THEORY OF COSMOLOGICAL PERTURBATIONS

V.F. MUKHANOV[a,b], H.A. FELDMAN[c] and R.H. BRANDENBERGER[a]

[a]*Physics Department, Brown University, Providence, RI 02912, USA*
[b]*Institute for Nuclear Research, Academy of Sciences, Moscow 117 312, Russia*
[c]*Canadian Institute for Theoretical Astrophysics, University of Toronto, Toronto ON, M5S 1A1, Canada*

NORTH-HOLLAND

PHYSICS REPORTS (Review Section of Physics Letters) 215, Nos. 5 & 6 (1992) 203–333. North-Holland PHYSICS REPORTS

Theory of cosmological perturbations

V.F. Mukhanov[a,b], H.A. Feldman[c,*] and R.H. Brandenberger[a]

[a]Physics Department, Brown University, Providence, RI 02912, USA
[b]Institute for Nuclear Research, Academy of Sciences, Moscow 117 312, Russia
[c]Canadian Institute for Theoretical Astrophysics, University of Toronto, Toronto ON, M5S 1A1, Canada

Received November 1991; editor: D.N. Schramm

Contents:

* Address beginning September 1, 1991: Physics Department, University of Michigan, 500 East University, Ann Arbor, MI 48109, USA.

Abstract:

We present in a manifestly gauge-invariant form the theory of classical linear gravitational perturbations in part I, and a quantum theory of cosmological perturbations in part II. Part I includes applications to several important examples arising in cosmology: a universe dominated by hydrodynamical matter, a universe filled with scalar-field matter, and higher-derivative theories of gravity. The growth rates of perturbations are calculated analytically in most interesting cases. The analysis is applied to study the evolution of fluctuations in inflationary universe models. Part II includes a unified description of the quantum generation and evolution of inhomogeneities about a classical Friedmann background. The method is based on standard canonical quantization of the action for cosmological perturbations which has been reduced to an expression in terms of a single gauge-invariant variable. The spectrum of density perturbations originating in quantum fluctuations is calculated in universes with hydrodynamical matter, in inflationary universe models with scalar-field matter, and in higher-derivative theories of gravity.

The gauge-invariant theory of classical and quantized cosmological perturbations developed in parts I and II is applied in part III to several interesting physical problems. It allows a simple derivation of the relation between temperature anisotropies in the cosmic microwave background radiation and the gauge-invariant potential for metric perturbations. The generation and evolution of gravitational waves is studied. As another example, a simple analysis of entropy perturbations and non-scale-invariant spectra in inflationary universe models is presented. The gauge-invariant theory of cosmological perturbations also allows a consistent and gauge-invariant definition of statistical fluctuations.

General introduction

The theory of linearized gravitational perturbations in an expanding universe – cosmological perturbations for short – has become a cornerstone of modern cosmology. It is used to describe the growth of structure in the universe, to calculate the predicted microwave background fluctuations, and in many other considerations.

Over the past decade, new methods to study linearized gravitational perturbations have been developed, most notably a gauge-invariant approach which is simpler and easier to apply than previous methods. The goal of this review article is to give a survey of the theories of classical and quantum cosmological perturbations, with particular emphasis on the gauge-invariant approach. We present a general discussion of cosmological perturbations and of the definition and meaning of gauge-invariant metric perturbation. Further, a general set of equations of motion for these quantities are derived and applied to the study of the evolution of cosmological perturbations in the most important cosmological models. The issue of initial conditions is addressed with particular care. Both quantum generation and evolution of the fluctuations is discussed in a unified manner.

There is good observational evidence that the universe was very homogeneous and isotropic on all scales at early times. It is usually assumed that there exist small primordial perturbations which slowly increase in amplitude due to gravitational instability to form the structures we observe at the present time on the scales of galaxies and galaxy clusters. The growth of these primordial perturbations is a problem ideally suited to be solved by applying linear gravitational perturbation theory.

The best evidence for homogeneity and isotropy of the universe at early times comes from the isotropy of the microwave background radiation. The present upper limits on the temperature fluctuations are less than 1 part in 10^4 on all angular scales [1]. In many cosmological models, the present temperature fluctuations are directly related to the density perturbations at the time of recombination, thus providing an upper bound on the amplitude of the density fluctuations at that time [2].

However, in order to explain nonlinear structures today on the scale of galaxies and clusters, we require initial perturbations. It is rather natural to assume (and this is indeed realized in most models of structure formation) that the perturbations start out at a very early time with a small amplitude and gradually grow in time.

The growth of perturbations in an expanding universe is a consequence of gravitational instability. A small overdensity will exert an extra gravitational attractive force on the surrounding matter. Consequently, the perturbation will increase and will in turn produce a larger attractive force. In a nonexpanding background, this would lead to an exponential instability. In an expanding universe, however, the increase in force is partially counteracted by the expansion. This, in general, results in power-law growth rather than exponential growth of the perturbations.

Mathematically, the problem of describing the growth of small perturbations in the context of general relativity reduces to solving the Einstein equations linearized about an expanding background. In principle, this sounds like a straightforward (albeit maybe rather tedious) task. However, there are complicating issues related to the freedom of gauge, i.e., the choice of background coordinates. Physical variables must be independent of this choice. In this review we develop a formalism which uses variables that are independent of the background coordinates, i.e., gauge-invariant variables. In this framework the physical interpretation of the results is unambiguous. (In electromagnetism this corresponds to working with electric and magnetic fields instead of vector potentials.)

