dx} \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n+1)!} \\ &= x^l \sum_{n=l}^{\infty} \frac{2n(2n-2)\cdots(2n-2(l-1))}{(2n+1)!} (-1)^{n+l} \cdot x^{2(n-l)}. \end{aligned}$$

Hence, the behavior as $x \rightarrow 0$ is

$$j_l(x) = \frac{2^l l!}{(2l+1)!} x^l \cdot [1 + \mathcal{O}(x^2)] = \frac{x^l}{(2l+1)!!} \cdot [1 + \mathcal{O}(x^2)]. \quad (\text{F.8})$$

We see that all spherical Bessel functions, except for the zeroth order function (F.5), rapidly tend to zero as $x \rightarrow 0$, while $j_0(0) = 1$.

We are particularly interested in the spherical Bessel functions of high order, $l \gg 1$. They are exponentially small at $x < (l + 1/2)$,

$$j_l(x) \simeq \frac{\exp \left\{ -\left(l + \frac{1}{2}\right) \left[\operatorname{Arch} \left(\frac{l+\frac{1}{2}}{x} \right) - \sqrt{1 - \frac{x^2}{(l+\frac{1}{2})^2}} \right] \right\}}{2\sqrt{x} \left(\left(l - \frac{1}{2}\right)^2 - x^2 \right)^{1/4}}. \quad (\text{F.9})$$

The way to obtain this asymptotics is briefly discussed below.

We now turn to the behavior of the spherical Bessel functions at large argument, $x \gg l$. We see from (F.6) that the term with all derivatives acting on sine dominates there. Since $-d\sin x/dx = \sin(x - \pi/2)$, we conclude that the asymptotic behavior at $x \gg l$ is

$$j_l(x) \approx \frac{\sin(x - \frac{\pi l}{2})}{x}. \quad (\text{F.10})$$

The asymptotics for the amplitude of oscillations $\propto 1/x$ is achieved fairly quickly, see below, while the phases of the exact function and the asymptotics (F.10) coincide at very large argument only, $x \gg l$.

The spherical Bessel function of large enough order grows with x , reaches its maximum, and then oscillates, see Fig. F.1. The position of the first maximum $x = x_*$ can be found for $l \gg 1$ by considering the region $x \sim l \gg 1$. We introduce the new argument $\xi = x/l$ and write Eq. (F.1) in terms of the function $f(\xi) = j_l(l\xi)$,

$$\frac{1}{l^2} \frac{d^2 f}{d\xi^2} + \frac{2}{l^2} \frac{df}{\xi d\xi} + \left(1 - \frac{1 + \frac{1}{l}}{\xi^2}\right) f = 0. \quad (\text{F.11})$$

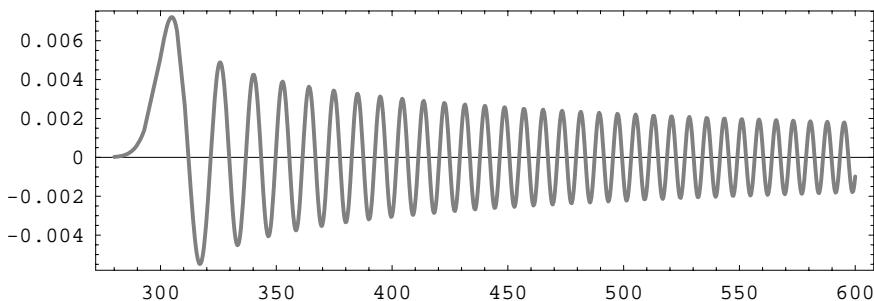


Fig. F.1 The spherical Bessel function $j_l(x)$ of order $l = 300$. The amplitude of oscillations behaves as $\propto 1/x$ at $x > l + 1/2$.

The position of the maximum $\xi = \xi_*$ corresponds to $d\mathcal{f}/d\xi = 0$, so we see from (F.11) that if the second derivative is not very large at the maximum, the expression in parenthesis must vanish. This gives

$$x_{*,l} = \left(l + \frac{1}{2}\right) \left[1 + \mathcal{O}\left(\left(l + \frac{1}{2}\right)^{-2/3}\right)\right]; \quad (\text{F.12})$$

corrections to this formula can be found in the book [13]. The height of the first maximum is given by [13]

$$j_l(x_{*,l}) = 0.85 \left(l + \frac{1}{2}\right)^{-5/6} \cdot \left\{1 + \mathcal{O}\left[\left(l + \frac{1}{2}\right)^{-2/3}\right] + \dots\right\}. \quad (\text{F.13})$$

We repeatedly use in this book the asymptotics of the spherical Bessel function in the regime where both the argument and order are large, $x, l \rightarrow \infty, x/(l + 1/2) = \text{const} > 1$. In this regime, the spherical Bessel function oscillates with sizeable amplitude. To obtain the behavior in this regime, we use the well known result for the similar asymptotics of the usual Bessel function $J_\nu(x)$ (see [13]):

$$\begin{aligned} J_\nu(x) &= \sqrt{\frac{2}{\nu\pi\tan\beta}} \left\{ \cos\left(\nu\tan\beta - \nu\beta - \frac{\pi}{4}\right) \cdot \left[1 + \mathcal{O}\left(\frac{1}{\nu^2}\right)\right] \right. \\ &\quad \left. + \frac{1}{\nu} \cdot \frac{1 + \frac{5}{3\tan^2\beta}}{8\tan\beta} \sin\left(\nu\tan\beta - \nu\beta - \frac{\pi}{4}\right) \cdot \left[1 + \mathcal{O}\left(\frac{1}{\nu^2}\right)\right] \right\}, \end{aligned} \quad (\text{F.14})$$

where $\cos\beta \equiv \nu/x$. The parameter $\tan\beta$ is not very small in the regime we discuss, so the second term in (F.14) can be neglected. Then we obtain from (F.14), (F.2)

$$\begin{aligned} j_l(x) &= \frac{1}{\sqrt{x}} \frac{1}{\left[x^2 - \left(l + \frac{1}{2}\right)^2\right]^{1/4}} \\ &\quad \times \cos\left[\sqrt{x^2 - \left(l + \frac{1}{2}\right)^2} - \left(l + \frac{1}{2}\right) \arccos\left(\frac{l + \frac{1}{2}}{x}\right) - \frac{\pi}{4}\right]. \end{aligned} \quad (\text{F.15})$$

As can be seen from Fig. F.2, this is a very good approximation at $l \gtrsim 5$. The formula (F.15) shows that the asymptotics (F.10) is restored at $x \gg l$. In the region of the first peaks $x = (l + 1/2)(1 + \epsilon)$, $\epsilon \ll 1$, we find from (F.15) to the leading order in ϵ ,

$$j_l[(l + 1/2)(1 + \epsilon)] = \frac{1}{l + \frac{1}{2}} \frac{1}{(2\epsilon)^{1/4}} \cos\left[\frac{2\sqrt{2}}{3}\epsilon\sqrt{\epsilon} \cdot \left(l + \frac{1}{2}\right) - \frac{\pi}{4}\right]. \quad (\text{F.16})$$

Hence, the widths of the first peaks are of order $\Delta\epsilon \sim l^{-2/3}$, and

$$\Delta x \sim l^{1/3}. \quad (\text{F.17})$$

Note that the asymptotic formula (F.15) can be analytically continued to the region $x \ll l$. Without going into details, we state the result that the oscillating cosine in

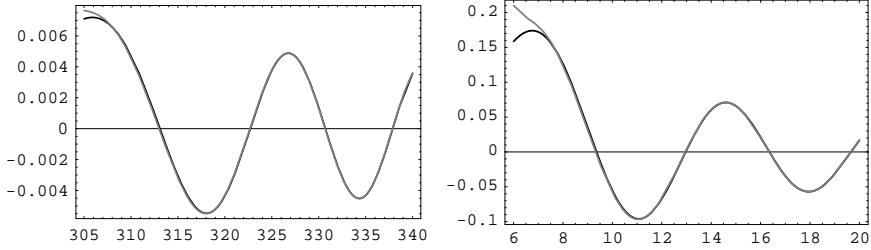


Fig. F.2 Exact spherical Bessel functions (black lines) of the order $l = 300$ (left) and $l = 5$ (right) and approximations (F.15) (gray lines). The curves coincide almost everywhere, showing that the approximation (F.15) is very accurate.

(F.15) becomes the exponential function which *decays* towards small x . This yields the behavior (F.9).