Although there is no consistent quantum theory of gravity, gravitational waves can be quantized in a consistent manner. In this review we generalize this construction to linearized density perturbations. The resulting theory can be used to describe the origin of structure and its early evolution in inflationary universe models.

The review article is organized in three parts. Each part has an abstract, an introduction, the main part, a conclusion and is reasonably self-contained. Part I is devoted to the formulation and solution of the classical equations for cosmological perturbations. In part II, the quantum origin of fluctuations is studied, and the spectrum of perturbations is calculated in models of cosmological interest. In the third part we apply the theory to various important cosmological problems. The entire analysis uses a gauge invariant formalism.

The goal of the first part of this review article is to provide a new exposition of gauge-invariant linear perturbation theory and to find the solutions of the gauge-invariant equations of motion in the most interesting cases. The derivation of the equations of motion in a new and simple form is presented. The formalism is applied to a hydrodynamical universe, to a universe dominated by scalar fields (with application to inflationary universe models), and is extended to analyze perturbations in higher-derivative theories of gravity. The growth rate of perturbations is calculated analytically in many cases.

Our aim in part II is to develop the quantum theory of cosmological perturbations, and to calculate the spectrum of metric and density perturbations, starting with initial quantum fluctuations, in general cosmological models and, in particular, for the cases analyzed in part I. The theory developed here provides the initial conditions for the classical evolution of inhomogeneities.

In part III, we apply the theory developed in parts I and II to problems in physical cosmology. We consider four important physical applications of cosmological perturbation theory: microwave background anisotropies, gravitational waves, entropy perturbations and statistical fluctuations. All of these topics address issues which directly relate primordial perturbations to observable quantities, hence providing constraints on models of the early universe. The gauge-invariant approach simplifies and clarifies the analysis in all four cases.

In the appendix we explain some general rules, give units and define constants that are used throughout the article. In tables 1 and 2 we summarize the most important equations.

PART I. CLASSICAL PERTURBATIONS

1. Introduction

In this part of the review article we develop the classical theory of cosmological perturbations in a manifestly gauge-invariant framework. We then apply the formalism to the three models of greatest interest for cosmology: hydrodynamical perturbations, fluctuations in a scalar-field dominated universe and inhomogeneities in higher-derivative gravity models. Most other cosmological theories can be reduced to the above three models.

Mathematically, the problem of describing the growth of small perturbations in the context of general relativity reduces to solving the Einstein equations linearized about an expanding background. In principle, this sounds like a straightforward (albeit maybe rather tedious) task. However, there are complicating issues related to the freedom of gauge. Not all perturbed metrics correspond to perturbed space–times. It is possible to obtain an inhomogeneous form for the metric $g_{\mu\nu}(\boldsymbol{x}, t)$ in a homogeneous and isotropic space–time by an inconvenient choice of coordinates. Hence, it is important to be able to distinguish between physical (geometrical) inhomogeneities and mere coordinate artifacts.

There are several approaches to this problem. One may fix the coordinate system uniquely based on some specific geometrical requirements. One may pick simple coordinate conditions (which in general do not totally fix the coordinates) and keep careful track of physical modes and coordinate artifacts (gauge modes) – the usual synchronous gauge approach is an example of this procedure. However, most convenient is the gauge-invariant approach in which one considers metric variables which are independent of the choice of coordinates.

As we hope to demonstrate in this review, the gauge-invariant approach is easier to work with than other methods. Further, all physical quantities are gauge invariant. Thus, our framework has a clear physical interpretation which is particularly apparent in the derivation we follow. The Newtonian limit of relativistic gravitational perturbation theory follows in a natural way. Finally, when using a gauge-invariant approach we avoid the danger of mistaking coordinate artifacts as physical effects – a danger which is particularly great when working with approximate solutions.

The method of working with gauge-invariant variables is well known from other physical theories. For example, in classical electrodynamics it is usually more physical to work in terms of the gauge-invariant electric and magnetic fields rather than in terms of the gauge-dependent scalar and vector potentials. However, the mathematical structure of electrodynamics sometimes makes the equations simpler using the potentials. This is because by introducing potentials we have automatically solved the homogeneous Maxwell equations. For linearized gravity, there is no such simplification and it is both more physical and more convenient to use gauge-invariant variables.

As for electrodynamics, also for linearized gravity there are an infinite number of gauge-invariant variables. We will choose a basis of gauge-invariant variables for which the equations of motion take on a particularly simple form. This is similar to choosing the electric and magnetic fields as the basic variables in electrodynamics. The variables we choose will coincide with the two functions used to describe metric perturbations in a particular gauge, the longitudinal (or conformal-Newtonian) gauge. The gauge-invariant equations of motion are identical to the perturbed Einstein equations in this gauge.

The pioneering work on density perturbations in Friedmann–Robertson–Walker (FRW) cosmological models is that of Lifshitz [3] as summarized by Lifshitz and Khalatnikov [4]. After that, the subject was studied by many authors. The textbooks by Weinberg [5], Peebles [6], and Zel'dovich and Novikov [7] treat cosmological perturbations in some detail. Although thus standarized, the subject continued to be plagued by difficulties in interpretation. The early papers did not adequately address the gauge freedom. Most treatments used a particular gauge choice, "synchronous gauge", which does not totally fix the coordinates.

In synchronous gauge, the interpretation of perturbations whose wavelength is larger than the horizon size is not always straightforward. Press and Vishniac [8] proposed a scheme for keeping track of gauge modes while continuing to calculate in standard synchronous gauge. However, this procedure does not eliminate the problems of this gauge.

A more elegant way to deal with the gauge problem is to eliminate the gauge dependence entirely rather than to just specify and understand it. The gauge-invariant approach to gravitational perturbations was pioneered by Bardeen [9, 10] and Gerlach and Sengupta [11]. It is based on previous work by Hawking [12], Field and Shepley [13], Harrison [14] and Olson [15]. Bardeen [10] applied the gauge-invariant approach to cosmological perturbations.

Initially, the gauge-invariant approach appeared to involve a computational tour de force. However, the increasing interest in cosmological perturbations generated by the advent of inflationary universe models created a lot of activity in this area. In ref. [16], it was shown that Bardeen's equations for the gauge-invariant variables can be derived in a straightforward manner starting from synchronous gauge. The gauge-invariant approach to cosmological perturbations was studied extensively in refs. [17–19]. It has been applied to construct a self-consistent quantum theory of metric perturbations [20, 21], to investigate eternal [22] and stochastic inflation [23], to follow the dynamics of inflationary universe models [24], and to analyze the stability of inflation in higher-derivative theories of gravity [25]. Other applications include a gauge-invariant analysis of perturbations through the decoupling epoch [26] and a study of fluctuations [27] in the Cold Dark Matter model of structure formation. Den and Tomita [28] have extended the gauge-invariant formalism to anisotropic cosmologies.

In our opinion, the simplest derivation of the equations of motion for gauge-invariant variables is obtained by working in longitudinal (conformal–Newtonian) gauge, in which Bardeen's gauge-invariant variables are identical to the remaining metric variables. This approach was pioneered in ref. [29].

The goal of the first part of this review article is to provide a new exposition of gauge-invariant linear perturbation theory and to find the solutions of the gauge-invariant equations of motion in the most interesting cases. The derivation of the equations of motion in a new and simple form is presented. The formalism is applied to a hydrodynamical universe, to a universe dominated by scalar fields (with application to inflationary universe models), and is extended to analyze perturbations in higher-derivative theories of gravity. The growth rate of perturbations can be calculated analytically in many cases.

Recently, there has been a lot of work on alternative approaches to cosmological perturbations. A gauge-invariant formalism based on a covariant approach was elaborated by Bruni et al. [30] (see also refs. [31–33], drawing on previous work by Hawking [12] and Olson [15]). A gauge-invariant formalism based on the $3+1$ Hamiltonian form of general relativity was developed by Durrer and Straumann [34]. Several authors have taken up the old synchronous-gauge approach and have refined it in order to take the gauge ambiguity into account [35, 36].

It is important to point out a fundamental limitation of linearized gravitational perturbation theory. The concepts are restricted to small perturbations of FRW space–times. If ε is a measure of the amplitude of the perturbation, then there are correction terms of the order ε^2 to all the equations. In particular, the gauge-invariant variables used here are invariant only to order ε. The only perturbation variables which are invariant under large gauge transformations are perturbations of quantities which are constant in both space and time on the background manifold (see ref. [38] for a discussion of this point). The authors of ref. [30] have recently developed a formalism based on gauge-invariant variables which are perturbations of quantities which vanish on the background. Hence, this formalism can in principle be extended to arbitrary orders in perturbation theory. However, this framework appears cumbersome. In ref. [39] the equivalence of the approach developed in ref. [30] with the formalism developed here has been shown. The Ellis and Bruni methods have been extended to multi-fluid and scalar-field models [40].

One of the main reasons for the upsurge of interest in linearized gravitational perturbation theory is its application to new cosmological theories, in particular the theory of the inflationary universe [41]. These theories were developed starting in about 1980 and for the first time allow a causal explanation of the origin of structures in the universe such as galaxies, clusters and even larger entities.

Two ideas were crucial in developing a causal theory of structure formation in inflationary universe models. First, it was realized in 1980 independently by Chibisov and Mukhanov [42] and by Lukash

[43], that quantum fluctuations in an expanding universe can lead to classical energy-density perturbations. (First attempts in this direction were made by Sakharov [44] in 1965.) Secondly, it was realized [45, 46, 42, 43] that in an inflationary universe, scales inside the Hubble radius $H^{-1}(t) = [\dot{a}(t)/a(t)]^{-1}$ at the beginning of the period of inflation will expand exponentially and reenter the Hubble radius at late time as large-scale cosmological perturbations with a scale-invariant [47] spectrum.

Mukhanov and Chibisov [48] calculated the spectrum of density perturbations in a model in which higher-derivative terms in the gravitational action lead to a period of exponential expansion [49]. The spectrum of density perturbations in New Inflationary Universe models was first estimated by Starobinsky [50], Hawking [51], Guth and Pi [52], Bardeen et al. [53] and Mukhanov and Chibisov [54]. Subsequently, the evolution of fluctuations in new inflation was studied intensively [55–60]. The analysis was extended to chaotic inflation [61]. The gauge-invariant approach makes the analysis more tractable.