In this book we encounter convolutions of the spherical Bessel functions of high orders with slowly varying functions,

$$\int f(k) j_l(k\eta) dk, \quad l \gg 1. \quad (\text{F.18})$$

Since $j_l(k\eta)$ rapidly oscillates at $k\eta \gg l$ and is exponentially small at $k\eta \ll l$, the integral is saturated at $k\eta \sim l$, i.e., in the region of the first peaks. Let $f(k)$ vary slowly in that region, more precisely, let $df/dk \cdot (\Delta(k\eta)/\eta) \ll f$, where $\Delta(k\eta)$ is estimated by (F.17). In other words, let $df/dk \ll \eta f/l^{1/3}$ at $k \sim l/\eta$. Then $f(k)$ is approximately constant in the relevant region, and we write

$$\int f(k) j_l(k\eta) dk \simeq f\left(\frac{l + \frac{1}{2}}{\eta}\right) \cdot \int dk j_l(k\eta).$$

Since

$$\int dk j_l(k\eta) = \frac{\sqrt{\pi}}{2\eta} \frac{\Gamma\left(\frac{l+1}{2}\right)}{\Gamma\left(\frac{l+2}{2}\right)} = \sqrt{\frac{\pi}{2l}} \frac{1}{\eta},$$

where the last equality is valid for $l \gg 1$, we find the approximate relation

$$\int f(k) j_l(k\eta) dk \simeq f\left(\frac{l + \frac{1}{2}}{\eta}\right) \sqrt{\frac{\pi}{2l}} \frac{1}{\eta}. \quad (\text{F.19})$$

Let us now consider the integral of slowly varying function with two spherical Bessel functions of the same order $l \gg 1$,

$$\int dk \cdot k^2 f(k) j_l(k\eta) j_l(k\eta').$$

Because of rapid oscillations of the spherical Bessel functions at $k\eta > l$ and their exponential decay at $k\eta < l$, this integral is substantially different from zero only if

η' is close to η . Therefore, we can write

$$\int dk \cdot k^2 f(k) j_l(k\eta) j_l(k\eta') \propto \delta(\eta' - \eta), \quad (\text{F.20})$$

where the coefficient of proportionality has yet to be found. To this end, we integrate the left hand side of Eq. (F.20) with a slowly varying function $g(\eta')$. Making use of (F.19) to evaluate the integral over η' first, and then the integral over k , we find

$$\begin{aligned} & \int dk \cdot k^2 f(k) j_l(k\eta) \int d\eta' g(\eta') j_l(k\eta') \\ & \approx \int dk k^2 f(k) j_l(k\eta) g\left(\frac{l+1/2}{k}\right) \sqrt{\frac{\pi}{2l}} \frac{1}{k} \\ & \approx \frac{\pi}{2l\eta} k f(k) g\left(\frac{l+1/2}{k}\right) \Big|_{k=(l+1/2)/\eta} \\ & \approx \frac{\pi}{2\eta^2} f\left(\frac{l+1/2}{\eta}\right) g(\eta). \end{aligned}$$

This gives the final approximate formula

$$\int dk \cdot k^2 f(k) j_l(k\eta) j_l(k\eta') \approx \frac{\pi}{2\eta^2} f\left(\frac{l+\frac{1}{2}}{\eta}\right) \cdot \delta(\eta - \eta'). \quad (\text{F.21})$$

To end this Section, let us consider the first derivative of the spherical Bessel function $j'_l(x)$. We obtain from (F.8) that the behavior of the derivative near the origin is

$$j'_l(x) = \frac{lx^{l-1}}{(2l+1)!!}.$$

Thus, all derivatives, except for the derivative of the function of the first order, vanish at $x = 0$. For large l , the derivative is exponentially small at $x < (l+1/2)$, similarly to the spherical Bessel function itself, see (F.9). The asymptotics at $l \gg 1$, $x \gg 1$, $x/(l+1/2) = \text{const} > 1$ is obtained by differentiating (F.15). The leading asymptotics comes about from the differentiation of the cosine, and we get

$$\begin{aligned} j'_l(x) &= -\frac{\left[x^2 - (l + \frac{1}{2})^2\right]^{1/4}}{x\sqrt{x}} \\ &\times \sin\left[\sqrt{x^2 - \left(l + \frac{1}{2}\right)^2} - \left(l + \frac{1}{2}\right) \arccos\left(\frac{l + \frac{1}{2}}{x}\right) - \frac{\pi}{4}\right]. \end{aligned} \quad (\text{F.22})$$

The derivative is suppressed at small $(x-l)$ as compared to the function itself. This has the effect that the first peak is suppressed, and the amplitude of oscillations remains nearly constant in a fairly large interval of the argument. This behavior is due to the competition between the growing factor $[x^2 - (l+1/2)^2]^{1/4}$ and decreasing factor $x^{-3/2}$. The result of this competition is shown in Fig. F.3.

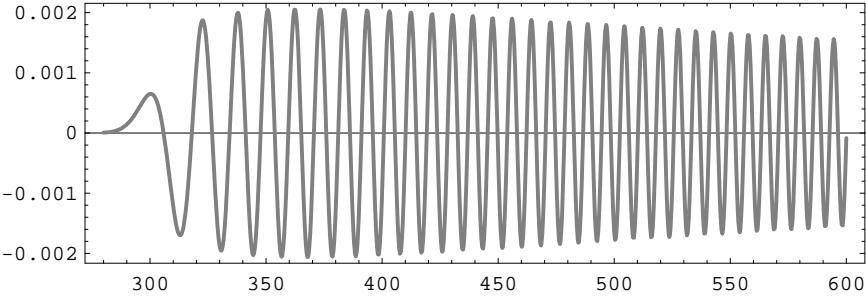


Fig. F.3 The derivative of the spherical Bessel function $j'_l(x)$ of order $l = 300$. The first peak is suppressed, and the amplitude is nearly constant at $x \simeq 300 - 450$.

F.2 Legendre Polynomials $P_n(x)$ and Spherical Harmonics $Y_{lm}(n)$

The Legendre polynomial of order n is defined as the finite in the interval $-1 \leq x \leq 1$ solution to the equation

$$(1 - x^2) \frac{d^2 P_n}{dx^2} - 2x \frac{dP_n}{dx} + n(n+1) P_n = 0, \quad (\text{F.23})$$

where n is a non-negative integer. The normalization is such that

$$P_n(1) = 1. \quad (\text{F.24})$$

Then the following recurrence relations hold,

$$(2n+1)xP_n - nP_{n-1} = (n+1)P_{n+1}, \quad (\text{F.25})$$

$$nP_n - nP_{n-1} = (x^2 - 1) \frac{dP_n}{dx}. \quad (\text{F.26})$$

Problem F.5. Find the recurrence relations for the coefficients in the Taylor series for $P_n(x)$. Show that these relations are consistent with (F.25), (F.26).

Problem F.6. The recurrence relations (F.25), (F.26) enable one to express P_{n+1} through P_n and dP_n/dx . Show that P_{n+1} obtained in this way solves Eq. (F.23).

It is clear from (F.23) and (F.24) that

$$P_0(x) = 1. \quad (\text{F.27})$$

Then the recurrence relation (F.25) shows that $P_n(x)$ are indeed polynomials. The first three Legendre polynomials are obtained by using (F.25),

$$P_1(x) = x, \quad P_2(x) = \frac{3x^2 - 1}{2}, \quad P_3(x) = \frac{5x^3 - 3x}{2}. \quad (\text{F.28})$$

The recurrence relations can be used to derive the Rodrigues representation,

$$P_n(x) = \frac{1}{n!2^n} \frac{d^n}{dx^n} (x^2 - 1)^n. \quad (\text{F.29})$$

Among other things, it implies the transformation property

$$P_n(-x) = (-1)^n \cdot P_n(x).$$

Problem F.7. Show directly that the function (F.29) is the solution to Eq. (F.23).

Problem F.8. Show by induction that the polynomials (F.29) obey the recurrence relations (F.25), (F.26).

It follows from (F.29) that the order of the polynomial $P_n(x)$ is n . The Rodrigues formula can also be used for obtaining the explicit form of the polynomials. To this end, we note that

$$\begin{aligned} \frac{d^2}{dx^2} (x^2 - 1)^n &= \frac{d^2}{dx^2} \sum_{m=0}^n (-1)^{n-m} \frac{n!x^{2m}}{m!(n-m)!} \\ &= \sum_{m=1}^n (-1)^{n-m} \frac{n!2m(2m-1)}{m!(n-m)!} x^{2m-2}. \end{aligned} \quad (\text{F.30})$$

Hence, the formula (F.29) gives

$$P_n(x) = \frac{1}{2^n} \sum_{m=[n/2]}^n (-1)^{n-m} \frac{2m(2m-1)\cdots(2m-n+1)}{m!(n-m)!} x^{2m-n},$$

where $[n/2]$ denotes the integer part of $n/2$. To obtain a more compact form, we write the sum in terms of $k = n - m$, and multiply the numerator and denominator in each term by $(n - 2k)!$. This gives finally

$$P_n(x) = \frac{1}{2^n} \sum_{k=0}^{[n/2]} \frac{(-1)^k (2n-2k)!x^{n-2k}}{k!(n-k)!(n-2k)!}. \quad (\text{F.31})$$

The formula (F.29) enables one to calculate integrals involving the Legendre polynomials. As an example, let us see that the polynomials are orthogonal to each other. To this end, we first calculate the integral of the polynomial of order n with weight x^m for $m < n$. Making use of (F.29), we write

$$\int_{-1}^1 x^m P_n(x) dx = \frac{1}{n!2^n} \int_{-1}^1 x^m \frac{d^n}{dx^n} (x^2 - 1)^n dx.$$

We integrate by parts m times, observe that the boundary terms vanish due to the factor $(x^2 - 1)$, and obtain

$$\int_{-1}^1 x^m P_n(x) dx = (-1)^m \frac{m!}{n!2^n} \int_{-1}^1 dx \frac{d^{n-m}}{dx^{n-m}} (x^2 - 1)^n = 0.$$