The outline of this part of our review article is as follows. In chapter 2, the metrics for the background models and perturbations are described. In chapter 3 we discuss gauge invariance and introduce the gauge-invariant variables. Chapter 4 is a derivation of the first-order perturbation of the Einstein tensor and Einstein equations, first for a general perturbed metric and then in gauge-invariant form. Chapters 2–4 are all short and provide the background needed to analyze the evolution of perturbations in the three important cosmological models studied in chapters 5–7.

Hydrodynamical perturbations are discussed in chapter 5. We derive the perturbed energy–momentum tensor $T_{\mu\nu}$ and write down the equations of motion. Next, a detailed study of adiabatic and entropy perturbations is presented, applying the gauge-invariant equations. In the cases of radiation- and matter-dominated universes, exact solutions for linear perturbations are derived. In a universe containing both matter and radiation it is possible to find exact solutions for long-wavelength perturbations.

Chapter 6 is a survey of perturbations in a theory in which matter is represented by a scalar field (as in inflationary universe models). First, the solutions of the background equations are determined and the results presented in terms of a phase space analysis. Then, we write down the perturbed $T_{\mu\nu}$ and derive the equations of motion for perturbations. These equations are then applied to study classical density perturbations in inflationary universe models.

Finally, in chapter 7, the perturbation equations for higher-derivative gravity theories are determined. We discuss the background solutions, derive the perturbed Einstein tensor, and find solutions of the gauge-invariant equations of motion.

Units are used in which $c = \hbar = k_B = 1$. Greek indices run from 0 to 3, Latin indices only over the spatial degrees of freedom. The Einstein summation convention is assumed. $a(t)$ is the scale factor of the background Friedmann–Robertson–Walker (FRW) model. G is Newton's gravitational constant. We draw the attention of the reader to the difference between φ (scalar field), ϕ (Newtonian gravitational potential) and Φ (gauge-invariant variable). In longitudinal gauge, ϕ and Φ are equal.

2. Background model and perturbations

Everywhere in this article we shall assume that space–time deviates only by a small amount from a homogeneous, isotropic idealized space–time which is defined to be the background. In this case, it is convenient to split the metric into two parts, the first being the background metric, the other describing how the "real" space–time deviates from the idealized background model. The second part is called the perturbation. The observational fact that the universe on large scales is nearly homogeneous and isotropic makes this approach reasonable. Further, it was shown [37] that in Robertson–Walker universes solutions of the linearized field equations can be viewed as linearizations of solutions of the full nonlinear equations. Hence linear perturbation theory is mathematically well defined.

The background line element is

$$\mathrm{d}s^2 = {}^{(0)}g_{\mu\nu}(x)\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = \mathrm{d}t^2 - a^2(t)\gamma_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j = a^2(\eta)(\mathrm{d}\eta^2 - \gamma_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j)\,, \tag{2.1}$$

where η is the conformal time $\mathrm{d}\eta = a^{-1}\,\mathrm{d}t$. We choose the background metric to be the Friedmann–Robertson–Walker (FRW) metric, in which case

$$\gamma_{ij} = \delta_{ij}[1 + \tfrac{1}{4}\mathcal{K}(x^2 + y^2 + z^2)]^{-2}\,, \tag{2.2}$$

where $\mathcal{K} = 0, 1, -1$ depending on whether the three-dimensional space corresponding to the hypersurface $\eta = \text{const.}$ is flat, closed or open. The Einstein equations are

$$G^\mu_\nu = R^\mu_\nu - \tfrac{1}{2}\delta^\mu_\nu R = 8\pi G T^\mu_\nu\,, \tag{2.3}$$

where R^μ_ν is the Ricci tensor, $R \equiv R^\mu_\mu$ is the Ricci (curvature) scalar and T^μ_ν is the energy–momentum tensor. For the background metric in eq. (2.1) in conformal time, the Einstein equations reduce to the 0–0 equation,

$$a' + \mathcal{K}a^2 = \tfrac{8}{3}\pi G T^0_0 a^4\,, \quad a' \equiv \mathrm{d}a/\mathrm{d}\eta \tag{2.4a, b}$$

and the i–i equation,

$$a'' + \mathcal{K}a = \tfrac{4}{3}\pi G T a^3\,, \quad T \equiv T^\mu_\mu \tag{2.5a, b}$$

For the background metric (2.1), the space–space part of the Ricci tensor R^i_j is proportional to δ^i_j. Thus, for an isotropic universe, which the background is constructed to be, the energy–momentum tensor must also be spatially diagonal, i.e. $T^i_j \propto \delta^i_j$, in order that the Einstein equations are satisfied. Differentiating (2.4a) with respect to η and subtracting $2a'$ times (2.5a) we get the continuity equation for matter $T^\mu_{0;\mu} = 0$ or

$$\mathrm{d}T^0_0 = -(4T^0_0 - T)\,\mathrm{d}\ln a\,. \tag{2.6}$$

To model the universe more realistically, we must include the perturbations. The full line element may be represented by

$$\mathrm{d}s^2 = {}^{(0)}g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu + \delta g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu\,, \tag{2.7}$$

where the $\delta g_{\mu\nu}$ describe the perturbation. The full metric has been decoupled into its background and perturbation parts

$$g_{\mu\nu} = {}^{(0)}g_{\mu\nu} + \delta g_{\mu\nu}\,. \tag{2.8}$$

The metric perturbations may be categorized into three distinct types: scalar, vector and tensor perturbations. This classification refers to the way in which the fields from which $\delta g_{\mu\nu}$ are constructed transform under three-space coordinate transformations on the constant-time hypersurface. (Recently, Stewart [62] has given a covariant description of this tensor decomposition.) Both vector and tensor perturbations exhibit no instabilities. Vector perturbations decay kinematically in an expanding universe whereas tensor perturbations lead to gravitational waves which do not couple to energy density and pressure inhomogeneities. However, scalar perturbations may lead to growing inhomogeneities which, in turn, have an important effect on the dynamics of matter.

Scalar perturbations. There are two possible ways that scalar quantities may enter into δg_{ij}; either by multiplying the tensor γ_{ij} with a scalar, or by taking covariant derivatives of a scalar function, the covariant derivative being with respect to the background metric γ_{ij} of the constant-time hypersurface. In a spatially flat universe ($\mathcal{K}=0$), these covariant derivatives become ordinary partial derivatives, denoted by a comma with corresponding index. In general, we denote the (background) three-dimensional covariant derivative of a function f with respect to some coordinate i by $f_{|i}$.

To complete the specification of a general scalar metric perturbation, we need two more scalar functions. The first gives δg_{00}, and the three-dimensional background covariant derivative of the second gives δg_{i0}. Thus, the most general form of the scalar metric perturbations is constructed using four scalar quantities ϕ, ψ, B and E which are functions of space and time coordinates,

$$\delta g_{\mu\nu}^{(s)} = a^2(\eta)\begin{pmatrix} 2\phi & -B_{|i} \\ -B_{|i} & 2(\psi\gamma_{ij} - E_{|ij}) \end{pmatrix} . \tag{2.9}$$

From (2.7) and (2.9), we get the most general form of the line element for the background and scalar metric perturbations to be

$$\mathrm{d}s^2 = a^2(\eta)\{(1+2\phi)\,\mathrm{d}\eta^2 - 2B_{|i}\,\mathrm{d}x^i\,\mathrm{d}\eta - [(1-2\psi)\gamma_{ij} + 2E_{|ij}]\,\mathrm{d}x^i\,\mathrm{d}x^j\} . \tag{2.10}$$

Vector perturbations. The vector perturbations are constructed using two three-vectors S_i and F_i satisfying the constraints

$$S_i^{|i} = F_i^{|i} = 0 , \tag{2.11}$$

where we shift from upper to lower three-space indices and vice versa by using the spatial background metric tensor γ_{ij} and its inverse γ^{ij}. That the above constraints are necessary can be seen as follows: if the three-divergence of each of the vectors did not vanish, we could separate them into a divergenceless vector and the gradient of a scalar. Hence, we would not be dealing with a pure vector perturbation. These considerations lead to the following metric for vector perturbations:

$$\delta g_{\mu\nu}^{(v)} = -a^2(\eta)\begin{pmatrix} 0 & -S_i \\ -S_i & F_{i|j} + F_{j|i} \end{pmatrix} . \tag{2.12}$$

Tensor perturbations. Tensor perturbations are constructed using a symmetric three-tensor h_{ij} which satisfies the constraints

$$h_i^i = 0 , \qquad h_{ij}^{\ \ |j} = 0 . \tag{2.13}$$

These constraints mean that h_{ij} does not contain pieces which transform as scalars or vectors. Thus, the metric for tensor perturbations is

$$\delta g^{(t)}_{\mu\nu} = -a^2(\eta)\begin{pmatrix} 0 & 0 \\ 0 & h_{ij} \end{pmatrix} . \tag{2.14}$$

Counting the number of independent functions we used to form $\delta g_{\mu\nu}$, we find that we have four functions for scalar perturbations, four functions for vector perturbations (two three-vectors with one constraint each) and two functions for tensor perturbations (a symmetric three-tensor has six independent components and there are four constraints). Thus we have ten functions all together. This number coincides with the number of indepenent components of $\delta g_{\mu\nu}$.

In the linear approximation, scalar, vector and tensor perturbations evolve independently and thus can be considered separately. In the following we concentrate on scalar perturbations since they are the ones that exhibit instabilities and may lead to the formation of structure.

3. Gauge-invariant variables and their physical meaning

First of all, we shall explain the definition and meaning of gauge transformations in the context of small perturbations of a homogeneous and isotropic background space–time. There are two mathematically equivalent approaches to the problem – the passive and active ones. In the passive approach we consider a physical space–time manifold $\mathcal{M}$ and choose some system of coordinates x^α on $\mathcal{M}$. A background model is defined by assigning to all functions Q on $\mathcal{M}$ a previously given function ${}^{(0)}Q(x^\alpha)$. The functions Q may be scalar, vector or tensor quantities. ${}^{(0)}Q(x^\alpha)$ are fixed functions of the coordinates (they are not geometrical quantities). Therefore, in a second coordinate system $\tilde{x}^\alpha$ the background functions ${}^{(0)}Q(\tilde{x}^\alpha)$ will have exactly the same functional dependence on $\tilde{x}^\alpha$. The perturbation δQ of the quantity Q in the system of coordinates x^α is defined as

$$\delta Q(p) = Q(x^\alpha(p)) - {}^{(0)}Q(x^\alpha(p)) , \tag{3.1}$$

and can be evaluated for any point $p \in \mathcal{M}$ with associated coordinates $x^\alpha(p)$. Similarly, in the second system of coordinates, the perturbation of Q is

$$\widetilde{\delta Q}(p) = \tilde{Q}(\tilde{x}^\alpha(p)) - {}^{(0)}Q(\tilde{x}^\alpha(p)) . \tag{3.2}$$