Hence, the polynomial P_n is orthogonal to all x^m with $m < n$. As we know, P_n is a polynomial in x of order n . We immediately see that the Legendre polynomials of

different orders are orthogonal,

$$\int_{-1}^1 P_n(x) P_m(x) dx = 0, \quad m \neq n.$$

Let us calculate the integral of the Legendre polynomial squared. We again use the Rodrigues formula (F.29) and integrate n times by parts,

$$\int_{-1}^1 P_n^2(x) dx = \frac{(-1)^n}{4^n (n!)^2} \int_{-1}^1 dx (x^2 - 1)^n \frac{d^{2n}}{dx^{2n}} (x^2 - 1)^n. \quad (\text{F.32})$$

Making use of (F.30) we obtain

$$\frac{d^{2n}}{dx^{2n}} (x^2 - 1)^n = (2n)!.$$

The remaining integral in (F.32) is

$$\int_{-1}^1 (x^2 - 1)^n dx = (-1)^n \frac{\sqrt{\pi} \Gamma(n+1)}{\Gamma(\frac{3}{2} + n)}.$$

Collecting all factors, we find

$$\int_{-1}^1 dx P_n^2(x) = \frac{2}{2n+1}. \quad (\text{F.33})$$

Hence the Legendre polynomials are orthogonal in the interval $-1 \leq x \leq 1$ and their norm is given by (F.33). With this normalization, they make a complete orthogonal set of functions in the interval $-1 \leq x \leq 1$,

$$\int_{-1}^1 dx P_n(x) P_m(x) = \frac{2\delta_{nm}}{2n+1}. \quad (\text{F.34})$$

We use in the main text the following series expansion of a plane wave in the Legendre polynomials,

$$e^{-ix \cos \theta} = \sum_{l=0}^{\infty} (2l+1) (-i)^l j_l(x) P_l(\cos \theta), \quad (\text{F.35})$$

where $j_l(x)$ is the spherical Bessel function of order l , see Section F.1. We introduce the variable $y = \cos \theta$ and use the orthogonality of the Legendre polynomials to rewrite the formula (F.35) in the equivalent form,

$$\int_{-1}^1 e^{-ixy} P_n(y) dy = 2(-i)^n j_n(x). \quad (\text{F.36})$$

This formula shows that modulo the numerical factor, the spherical Bessel function is the Fourier transform of the Legendre polynomial of the same order.

Problem F.9. Derive the formula (F.36) making use of the Rodrigues representation (F.29). Hint: Use the representation for the spherical Bessel functions obtained in problem F.4.

Let us prove the relation (F.36) by induction. For $n = 0$ we have $P_0(x) = 1$, $j_0(x) = \sin x/x$, so the equality (F.36) holds. For $n = 1$ we have $P_1(y) = y$, and the left hand side of (F.36) can be written as

$$\frac{1}{-i} \frac{d}{dx} \int_{-1}^1 e^{-ixy} P_0(y) dy = \frac{2}{-i} \frac{d}{dx} j_0(x).$$

Hence, for $n = 1$ the relation (F.36) is equivalent to

$$\frac{2}{-i} \frac{d}{dx} j_0(x) = -2ij_1(x),$$

which indeed holds in view of (F.4). So, the formula (F.36) is valid also for $n = 1$.

Let the formula (F.36) be valid for $n = l$. Let us show that it is then valid for $n = l + 1$. To this end, we express the polynomial $P_{l+1}(y)$ in the integrand of (F.36) in terms of the linear combination of polynomials of lower orders, using the recurrence relation (F.25),

$$\int_{-1}^1 e^{-ixy} P_{l+1}(y) dy = \frac{2l+1}{l+1} \int_{-1}^1 y e^{-ixy} P_l(y) dy - \frac{l}{l+1} \int_{-1}^1 e^{-ixy} P_{l-1}(y) dy. \quad (\text{F.37})$$

The second term here is known by the assumption of induction, while the first one is expressed through the derivative of $j_n(x)$,

$$\int_{-1}^1 y e^{-ixy} P_l(y) dy = \frac{1}{-i} \frac{d}{dx} \int_{-1}^1 e^{-ixy} P_l(y) dy = 2i(-i)^l \frac{d}{dx} j_l(x).$$

Collecting all terms, we obtain

$$\int_{-1}^1 e^{-ixy} P_{l+1}(y) dy = \frac{2}{l+1} (-i)^{l+1} \left(l j_{l-1}(x) - (2l+1) \frac{d}{dx} j_l(x) \right) = 2(-i)^{l+1} j_{l+1}(x),$$

where the last equality follows from the recurrence relation (F.4). Hence, the formula (F.36) is valid for $n = l + 1$, and by induction for all n .

At large n , the Legendre polynomials exhibit the asymptotics (see, e.g., [13])

$$P_n(\cos \theta) = \sqrt{\frac{2}{n\pi \sin \theta}} \cos \left[\left(n + \frac{1}{2} \right) \theta - \frac{\pi}{4} \right] + \mathcal{O}\left(\frac{1}{n^{3/2}}\right). \quad (\text{F.38})$$

It shows that the functions $P_n(\cos \theta)$ oscillate with θ at large n with the period $2\pi/n$.

We use in the main text the expansions of functions on a unit sphere in the spherical harmonics $Y_{lm}(\mathbf{n}) \equiv Y_{lm}(\theta, \phi)$, where ϕ and θ are the angles in the spherical coordinate frame. The spherical harmonics form the complete orthonormalized system of functions on the unit sphere (see, e.g., the book [61] for details),

$$\int d\mathbf{n} Y_{lm}^*(\mathbf{n}) Y_{l'm'}(\mathbf{n}) = \int \sin \theta \, d\theta \, d\phi \, Y_{lm}^*(\theta, \phi) Y_{l'm'}(\theta, \phi) = \delta_{ll'} \delta_{mm'}.$$

We recall that the spherical harmonics are expressed in terms of the Legendre polynomials. Namely,²

$$Y_{lm}(\theta, \phi) = (-1)^{\frac{m+|m|}{2}} \cdot \sqrt{\frac{2l+1}{4\pi} \frac{(l-|m|)!}{(l+|m|)!}} \sin^{|m|} \theta \cdot \frac{d^{|m|} P_l(\cos \theta)}{(d \cos \theta)^{|m|}} \cdot e^{im\phi}. \quad (\text{F.39})$$

The spherical harmonics are the solutions of the equation

$$\frac{\partial^2 Y_{lm}}{\partial \theta^2} + \frac{\cos \theta}{\sin \theta} \frac{\partial Y_{lm}}{\partial \theta} + l(l+1) Y_{lm} - \frac{m^2}{\sin^2 \theta} Y_{lm} = 0. \quad (\text{F.40})$$

Problem F.10. Show that the spherical harmonics defined according to (F.39) solve equation (F.40).

The asymptotic formula for the spherical harmonics, analogous to (F.38), reads

$$Y_{lm}(\theta, \phi) = \frac{1}{\pi \sqrt{\sin \theta}} \cos \left[\left(l + \frac{1}{2} \right) \theta - \frac{\pi}{4} + \frac{\pi m}{2} \right] \cdot e^{im\phi} + O\left(\frac{1}{l}\right). \quad (\text{F.41})$$

The latter formula is valid at $l \gg 1$, $l \gg m$. Note that the formulas (F.38) and (F.41) are *not* valid near the poles of the sphere, more precisely, at $\sin \theta \lesssim n^{-1}$ and $\sin \theta \lesssim l^{-1}$, respectively.

²It is worth noting that the formula (F.39) differs from the definition given in Ref. [61], where the analogous formula contains an extra factor i^l .

This page is intentionally left blank

Books and Reviews

We give here an (incomplete) list of books and reviews where various aspects of cosmological perturbations, inflation and related issues are discussed.

Books

- Ya. B. Zeldovich, I. D. Novikov, *The Structure and Evolution of the Universe (Relativistic Astrophysics, volume 2)*, University of Chicago Press, 1983.
- N. D. Birrell, P. C. W. Davies, *Quantum Fields In Curved Space*, Cambridge University Press, 1982.
- A. D. Linde, *Particle Physics and Inflationary Cosmology*, Harwood, Chur, 1990.
- E. W. Kolb, M. S. Turner, *The Early Universe*, Addison-Wesley, Redwood City, 1990 — Frontiers in physics, 69.
- P. J. E. Peebles, *Principles of Physical Cosmology*, Princeton University Press, 1993.
- J. A. Peacock, *Cosmological Physics*, Cambridge University Press, 1999.
- A. R. Liddle, D. H. Lyth, *Cosmological Inflation and Large Scale Structure*, Cambridge University Press, 2000.
- P. Coles and F. Lucchin, *Cosmology: The Origin and Evolution of Cosmic Structure*, Chichester, UK: Wiley, 2002.
- S. Dodelson, *Modern Cosmology*, Academic Press, Amsterdam, 2003.
- V. Mukhanov, *Physical Foundations of Cosmology*, Cambridge University Press, 2005.
- P. Naselsky, D. Novikov, I. Novikov, *The Physics of Cosmic Microwave Background*, Cambridge University Press, 2006.
- S. Weinberg, *Cosmology*, Oxford University Press, 2008.
- M. Giovannini, *A Primer on the Physics of the Cosmic Microwave Background*, World Scientific, Singapore, 2008.
- R. Durrer, *Cosmic Microwave Background*, Cambridge University Press, 2008.
- A. R. Liddle, D. H. Lyth, *The Primordial Density Perturbation: Cosmology, Inflation and the Origin of Structure*, Cambridge University Press, 2009.

Reviews

If necessary, relevant Sections are indicated in parenthesis.

- H. Kodama and M. Sasaki, Cosmological Perturbation Theory, *Prog. Theor. Phys. Suppl.* **78** (1984) 1.

- R. H. Brandenberger, Quantum Field Theory Methods And Inflationary Universe Models, *Rev. Mod. Phys.* **57** (1985) 1.
- V. F. Mukhanov, H. A. Feldman and R. H. Brandenberger, Theory of Cosmological Perturbations, *Phys. Rept.* **215** (1992) 203.
- V. N. Lukash and I. D. Novikov, Lectures on the Very Early Universe, in ‘*Observational and Physical Cosmology*’, eds. F. Sanchez *et al.*, Cambridge University Press (1992) 3.
- A. R. Liddle and D. H. Lyth, The Cold Dark Matter Density Perturbation, *Phys. Rept.* **231** (1993) 1 [arXiv:astro-ph/9303019].
- S. D. M. White, Formation and Evolution of Galaxies: Lectures given at Les Houches, August 1993, 349–430 in *Cosmology and Large scale structure* / Eds. Schaeffer R., Silk J., Spiro M., Zinn-Justin. J., N. Y., Elsevier, 1996. 968p.; arXiv:astro-ph/9410043 (Chapter 7).
- A. V. Gurevich and K. P. Zybin, Large-Scale Structure of the Universe. Analytic Theory. *Physics Uspekhi*, **38** (1995) 687 (Chapter 1).
- A. Kosowsky, Cosmic Microwave Background Polarization, *Annals Phys.* **246** (1996) 49 [arXiv:astro-ph/9501045] (Chapter 10).
- V. N. Lukash, Formation of the Large Scale Structure of the Universe, in ‘*Cosmology and Gravitation, II*’, ed. M. Novello, Editions Frontieres (1996) 288.
- V. N. Lukash, Physics of the Early Universe, in ‘*Cosmology: The Physics of the Universe*’, eds. B.A. Robson *et al.*, World Scientific (1996) 213.
- J. A. Peacock, Cosmology and Particle Physics, Proc. 1998 European School of High-Energy Physics, St. Andrews, Scotland, 23 Aug–5 Sep 1998.
- M. S. Turner and J. A. Tyson, Cosmology at the Millennium, *Rev. Mod. Phys.* **71** (1999) S145 [arXiv:astro-ph/9901113].
- A. Riotto and M. Trodden, Recent Progress in Baryogenesis, *Ann. Rev. Nucl. Part. Sci.* **49** (1999) 35 [arXiv:hep-ph/9901362] (Chapter 15).
- D. H. Lyth and A. Riotto, Particle Physics Models of Inflation and the Cosmological Density Perturbation, *Phys. Rept.* **314** (1999) 1 [arXiv:hep-ph/9807278].
- V. N. Lukash, The Very Early Universe, [arXiv:astro-ph/9910009].
- A. D. Linde, Inflationary Cosmology, *Phys. Rept.* **333** (2000) 575.
- E. Bertschinger, Cosmological Perturbation Theory and Structure Formation, 1.1 in *Proceedings of Cosmology 2000* / Eds. Bento M. C., Bertolami O., Teodoro L. [arXiv:astro-ph/0101009].
- A. D. Dolgov, Cosmological Implications of Neutrinos, *Surveys High Energy Phys.* **17** (2002) 91 [arXiv:hep-ph/0208222] (Chapter 8).
- A. D. Dolgov, Neutrinos in Cosmology, *Phys. Rept.* **370** (2002) 333 [arXiv:hep-ph/0202122] (Chapter 8).
- K. Enqvist and A. Mazumdar, Cosmological Consequences of MSSM Flat Directions, *Phys. Rept.* **380** (2003) 99 [arXiv:hep-ph/0209244] (Chapters 12, 13, 14).
- W. L. Freedman and M. S. Turner, Measuring and Understanding the Universe, *Rev. Mod. Phys.* **75** (2003) 1433 [arXiv:astro-ph/0308418].
- I. I. Tkachev, Astroparticle Physics, Proc. 2003 European School on High-Energy Physics, Tsakhkadzor, Armenia, 24 Aug–6 Sep 2003 [arXiv:hep-ph/0405168].
- M. Gasperini and G. Veneziano, The Pre-big Bang Scenario in String Cosmology, *Phys. Rept.* **373** (2003) 1 [arXiv:hep-th/0207130] (Chapter 16).
- N. Bartolo, E. Komatsu, S. Matarrese and A. Riotto, Non-Gaussianity from Inflation: Theory and Observations, *Phys. Rept.* **402** (2004) 103 [arXiv:astro-ph/0406398] (Chapters 9, 14).

- A. Challinor, Anisotropies in the Cosmic Microwave Background, *Lect. Notes Phys.* 2004. **653**. 71 [arXiv:astro-ph/0403344] (Chapters 9, 10).
- V. Rubakov, *Introduction to Cosmology*, PoS **RTN2005** (2005) 003.
- M. Giovannini, Theoretical Tools for the Physics of CMB Anisotropies, *Int. J. Mod. Phys. D* **14** (2005) 363 [arXiv:astro-ph/0412601] (Chapters 9, 10).
- B. A. Bassett, S. Tsujikawa and D. Wands, Inflation Dynamics and Reheating, *Rev. Mod. Phys.* **78** (2006) 537 [arXiv:astro-ph/0507632].
- A. Lewis and A. Challinor, Weak Gravitational Lensing of the CMB, *Phys. Rept.* **429** (2006) 1 [arXiv:astro-ph/0601594] (Chapters 9, 10).
- S. Hannestad, Primordial Neutrinos, *Ann. Rev. Nucl. Part. Sci.* **56** (2006) 137 [arXiv:hep-ph/0602058] (Chapter 8).
- J. Lesgourgues and S. Pastor, Massive Neutrinos and Cosmology, *Phys. Rept.* **429**, 307 (2006) [arXiv:astro-ph/0603494] (Chapter 8).
- R. Brandenberger, Topics in Cosmology, PoS **P2GC** (2006) 007 [arXiv:hep-th/0701157].
- N. Aghanim, S. Majumdar and J. Silk, Secondary Anisotropies of the CMB, *Rept. Prog. Phys.* **71** (2008) 066902 [arXiv:0711.0518 [astro-ph]] (Chapters 9, 10).
- C. G. Tsagas, A. Challinor and R. Maartens, Relativistic Cosmology and Large-Scale Structure, *Phys. Rept.* **465** (2008) 61 [arXiv:0705.4397 [astro-ph]].
- L. Kofman, Preheating after Inflation, *Lect. Notes Phys.* **738** (2008) 55 (Chapter 15).
- A. Linde, Inflationary Cosmology, *Lect. Notes Phys.* **738** (2008) 1 [arXiv:0705.0164 [hep-th]].
- J. L. Lehners, Ekpyrotic and Cyclic Cosmology, *Phys. Rept.* **465**, 223 (2008) [arXiv:0806.1245 [astro-ph]] (Chapter 16).
- M. Novello and S. E. P. Bergliaffa, Bouncing Cosmologies, *Phys. Rept.* **463**, 127 (2008) [arXiv:0802.1634 [astro-ph]] (Chapter 16).
- A. Mazumdar and J. Rocher, Particle Physics Models of Inflation and Curvaton Scenarios, arXiv:1001.0993 [hep-ph].
- R. H. Brandenberger, Cosmology of the Very Early Universe, arXiv:1003.1745 [hep-th] (Chapter 16).
- R. Allahverdi, R. Brandenberger, F. Y. Cyr-Racine and A. Mazumdar, Reheating in Inflationary Cosmology: Theory and Applications, arXiv:1001.2600 (Chapter 15).

This page is intentionally left blank

Bibliography

- [1] D. S. Gorbunov and V. A. Rubakov, *Introduction to the Theory of the Early Universe: Hot Big Bang Theory* — World Scientific, 2011, 473 p.
- [2] E. Komatsu *et al.* [WMAP Collaboration], *Astrophys. J. Suppl.* **180**, 330 (2009) [arXiv:0803.0547 [astro-ph]].
- [3] <http://pdg.lbl.gov>
- [4] M. Khlopov, B. A. Malomed and I. B. Zeldovich, *Mon. Not. Roy. Astron. Soc.* **215**, 575 (1985).
- [5] W. Hu, R. Barkana and A. Gruzinov, *Phys. Rev. Lett.* **85**, 1158 (2000) [arXiv:astro-ph/0003365], A. Arbey, J. Lesgourgues and P. Salati, *Phys. Rev. D* **65**, 083514 (2002) [arXiv:astro-ph/0112324].
- [6] F. Piazza and M. Pospelov, “Sub-eV scalar dark matter through the super-renormalizable Higgs portal,” arXiv:1003.2313.
- [7] M. Bucher, K. Moodley and N. Turok, *Phys. Rev. D* **62**, 083508 (2000) [arXiv:astro-ph/9904231].
- [8] N. Bartolo, E. Komatsu, S. Matarrese and A. Riotto, *Phys. Rept.* **402**, 103 (2004) [arXiv:astro-ph/0406398].
- [9] S. Weinberg, *Phys. Rev. D* **67**, 123504 (2003) [arXiv:astro-ph/0302326].
- [10] E. R. Harrison, *Phys. Rev. D* **1**, 2726 (1970).
- [11] Y. B. Zeldovich, *Mon. Not. Roy. Astron. Soc.* **160**, 1P (1972).
- [12] E. Komatsu and D. N. Spergel, *Phys. Rev. D* **63**, 063002 (2001) [arXiv:astro-ph/0005036].
- [13] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York: Dover Publications, eds. 1972.
- [14] V. J. Martinez, *Lect. Notes Phys.* **665**, 269 (2009) [arXiv:0804.1536 [astro-ph]].
- [15] M. Tegmark *et al.* [SDSS Collaboration], *Astrophys. J.* **606**, 702 (2004) [arXiv:astro-ph/0310725].
- [16] R. A. Sunyaev and Y. B. Zeldovich, *Astrophys. Space Sci.* **7**, 3 (1970).
- [17] P. J. E. Peebles and J. T. Yu, *Astrophys. J.* **162**, 815 (1970).
- [18] J. R. Bond and G. Efstathiou, *Astrophys. J.* **285**, L45 (1984).
- [19] J. A. Holzman, *Astrophys. J. Suppl.* **71**, 1 (1989).
- [20] A. D. Sakharov, *Zh. Eksp. Teor. Fiz.* **49**, 345 (1965) [Sov. Phys. JETP **22**, 241 (1966)].
- [21] D. J. Eisenstein *et al.* [SDSS Collaboration], *Astrophys. J.* **633**, 560 (2005) [arXiv:astro-ph/0501171].