Here, $\tilde{Q}(\tilde{x}^\alpha(p))$ is the value of Q in the new coordinate system at the same point p of $\mathcal{M}$, and, as we stress again, ${}^{(0)}Q(\tilde{x}^\alpha(p))$ is the same function of $\tilde{x}^\alpha$ as ${}^{(0)}Q(x^\alpha(p))$ is of x^α. The transformation $\delta Q(p) \to \widetilde{\delta Q}(p)$ is called the gauge transformation associated with the change of variables $x^\alpha \to \tilde{x}^\alpha$ on the manifold $\mathcal{M}$.

In the second approach (the active one) we consider two manifolds – the physical manifold $\mathcal{M}$ and a background space–time $\mathcal{N}$ on which coordinates $x^\alpha_{\rm b}$ are rigidly fixed, where the index b stands for "background". Any diffeomorphism $\mathcal{D}: \mathcal{N} \to \mathcal{M}$ induces a system of coordinates on $\mathcal{M}$ via $\mathcal{D}: x^\alpha_{\rm b} \to x^\alpha$ (see fig. 3.1). For a given diffeomorphism $\mathcal{D}$ we define the perturbation δQ of the function Q (defined on $\mathcal{M}$) as

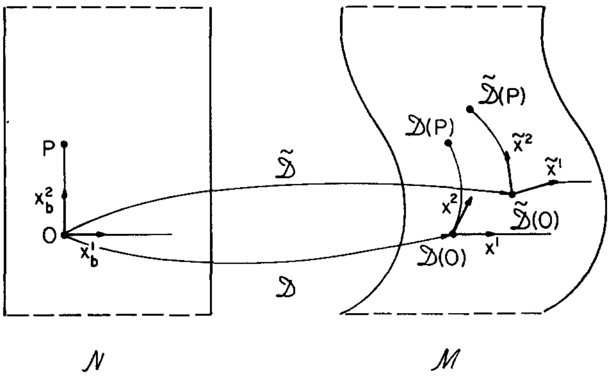

Fig. 3.1. Sketch of two diffeomorphisms $\mathscr{D}$ and $\tilde{\mathscr{D}}$ mapping a homogeneous background manifold $\mathscr{N}$ onto an inhomogeneous manifold $\mathscr{M}$. For fixed background coordinates x_b^α on $\mathscr{N}$, the diffeomorphisms $\mathscr{D}$ and $\tilde{\mathscr{D}}$ induce two coordinate systems x^α and $\tilde{x}^\alpha$ on $\mathscr{M}$. Also drawn are the integral curves of the basis tangent vectors of the three coordinate systems.

$$\delta Q(p) = Q(p) - {}^{(0)}Q(\mathscr{D}^{-1}(p)) , \tag{3.3}$$

where ${}^{(0)}Q$ is a fixed function defined on the background space-time. A second diffeomorphism $\tilde{\mathscr{D}}$: $\mathscr{N} \rightarrow \mathscr{M}$ induces a new set of coordinates $\tilde{x}^\alpha$ on $\mathscr{M}$ via $\tilde{\mathscr{D}}$: $x_b^\alpha \rightarrow \tilde{x}^\alpha$ and a different $\widetilde{\delta Q}$,

$$\widetilde{\delta Q}(p) = \tilde{Q}(p) - {}^{(0)}Q(\tilde{\mathscr{D}}^{-1}(p)) , \tag{3.4}$$

where $\tilde{Q}$ is the value of Q in $\tilde{x}^\alpha$ coordinates. In this approach, the gauge transformation $\delta Q(p) \rightarrow \delta\tilde{Q}(p)$ is generated by the change of correspondence $\mathscr{D} \rightarrow \tilde{\mathscr{D}}$ between the manifolds $\mathscr{N}$ and $\mathscr{M}$. We can associate with this change in correspondence the change of coordinates $x^\alpha \rightarrow \tilde{x}^\alpha$ induced on $\mathscr{M}$.

Both approaches are equivalent. From the point of view of physics, the first allows one to connect the gauge transformation with the choice of the system of coordinates on $\mathscr{M}$ in which the perturbations are described. The second view allows one to understand how the amplitudes of the perturbations depend on the correspondence between background manifold $\mathscr{N}$ and physical manifold $\mathscr{M}$. For an alternate coordinate-invariant definition of a gauge transformation see ref. [62].

In both of the approaches described above one may consider infinitesimal coordinate transformations

$$x^\alpha \rightarrow \tilde{x}^\alpha = x^\alpha + \xi^\alpha , \tag{3.5}$$

described by four functions ξ^α of space and time. This transformation leads to the following change in δQ:

$$\Delta Q = \widetilde{\delta Q} - \delta Q = \mathscr{L}_\xi Q , \tag{3.6}$$

where $\mathscr{L}_\xi$ is the Lie derivative in direction of the vector ξ. The transformations given by (3.5) and (3.6) form a group, the group of infinitesimal coordinate transformations or the gauge group of gravitation.

An arbitrary three-vector ξ^i [the space part of the four vector $\xi^\alpha = (\xi^0, \xi^i)$] can be written as the sum of two parts

$$\xi^i = \xi^i_{tr} + \gamma^{ij}\xi_{|j} , \tag{3.7}$$

where the function ξ is the solution of the equation

$$\xi^{|i}{}_{|i} = \xi^{i}{}_{|i} . \tag{3.8}$$

The "transverse" vector ξ^i_{tr} satisfies the condition

$$\xi^i_{tr|i} = 0 , \tag{3.9}$$

and contributes only to the vector-like metric perturbation. Hence, it is only the two functions ξ^0 and ξ which preserve the scalar nature of a metric perturbation. Note that the tensor part of a metric perturbation [see eq. (2.13)] is invariant under coordinate transformations (3.5).

We have seen that not all diffeomorphisms preserve the scalar nature of the metric fluctuations. The most general ones which do can be described in terms of only two independent functions ξ^0 and ξ,

$$\eta \to \tilde{\eta} = \eta + \xi^0(\eta, \boldsymbol{x}) , \qquad x^i \to \tilde{x}^i = x^i + \gamma^{ij}\xi_{|j}(\eta, \boldsymbol{x}) , \tag{3.10}$$

where the derivative of ξ is a covariant derivative with respect to the background space coordinates.

Clearly, the metric perturbation $\delta g_{\alpha\beta}$ is not invariant under this change of coordinates. The transformation (3.10) induces a change in $\delta g_{\alpha\beta}$,

$$\delta g_{\alpha\beta} \to \delta \tilde{g}_{\alpha\beta} = \delta g_{\alpha\beta} + \Delta g_{\alpha\beta} , \tag{3.11}$$

which can be rewritten as variations of the functions ϕ, ψ, B and E determining the perturbed metric [see eq. (2.9)]. One may verify that the new functions $\tilde{\phi}$, $\tilde{\psi}$, $\tilde{B}$ and $\tilde{E}$ are given by

$$\tilde{\phi} = \phi - (a'/a)\xi^0 - \xi^{0\prime} , \quad \tilde{\psi} = \psi + (a'/a)\xi^0 , \quad \tilde{B} = B + \xi^0 - \xi' , \quad \tilde{E} = E - \xi . \tag{3.12}$$

By taking combinations of ϕ, ψ, B and E we can construct gauge-invariant variables. The simplest gauge-invariant linear combinations of ϕ, ψ, B and E which span the two-dimensional space of gauge-invariant variables that can be constructed from metric variables alone are

$$\Phi = \phi + (1/a)[(B - E')a]' , \qquad \Psi = \psi - (a'/a)(B - E') . \tag{3.13}$$

The above variables were first introduced by Bardeen [10]. In his notation, $\Phi = \Phi_A$ and $\Psi = -\Phi_H$.

Of course, there are an infinite number of different gauge-invariant variables, since any combination of gauge-invariant variables will also be gauge invariant. The situation is similar to what is well known in gauge field theories. In gauge theories, the electric and magnetic fields take on a special role. Similarly, here the potentials Φ and Ψ play a special role since they are the simplest combinations of metric perturbations which are gauge invariant and since they satisfy simple equations of motion, as we shall show in the following chapter. The variables Φ and Ψ will play a crucial role in the rest of this review. It is important to stress that they are unchanged under all infinitesimal scalar coordinate transformations, but they are not invariant under finite coordinate changes.

As another useful example, let us consider a four-scalar $q(\eta, \boldsymbol{x})$ defined on the physical manifold $\mathcal{M}$. q can be split into its background value and a perturbation

$$q(\eta, \boldsymbol{x}) = q_0(\eta) + \delta q(\eta, \boldsymbol{x}) . \tag{3.14}$$

It is important to note that in general δq is not gauge invariant. Its change under the coordinate transformation (3.10) is given by

$$\delta q \rightarrow \widetilde{\delta q}(\eta, x) = \delta q(\eta, x) - q_0'(\eta)\xi^0 . \tag{3.15}$$

Only if q_0 is time-independent, then δq is gauge-invariant. Otherwise, the following combination of δq and metric perturbations is gauge-invariant:

$$\delta q^{(\mathrm{gi})} = \delta q + q_0'(B - E') . \tag{3.16}$$

The freedom of gauge choice can be used to impose two conditions on the four functions ϕ, ψ, B and E. This is possible since there are two functions ξ^0 and ξ which can be chosen appropriately. In picking a particular gauge, one has also specified in which coordinate system the perturbations are considered. In the following we shall consider two particular choices of gauge. First, the synchronous gauge, which is the one used most often in the literature. Second, the longitudinal (or conformal–Newtonian) gauge which can be used to verify the derivation in later sections [63, 64].