- [22] S. Bashinsky and E. Bertschinger, *Phys. Rev. Lett.* **87**, 081301 (2001) [arXiv:astro-ph/0012153]; S. Bashinsky and E. Bertschinger, *Phys. Rev. D* **65**, 123008 (2002) [arXiv:astro-ph/0202215].
- [23] W. J. Percival, S. Cole, D. J. Eisenstein, R. C. Nichol, J. A. Peacock, A. C. Pope and A. S. Szalay, *Mon. Not. Roy. Astron. Soc.* **381**, 1053 (2007) [arXiv:0705.3323 [astro-ph]].
- [24] W. H. Press and P. Schechter, *Astrophys. J.* **187**, 425 (1974).
- [25] L. Liberato and R. Rosenfeld, *JCAP* **0607**, 009 (2006) [arXiv:astro-ph/0604071].
- [26] J. M. Bardeen, J. R. Bond, N. Kaiser and A. S. Szalay, *Astrophys. J.* **304**, 15 (1986).
- [27] J. R. Bond, S. Cole, G. Efstathiou and N. Kaiser, *Astrophys. J.* **379**, 440 (1991).
- [28] R. K. Sheth, H. J. Mo and G. Tormen, *Mon. Not. Roy. Astron. Soc.* **323**, 1 (2001) [arXiv:astro-ph/9907024].
- [29] R. K. Sheth and G. Tormen, *Mon. Not. Roy. Astron. Soc.* **329**, 61 (2002) [arXiv:astro-ph/0105113].
- [30] A. Vikhlinin *et al.*, *Astrophys. J.* **692**, 1060 (2009) [arXiv:0812.2720 [astro-ph]].
- [31] L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics* in 10 volumes; Volume IV, V. B. Berestetskii, E. M. Lifshitz, L. P. Pitaevskii *Quantum Electrodynamics* — Second Edition — Reed Educational and Professional Publishing Ltd, 1982.
- [32] N. N. Bogolyubov and D. V. Shirkov, *Introduction to the theory of quantized fields*, Third edition — Wiley, New York, 1980.
- [33] C. Itzykson and J.B. Zuber, *Quantum Field Theory*, — McGraw-Hill, New York, 1980.
- [34] M.E. Peskin and D.V. Schroeder, *An Introduction to Quantum Field Theory* — Addison-Wesley Publishing Company, 1995.
- [35] S. Weinberg, *The Quantum Theory of Fields*; in 3 volumes — Cambridge University Press, 1996.
- [36] G. S. Bisnovaty-Kogan and Ya. B. Zeldovich, *Astron. Zh.* **47**, 942 (1970) [Sov. Astron. **14**, 758 (1971)].
- [37] D. Gorbunov, A. Khmelnitsky and V. Rubakov, *JHEP* **0812**, 055 (2008) [arXiv:0805.2836 [hep-ph]].
- [38] A. A. Klypin, A. V. Kravtsov, O. Valenzuela and F. Prada, *Astrophys. J.* **522**, 82 (1999) [arXiv:astro-ph/9901240].
- [39] B. Moore, S. Ghigna, F. Governato, G. Lake, T. R. Quinn, J. Stadel and P. Tozzi, *Astrophys. J.* **524**, L19 (1999).
- [40] A. Tikhonov and A. Klypin, *Mon. Not. Roy. Astron. Soc.* **395**, 1915 (2009) [arXiv:0807.0924 [astro-ph]].
- [41] A. V. Tikhonov, S. Gottloeber, G. Yepes and Y. Hoffman, “The sizes of mini-voids in the local universe: an argument in favor of a warm dark matter model?”, arXiv:0904.0175 [astro-ph.CO].
- [42] M. Viel, M. G. Haehnelt and V. Springel, *Mon. Not. Roy. Astron. Soc.* **354**, 684 (2004) [arXiv:astro-ph/0404600].
- [43] J. D. Simon and M. Geha, *Astrophys. J.* **670**, 313 (2007) [arXiv:0706.0516 [astro-ph]].
- [44] S. Tremaine and J. E. Gunn, *Phys. Rev. Lett.* **42**, 407 (1979).
- [45] A. Boyarsky, O. Ruchayskiy and D. Iakubovskyi, *JCAP* **0903**, 005 (2009) [arXiv:0808.3902 [hep-ph]].
- [46] D. Gorbunov, A. Khmelnitsky and V. Rubakov, *JCAP* **0810**, 041 (2008) [arXiv:0808.3910 [hep-ph]].

- [47] U. Seljak, A. Makarov, P. McDonald and H. Trac, *Phys. Rev. Lett.* **97**, 191303 (2006) [arXiv:astro-ph/0602430].
- [48] A. Boyarsky, D. Iakubovskyi, O. Ruchayskiy and V. Savchenko, *Mon. Not. Roy. Astron. Soc.* **387**, 1361 (2008) [arXiv:0709.2301 [astro-ph]].
- [49] R. Trotta and A. Melchiorri, *Phys. Rev. Lett.* **95**, 011305 (2005) [arXiv:astro-ph/0412066].
- [50] J. Lesgourges and S. Pastor, *Phys. Rept.* **429**, 307 (2006) [arXiv:astro-ph/0603494].
- [51] V. Barger, D. Marfatia and A. Tregre, *Phys. Lett. B* **595**, 55 (2004) [arXiv:hep-ph/0312065].
- [52] S. A. Thomas, F. B. Abdalla and O. Lahav, “Upper Bound of 0.28eV on the Neutrino Masses from the Largest Photometric Redshift Survey,” arXiv:0911.5291.
- [53] M. Tegmark *et al.* [SDSS Collaboration], *Phys. Rev. D* **74**, 123507 (2006) [arXiv:astro-ph/0608632].
- [54] J. R. Bond, in *Cosmology and Large Scale Structure*, Les Houches Session LX, Eds. R. Schaeffer, J. Silk, M. Spiro and J. Zinn-Justin. N.Y., Elsevier, 1996.
- [55] S. Weinberg, *Phys. Rev. D* **69**, 023503 (2004) [arXiv:astro-ph/0306304].
- [56] S. Bashinsky, “Coupled evolution of primordial gravity waves and relic neutrinos,” arXiv:astro-ph/0505502.
- [57] Y. Zhang, W. Zhao, T. Xia and Y. Yuan, *Phys. Rev. D* **74**, 083006 (2006) [arXiv:astro-ph/0508345].
- [58] J. Silk, *Astrophys. J.* **151**, 459 (1968).
- [59] S. P. Goldman, *Phys. Rev. A* **40**, 1185 (1989).
- [60] P. Naselsky, D. Novikov and I. Novikov, *The Physics of Cosmic Microwave Background*, Cambridge University Press, 2006.
- [61] L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics* in 10 volumes; Volume III, L. D. Landau, E. M. Lifshitz, *Quantum mechanics (Nonrelativistic Theory)* — Third Edition — Reed Educational and Professional Publishing Ltd, 1977.
- [62] H. K. Eriksen, D. I. Novikov, P. B. Lilje, A. J. Banday and K. M. Gorski, *Astrophys. J.* **612**, 64 (2004) [arXiv:astro-ph/0401276].
- [63] C. Copi, D. Huterer, D. Schwarz and G. Starkman, *Phys. Rev. D* **75**, 023507 (2007) [arXiv:astro-ph/0605135].
- [64] C. L. Bennett *et al.*, “Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Are There Cosmic Microwave Background Anomalies?” arXiv:1001.4758.
- [65] C. L. Reichardt *et al.*, *Astrophys. J.* **694**, 1200 (2009) [arXiv:0801.1491 [astro-ph]].
- [66] R. K. Sachs and A. M. Wolfe, *Astrophys. J.* **147**, 73 (1967) [Gen. Rel. Grav. **39**, 1929 (2007)].
- [67] <http://map.gsfc.nasa.gov>
- [68] G. Hinshaw *et al.* [WMAP Collaboration], *Astrophys. J. Suppl.* **180**, 225 (2009) [arXiv:0803.0732 [astro-ph]].
- [69] <http://cmb.phys.cwru.edu/boomerang>
- [70] <http://cosmology.berkeley.edu/group/swlh/acbar/index.html>
- [71] A. Cabre, E. Gaztanaga, M. Manera, P. Fosalba and F. Castander, *Mon. Not. Roy. Astron. Soc. Lett.* **372**, L23 (2006) [arXiv:astro-ph/0603690].
- [72] T. Giannantonio *et al.*, *Phys. Rev. D* **74**, 063520 (2006) [arXiv:astro-ph/0607572].
- [73] W. Hu, *Astrophys. J.* **506**, 485 (1998) [arXiv:astro-ph/9801234].
- [74] W. Hu, D. J. Eisenstein, M. Tegmark and M. J. White, *Phys. Rev. D* **59**, 023512 (1999) [arXiv:astro-ph/9806362].

- [75] B. Li, J. D. Barrow, D. F. Mota and H. Zhao, *Phys. Rev. D* **78**, 064021 (2008) [arXiv:0805.4400 [gr-qc]].
- [76] R. E. Lopez, S. Dodelson, R. J. Scherrer and M. S. Turner, *Phys. Rev. Lett.* **81**, 3075 (1998) [arXiv:astro-ph/9806116].
- [77] A. Challinor, “Anisotropies in the cosmic microwave background,” arXiv:astro-ph/0403344.
- [78] I. A. Strukov, A. A. Bryukhanov, D. P. Skulachev and M. V. Sazhin, *Pis'ma Astron. Zh.* **18**, 387 (1992) [Sov. Astron. Lett. **18**, 153 (1992)].
- [79] G. F. Smoot *et al.*, *Astrophys. J.* **396**, L1 (1992); C. L. Bennett *et al.*, *Astrophys. J.* **464**, L1 (1996) [arXiv:astro-ph/9601067].
- [80] N. Vittorio and J. Silk, *Astrophys. J.* **285**, L39 (1984).
- [81] E. Calabrese, A. Slosar, A. Melchiorri, G. F. Smoot and O. Zahn, *Phys. Rev. D* **77**, 123531 (2008) [arXiv:0803.2309 [astro-ph]].
- [82] R. A. Sunyaev and Ya. B. Zeldovich, *Astrophys. Space Sci.* **7**, 3 (1970); R. A. Sunyaev and Ya. B. Zeldovich, *Comm. Astrophys. Space Phys.* **4**, 173 (1972).
- [83] <http://lambda.gsfc.nasa.gov/toolbox>
- [84] M. J. Rees, *Astrophys. J.* **153**, L1 (1968); M. M. Basko and A. G. Polnarev, *Mon. Not. Roy. Astron. Soc.* **191**, 207 (1980); J. Negroponte and J. Silk, *Phys. Rev. Lett.* **44**, 1433 (1980); N. Kaiser, *Mon. Not. Roy. Astron. Soc.* **202**, 1169 (1983).
- [85] A. G. Polnarev, *Astron. Zh.* **62**, 1041 (1985) [Sov. Astron. **29**, 607 (1985)].
- [86] M. V. Sazhin and N. Benites, *Astro. Lett. Commun.* **32** 105 (1995).
- [87] R. G. Crittenden, D. Coulson and N. G. Turok, *Phys. Rev. D* **52**, 5402 (1995).
- [88] M. Kamionkowski, A. Kosowsky and A. Stebbins, *Phys. Rev. Lett.* **78**, 2058 (1997) [arXiv:astro-ph/9609132].
- [89] U. Seljak and M. Zaldarriaga, *Phys. Rev. Lett.* **78**, 2054 (1997) [arXiv:astro-ph/9609169].
- [90] R. F. Stark, *Mon. Not. Roy. Astron. Soc.* **195**, 127S (1981).
- [91] M. R. Nolta *et al.* [WMAP Collaboration], *Astrophys. J. Suppl.* **180**, 296 (2009) [arXiv:0803.0593 [astro-ph]].
- [92] M. Zaldarriaga and U. Seljak, *Phys. Rev. D* **55**, 1830 (1997) [arXiv:astro-ph/9609170].
- [93] M. Kamionkowski, A. Kosowsky and A. Stebbins, *Phys. Rev. D* **55**, 7368 (1997) [arXiv:astro-ph/9611125].
- [94] A. Challinor, “Constraining fundamental physics with the cosmic microwave background,” arXiv:astro-ph/0606548.
- [95] J. Kovac, E. M. Leitch, C. Pryke, J. E. Carlstrom, N. W. Halverson and W. L. Holzapfel, *Nature* **420**, 772 (2002) [arXiv:astro-ph/0209478].
- [96] M. L. Brown *et al.* [QUaD collaboration], *Astrophys. J.* **705**, 978 (2009) [arXiv:0906.1003 [astro-ph.CO]].
- [97] A. Challinor, *Lect. Notes Phys.* **653**, 71 (2004) [arXiv:astro-ph/0502093].
- [98] M. Kaplinghat, L. Knox and Y. S. Song, *Phys. Rev. Lett.* **91**, 241301 (2003) [arXiv:astro-ph/0303344].
- [99] U. Seljak and A. Slosar, *Phys. Rev. D* **74**, 063523 (2006) [arXiv:astro-ph/0604143].
- [100] A. A. Starobinsky, *JETP Lett.* **30**, 682 (1979) [*Pisma Zh. Eksp. Teor. Fiz.* **30**, 719 (1979)].
- [101] A. A. Starobinsky, *Phys. Lett. B* **91**, 99 (1980).
- [102] A. H. Guth, *Phys. Rev. D* **23**, 347 (1981).
- [103] C. Armendariz-Picon, T. Damour and V. F. Mukhanov, *Phys. Lett. B* **458**, 209 (1999) [arXiv:hep-th/9904075].
- [104] A. D. Linde, *Phys. Lett. B* **129**, 177 (1983).
- [105] D. H. Lyth and A. Riotto, *Phys. Rept.* **314**, 1 (1999) [arXiv:hep-ph/9807278].

- [106] S. Kachru, R. Kallosh, A. Linde and S. P. Trivedi, *Phys. Rev. D* **68**, 046005 (2003) [arXiv:hep-th/0301240].
- [107] S. Kachru, R. Kallosh, A. Linde, J. M. Maldacena, L. McAllister and S. P. Trivedi, *JCAP* **0310**, 013 (2003) [arXiv:hep-th/0308055].
- [108] A. D. Linde, *Phys. Lett. B* **108**, 389 (1982).
- [109] A. Albrecht and P. J. Steinhardt, *Phys. Rev. Lett.* **48**, 1220 (1982).
- [110] A. D. Linde, *Phys. Rev. D* **49**, 748 (1994) [arXiv:astro-ph/9307002].
- [111] V. F. Mukhanov and G. V. Chibisov, *JETP Lett.* **33**, 532 (1981) [*Pisma Zh. Eksp. Teor. Fiz.* **33**, 549 (1981)].
- [112] S. W. Hawking, *Phys. Lett. B* **115**, 295 (1982).
- [113] A. A. Starobinsky, *Phys. Lett. B* **117**, 175 (1982).
- [114] A. H. Guth and S. Y. Pi, *Phys. Rev. Lett.* **49**, 1110 (1982).
- [115] J. M. Bardeen, P. J. Steinhardt and M. S. Turner, *Phys. Rev. D* **28**, 679 (1983).
- [116] V. A. Rubakov, M. V. Sazhin and A. V. Veryaskin, *Phys. Lett. B* **115**, 189 (1982).
- [117] R. Fabbri and M. D. Pollock, *Phys. Lett. B* **125**, 445 (1983).
- [118] L. F. Abbott and M. B. Wise, *Nucl. Phys. B* **244**, 541 (1984).
- [119] A. A. Starobinsky, *Sov. Astron. Lett.* **11**, 133 (1985).
- [120] V. N. Lukash, *JETP Lett.* **31**, 596 (1980) [*Pisma Zh. Eksp. Teor. Fiz.* **31**, 631 (1980)].
- [121] V. N. Lukash, *Sov. Phys. JETP* **52**, 807 (1980) [*Zh. Eksp. Teor. Fiz.* **79**, 1601 (1980)].
- [122] L. P. Grishchuk, *Sov. Phys. JETP* **40**, 409 (1975) [*Zh. Eksp. Teor. Fiz.* **67**, 825 (1974)].
- [123] Mukhanov V., *Physical Foundations of Cosmology*. Cambridge, UK: University Press, 2005, 280 p.
- [124] F. L. Bezrukov and M. Shaposhnikov, *Phys. Lett. B* **659**, 703 (2008) [arXiv:0710.3755 [hep-th]].
- [125] T. L. Smith, M. Kamionkowski and A. Cooray, *Phys. Rev. D* **73**, 023504 (2006) [arXiv:astro-ph/0506422].
- [126] J. E. Lidsey, A. R. Liddle, E. W. Kolb, E. J. Copeland, T. Barreiro and M. Abney, *Rev. Mod. Phys.* **69**, 373 (1997) [arXiv:astro-ph/9508078].
- [127] D. S. Salopek and J. R. Bond, *Phys. Rev. D* **42**, 3936 (1990).
- [128] T. Falk, R. Rangarajan and M. Srednicki, *Astrophys. J.* **403**, L1 (1993) [arXiv:astro-ph/9208001].
- [129] J. M. Maldacena, *JHEP* **0305**, 013 (2003) [arXiv:astro-ph/0210603].
- [130] R. H. Brandenberger, “Inflationary cosmology: Progress and problems,” arXiv:hep-ph/9910410.
- [131] A. D. Linde, *Phys. Lett. B* **175**, 395 (1986).
- [132] A. S. Goncharov, A. D. Linde and V. F. Mukhanov, *Int. J. Mod. Phys. A* **2**, 561 (1987).
- [133] J. D. Barrow and F. J. Tipler, *The Anthropic Cosmological Principle*. Oxford, UK: Oxford University Press, 1986, 706 p.
- [134] S. Weinberg, *Rev. Mod. Phys.* **61**, 1 (1989).
- [135] A. Linde, “Inflation, quantum cosmology and the anthropic principle,” 426–458 in *Science and Ultimate Reality: From Quantum to Cosmos*. (eds.) Barrow J. D., Davies P. C. W. and Harper C. L., Cambridge University Press, 2003, [arXiv:hep-th/0211048].
- [136] J. Polchinski, “The cosmological constant and the string landscape,” 216–236 in *Proc. 23rd Solvay Conference in Physics: The Quantum Structure of Space and Time*. (eds.) Gross D., Henneaux M. and Sevrin A., World Scientific, 2007, 272 p [arXiv:hep-th/0603249].
- [137] S. Mollerach, *Phys. Rev. D* **42**, 313 (1990).

- [138] A. D. Linde and V. F. Mukhanov, *Phys. Rev. D* **56**, 535 (1997) [arXiv:astro-ph/9610219].
- [139] D. H. Lyth and D. Wands, *Phys. Lett. B* **524**, 5 (2002) [arXiv:hep-ph/0110002].
- [140] T. Moroi and T. Takahashi, *Phys. Lett. B* **522**, 215 (2001) [Erratum-ibid. B **539**, 303 (2002)] [arXiv:hep-ph/0110096].
- [141] K. Enqvist and M. S. Sloth, *Nucl. Phys. B* **626**, 395 (2002) [arXiv:hep-ph/0109214].
- [142] D. H. Lyth, C. Ungarelli and D. Wands, *Phys. Rev. D* **67**, 023503 (2003) [arXiv:astro-ph/0208055].
- [143] T. Moroi and T. Takahashi, *Phys. Rev. D* **66**, 063501 (2002) [arXiv:hep-ph/0206026].
- [144] T. S. Bunch and P. C. W. Davies, *Proc. Roy. Soc. Lond. A* **360**, 117 (1978).
- [145] A. Vilenkin and L. H. Ford, *Phys. Rev. D* **26**, 1231 (1982).
- [146] A. D. Linde, *Phys. Lett. B* **116**, 335 (1982).
- [147] A. D. Linde, *Particle Physics and Inflationary Cosmology*, Harwood Academic Publishers, Chur, Switzerland, 1990.
- [148] A. D. Linde, *Phys. Lett. B* **158**, 375 (1985).
- [149] D. Seckel and M. S. Turner, *Phys. Rev. D* **32**, 3178 (1985).
- [150] M. S. Turner, *Phys. Rept.* **197**, 67 (1990).
- [151] E. P. S. Shellard and R. A. Battye, *Phys. Rept.* **307**, 227 (1998) [arXiv:astro-ph/9808220].
- [152] A. D. Linde, *Phys. Lett. B* **259**, 38 (1991).
- [153] M. Dine and A. Anisimov, *JCAP* **0507**, 009 (2005) [arXiv:hep-ph/0405256].
- [154] D. H. Lyth and E. D. Stewart, *Phys. Rev. D* **46**, 532 (1992).
- [155] A. D. Dolgov and D. P. Kirilova, *Yad. Fiz.* **51**, 273 (1990) [Sov. J. Nucl. Phys. **51**, 172 (1990)].
- [156] J. H. Traschen and R. H. Brandenberger, *Phys. Rev. D* **42**, 2491 (1990).
- [157] L. Kofman, A. D. Linde and A. A. Starobinsky, *Phys. Rev. Lett.* **73**, 3195 (1994) [arXiv:hep-th/9405187].
- [158] Y. Shtanov, J. H. Traschen and R. H. Brandenberger, *Phys. Rev. D* **51**, 5438 (1995) [arXiv:hep-ph/9407247].
- [159] D. Boyanovsky, H. J. de Vega, R. Holman, D. S. Lee and A. Singh, *Phys. Rev. D* **51**, 4419 (1995) [arXiv:hep-ph/9408214].
- [160] M. Yoshimura, *Prog. Theor. Phys.* **94**, 873 (1995) [arXiv:hep-th/9506176].
- [161] S. Y. Khlebnikov and I. I. Tkachev, *Phys. Lett. B* **390**, 80 (1997) [arXiv:hep-ph/9608458].
- [162] L. Kofman, A. D. Linde and A. A. Starobinsky, *Phys. Rev. D* **56**, 3258 (1997) [arXiv:hep-ph/9704452].
- [163] R. Micha and I. I. Tkachev, *Phys. Rev. D* **70**, 043538 (2004) [arXiv:hep-ph/0403101].
- [164] S. Y. Khlebnikov and I. I. Tkachev, *Phys. Rev. Lett.* **77**, 219 (1996) [arXiv:hep-ph/9603378].
- [165] L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics* in 10 volumes; Volume I, L. D. Landau, E. M. Lifshitz, *Mechanics* — Third Edition — Reed Educational and Professional Publishing Ltd, 1976.
- [166] P. B. Greene, L. Kofman, A. D. Linde and A. A. Starobinsky, *Phys. Rev. D* **56**, 6175 (1997) [arXiv:hep-ph/9705347].
- [167] G. F. Giudice, M. Peloso, A. Riotto and I. Tkachev, *JHEP* **9908**, 014 (1999) [arXiv:hep-ph/9905242].
- [168] E. W. Kolb, A. D. Linde and A. Riotto, *Phys. Rev. Lett.* **77**, 4290 (1996) [arXiv:hep-ph/9606260]; E. W. Kolb, A. Riotto and I. I. Tkachev, *Phys. Lett. B* **423**, 348 (1998) [arXiv:hep-ph/9801306].

- [169] J. Garcia-Bellido, D. Y. Grigoriev, A. Kusenko and M. E. Shaposhnikov, *Phys. Rev. D* **60**, 123504 (1999) [arXiv:hep-ph/9902449].
- [170] L. Boubekeur, S. Davidson, M. Peloso and L. Sorbo, *Phys. Rev. D* **67**, 043515 (2003) [arXiv:hep-ph/0209256].
- [171] E. J. Ahn and E. W. Kolb, *Phys. Rev. D* **74**, 103503 (2006) [arXiv:astro-ph/0508399].
- [172] L. Kofman, A. D. Linde and A. A. Starobinsky, *Phys. Rev. Lett.* **76**, 1011 (1996) [arXiv:hep-th/9510119].
- [173] I. I. Tkachev, *Phys. Lett. B* **376**, 35 (1996) [arXiv:hep-th/9510146].
- [174] S. Khlebnikov, L. Kofman, A. D. Linde and I. Tkachev, *Phys. Rev. Lett.* **81**, 2012 (1998) [arXiv:hep-ph/9804425].
- [175] S. Y. Khlebnikov and I. I. Tkachev, *Phys. Rev. D* **56**, 653 (1997) [arXiv:hep-ph/9701423].
- [176] J. F. Dufaux, G. N. Felder, L. Kofman and O. Navros, *JCAP* **0903**, 001 (2009) [arXiv:0812.2917 [astro-ph]].
- [177] G. Veneziano, *Phys. Lett. B* **265**, 287 (1991); M. Gasperini and G. Veneziano, *Astropart. Phys.* **1**, 317 (1993) [arXiv:hep-th/9211021].
- [178] M. Gasperini and G. Veneziano, *Phys. Rept.* **373**, 1 (2003) [arXiv:hep-th/0207130].
- [179] J. Khoury, B. A. Ovrut, P. J. Steinhardt and N. Turok, *Phys. Rev. D* **64**, 123522 (2001) [arXiv:hep-th/0103239].
- [180] J. M. Cline, S. Jeon and G. D. Moore, *Phys. Rev. D* **70**, 043543 (2004) [arXiv:hep-ph/0311312].
- [181] V. A. Rubakov and P. G. Tinyakov, *Phys. Usp.* **51**, 759 (2008) [arXiv:0802.4379 [hep-th]].
- [182] E. I. Buchbinder, J. Khoury and B. A. Ovrut, *Phys. Rev. D* **76**, 123503 (2007) [arXiv:hep-th/0702154]; E. I. Buchbinder, J. Khoury and B. A. Ovrut, *JHEP* **0711**, 076 (2007) [arXiv:0706.3903 [hep-th]].
- [183] P. Creminelli and L. Senatore, *JCAP* **0711**, 010 (2007) [arXiv:hep-th/0702165].
- [184] J. L. Lehners, *Phys. Rept.* **465**, 223 (2008) [arXiv:0806.1245 [astro-ph]].
- [185] D. Wands, *Phys. Rev. D* **60**, 023507 (1999) [arXiv:gr-qc/9809062].
- [186] F. Finelli and R. Brandenberger, *Phys. Rev. D* **65**, 103522 (2002) [arXiv:hep-th/0112249].
- [187] K. Koyama and D. Wands, *JCAP* **0704**, 008 (2007) [arXiv:hep-th/0703040].
- [188] J. L. Lehners, P. McFadden, N. Turok and P. J. Steinhardt, *Phys. Rev. D* **76**, 103501 (2007) [arXiv:hep-th/0702153].
- [189] D. Wands, “Cosmological perturbations through the big bang,” arXiv:0809.4556 [astro-ph].
- [190] V. A. Rubakov, *JCAP* **0909**, 030 (2009) [arXiv:0906.3693 [hep-th]].
- [191] E. M. Lifshitz and I. M. Khalatnikov, *Adv. Phys.* **12**, 185 (1963); V. A. Belinsky, I. M. Khalatnikov and E. M. Lifshitz, *Adv. Phys.* **19**, 525 (1970); V. A. Belinsky, E. M. Lifshits and I. M. Khalatnikov, *ZhETF* **62**, 1606 (1972) [Sov. Phys. JETP **36**, 591 (1973)].
- [192] L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics* in 10 volumes; Volume II, L. D. Landau, E. M. Lifshitz, *The Classical Theory of Fields* — Fourth Edition — Reed Educational and Professional Publishing Ltd, 1975.
- [193] J. K. Erickson, D. H. Wesley, P. J. Steinhardt and N. Turok, *Phys. Rev. D* **69**, 063514 (2004) [arXiv:hep-th/0312009].
- [194] V. A. Rubakov, *Classical Theory of Gauge Fields* — Princeton University Press, 2002.

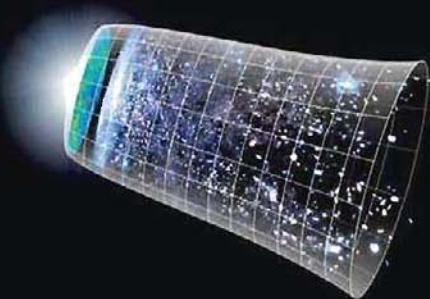
This page is intentionally left blank

Index

- acoustic (sound) horizon, 54, 105, 218
- acoustic oscillations, 56
- angular diameter distance, 119, 130
- anisotropic stress
 - potential, 148, 428
 - tensor, 23, 147, 427
- axion field, 37
- axion string, 366
- BAO distance scale, 119
- baryon asymmetry generation, 64
- baryon fraction parameter R_B , 92, 193
- baryon number conservation, 92
- baryon-to-photon ratio η_B , 5
- basis helicity-2 tensors, 25
- Belinsky–Lifshits–Khalatnikov theory, 403
- Bogoliubov coefficients, 376, 455
- Boltzmann equation, 138, 142, 190
 - linearized, 146, 192
- Bunch–Davis vacuum, 320
- Christoffel symbols
 - in linearized theory, 146, 427
- cluster counting, 113
- CMB
 - cross correlation spectra C_l^{TE} , C_l^{TB} , 249
 - polarization
 - B -mode angular spectrum C_l^{BB} , 249
 - E -mode angular spectrum C_l^{EE} , 249
 - bound on B -mode, 277
 - degree, 242
 - direction, 243
- secondary, 242
- temperature, 197
- temperature anisotropy
 - correlation function, 199
 - dipole component, 198
 - Gaussianity, 198
 - multipoles $C_l \equiv C_l^{TT}$, 198
 - quadrupole component, 211
 - secondary, 230
- collision integral, 143, 190
- Compton scattering, 239
- conformal
 - field, 449
 - symmetry, 386
- continuity equation, 3, 57
- correlation function
 - of galaxies, 112
 - of matter, 110, 116
- cosmic variance, 73, 199
- covariant conservation equations
 - linearized, 23, 30, 32, 61, 148
- curvaton field, 37
- curvature of space, bound, 15
- cyclic scenario, 403
- dark matter
 - cold, 6
 - hot, 149
 - warm, 45, 149
- dark matter generation, 64
- degeneracy in cosmological
 - parameters, 119
- density contrast
 - smoothed δ_R , 122
 - smoothed, variance σ_R , 124

- de Sitter space-time, 356
- Doppler effect, 204
- e-foldings, 288, 299
 - minimal number, 288
- Einstein equations
 - linearized, 30, 31, 61, 148
- Einstein tensor, 20, 426
 - linearized, 27, 31
- ekpyrotic scenario, 401
- entropy modes, 64
- Euler equation, 3, 57
- first stars, 12, 121, 225
- free streaming, 149
 - length, 150
 - of photons, 186
- Friedmann equation, 14
- gauge transformations in General Relativity, 21
 - linearized, 21
- ghost field, 402
- gravitational lensing
 - weak, 113, 201, 230, 242
- gravitino, 159, 165
- gravity waves, 27, 29
 - relic, 336
 - energy density, 337
- Guth inflationary model, 292
- Harrison–Zeldovich spectrum, 73
- horizon entry, 36
- hypersurfaces
 - of comoving frame, 71
 - of constant energy density, 69
 - spatially flat, 70
- inflaton, 296
- isothermal modes, 64
- Jeans
 - length, 4, 150
 - mass, 5
 - time, 5
- Kasner metric, 404
- Kibble mechanism, 396
- Landau damping, 149
- Liouville equation, 138
- Lyman- α forest, 114, 159
- Mandelstam variables, 141
- Mathieu equation, 383
- mode of perturbations
 - constant, 49
 - decaying, 49
 - MD-entering, 36, 50
 - RD-entering, 36, 50
- monopole problem, 289
- Mukhanov–Sasaki variable, 329
- neutralino, 154
 - decoupling, 154
- neutrino
 - effect on gravity waves, 181
 - effective number of species, 166
 - fraction at radiation domination, 166
 - masses, 166
 - number density, 166
 - sterile, 159, 165
 - temperature, 165
- neutrino multipoles $F_{\nu,l}$, 172
- old inflation model, 292
- optical depth, 186, 226, 242
- parameter σ_8 , 126
- Peccei–Quinn symmetry, 365
- phantom energy, 401
- phantom matter, 401
- phase space density
 - coarse grained, 163
- photon multipoles $F_{\gamma,l}$, 192
- pre-Big Bang scenario, 401
- Press–Schechter formalism, 122
- reionization, 159, 225, 242, 254
- resonance
 - narrow, 382
 - stochastic, 377, 380
 - wide, 377
- Sachs–Wolfe effect, 203
 - integrated, 204
 - correlation with matter distribution, 206
 - early, 205
 - late, 205

- Sakharov oscillations, 115, 201
 scalar field, superlight, 45
 scalar perturbations
 adiabatic, 63
 bounds on non-Gaussianity
 parameter f_{NL} , 76
 non-Gaussianity parameter
 f_{NL} , 76
 normalization $A_{\mathcal{R}}$, 73
 primordial amplitude $\Delta_{\mathcal{R}}$, 73
 primordial gravitational
 potential spectrum P_{Φ} , 74
 primordial power spectrum $P_{\mathcal{R}}$,
 72, 324
 running index $dn_s/d \log k$, 73
 spectral tilt ($n_s - 1$), 73
 adiabatic-isocurvature correlation
 parameter β , 75
 baryon isocurvature mode, 63
 bounds on admixture of isocurvature
 modes, 75
 CDM isocurvature mode, 63
 Gaussianity, 72, 76
 neutrino isocurvature modes, 64
 of metric, action, 431
 Silk damping, 118, 189, 228, 241
 Silk length, 190, 195
 Solar mass, 5
 sound horizon, 54, 218
 at equality, 218
 at last scattering, 105, 115, 218
 angular size, 220
 sound speed, 3, 32, 44, 54
 in baryon-photon plasma, 93
 spherical harmonics of helicity
 ± 2 , ${}_s Y_{lm}(\mathbf{n})$, 249
 spin connection, 446
 Starobinsky inflationary model, 285, 310
 Stokes parameters, 243
 subhorizon regime, 33
 Sunyaev–Zeldovich effect, 231
 super-stiff equation of state, 404
 superhorizon regime, 33
 suppression factor due to dark energy
 $g(z)$, 59
 tensor perturbations, 75
 action, 28
 primordial amplitude, 332
 primordial power spectrum, 75, 331
 tensor-to-scalar ratio r , 76
 Thomson cross section, 185
 tight coupling approximation, 82, 91
 trans-Planckian problem, 345
 transfer function, 13, 116
 Tremaine–Gunn bound, 164
 variable \mathcal{S}_{λ} , 71
 variable ζ , 66
 variable \mathcal{R} , 68
 vector perturbations of metric
 action, 431
 velocity potential, 9, 26, 428
 effective, 327
 physical, 31
 vierbein, 144, 445
 visibility function $V(\eta_1, \eta_2)$, 186
 Vlasov equation, 139, 150
 Wick theorem, 315, 436
 WIMPs, 153
 window function, 122



INTRODUCTION TO THE THEORY OF THE EARLY UNIVERSE

Cosmological Perturbations
and Inflationary Theory

This book accompanies another book by the same authors, *Introduction to the Theory of the Early Universe: Hot Big Bang Theory* and presents the theory of the evolution of density perturbations and relic gravity waves, theory of cosmological inflation and post-inflationary reheating. Written in a pedagogical style, the main chapters give a detailed account of the established theory, with derivation of formulas. Being self-contained, it is a useful textbook for advanced undergraduate students and graduate students. Essential materials from General Relativity, theory of Gaussian random fields and quantum field theory are collected in the appendices. The more advanced topics are approached similarly in a pedagogical way. These parts may serve as a detailed introduction to current research.