Synchronous gauge. Synchronous gauge is defined by the conditions $\phi = 0$ and $B = 0$. From (3.12) it follows that given any initial system of coordinates one can find a coordinate transformation to synchronous gauge. To do this, we set $\tilde{\phi} = 0$ and $\tilde{B} = 0$ in (3.12) and solve the equations for ξ^0 and ξ. The result is

$$\eta \rightarrow \eta_\mathrm{s} = \eta + a^{-1} \int a\phi \, \mathrm{d}\eta , \qquad x^i \rightarrow x^i_\mathrm{s} = x^i + \gamma^{ij}\Big(\int B \, \mathrm{d}\eta + \int a^{-1} \mathrm{d}\eta \int a\phi \, \mathrm{d}\eta\Big)_{|j} , \tag{3.17}$$

where a subscript s denotes synchronous gauge. However, as can be seen from (3.12) [or from (3.17)], the synchronous coordinates are not totally fixed. Under the residual transformation given by

$$\eta \rightarrow \tilde{\eta} + a^{-1}C_1(x) , \qquad x^i \rightarrow \tilde{x}^i + \gamma^{ij}C_{1|j}(x) \int a^{-1} \mathrm{d}\eta + \gamma^{ij}C_{2|j}(x) , \tag{3.18}$$

where $C_1(x)$ and $C_2(x)$ are arbitrary functions of the spatial variables [they are integration constants in (3.17)], the synchronous-gauge conditions are maintained. This residual coordinate freedom in synchronous gauge leads to the appearance of unphysical gauge modes which render the interpretation of synchronous gauge calculations difficult, especially on scales larger than the Hubble radius (see also ref. [65]).

Longitudinal gauge. Longitudinal gauge is defined by the conditions $B = E = 0$. From (3.12), it follows that in longitudinal gauge the coordinates are totally fixed. The condition $E = 0$ determines ξ uniquely, and using this result, the condition $B = 0$ uniquely fixes ξ^0. Hence, in longitudinal gauge there are no complicating residual gauge modes. Starting from any system of coordinates (η, x^i), the longitudinal gauge conditions can be implemented by a coordinate transformation, namely by

$$\eta \rightarrow \eta_\mathrm{l} = \eta - (B - E') , \qquad x^i \rightarrow x^i_\mathrm{l} = x^i + \gamma^{ij}E_{|j} . \tag{3.19}$$

Under this transformation, the metric variables change as follows:

$$\phi \to \phi_l = \phi + a^{-1}[a(B - E')]' = \Phi , \quad \psi \to \psi_l = \psi - (a'/a)(B - E') = \Psi ,$$
$$B \to B_l = 0 , \quad E \to E_l = 0 , \tag{3.20}$$

We draw the important conclusion that in longitudinal gauge, ϕ and ψ coincide with the gauge-invariant variables Φ and Ψ respectively. This fact can be used to propose an elegant derivation of the equations of motion for Φ and Ψ. We go to longitudinal gauge and derive the equations of motion for ϕ_l and ψ_l which can then be abstracted to give the gauge-invariant equations of motion for Φ and Ψ if we replace ϕ_l and ψ_l by Φ and Ψ respectively.

In longitudinal gauge, the metric takes the form

$$ds^2 = a^2(\eta)[(1 + 2\Phi)\, d\eta^2 - (1 - 2\Psi)\gamma_{ij}\, dx^i\, dx^j] . \tag{3.21}$$

In the case when the spatial part of the energy–momentum tensor is diagonal, i.e. $\delta T^i_j \sim \delta^i_j$, it follows (see chapter 4) that $\phi_l = \psi_l$ or $\Phi = \Psi$. There remains only one free metric perturbation variable which is a generalization of the Newtonian gravitational potential ϕ. This explains the choice of the name "conformal–Newtonian" for this system. It is also the reason for our notation. As can be seen from (3.21), the gauge invariants Φ and Ψ have a very simple physical interpretation: they are the amplitudes of the metric perturbations in the conformal–Newtonian (or longitudinal) coordinate system.

As the final point of this chapter, we shall write down the formulas which connect synchronous and longitudinal coordinate systems. The coordinate transformation which leads from synchronous to longitudinal gauge is obtained from (3.19) upon inserting $\phi = B = 0$,

$$\eta_l = \eta_s + E_s' , \qquad x_l^i = x_s^i + \gamma^{ij} E_{s|j} , \tag{3.22}$$

where E_S denotes the metric variable E in synchronous gauge. The metric variables are connected as follows:

$$\phi_l = -(a'/a)E_s' - E_s'' , \qquad \psi_l = \psi_s + (a'/a)E_s' , \tag{3.23}$$

and the energy-density perturbations $\delta\varepsilon_s$ and $\delta\varepsilon_l$ are related [see eqs. (3.15) and (3.22)] via

$$\delta\varepsilon_l = \delta\varepsilon_s - \varepsilon_0' E_s' . \tag{3.24}$$

Conversely, given longitudinal coordinates one can obtain synchronous coordinates by the transformation [see eq. (3.17)]

$$\eta_s = \eta_l + a^{-1} \int a(\eta)\phi_l\, d\eta , \qquad x_s^i = x_l^i + \gamma^{ij}\left(\int d\eta\, a^{-1}(\eta) \int d\eta\, a(\eta)\phi_l\right)_{|j} , \tag{3.25}$$

under which the metric variables and energy-density perturbation are connected by

$$\psi_s = \psi_l + \frac{a'}{a^2} \int a\phi_l\, d\eta , \qquad E_s = -\int d\eta\, a^{-1}(\eta) \int d\eta\, a\phi_l , \qquad \delta\varepsilon_s = \delta\varepsilon_l - \varepsilon_0' a^{-1} \int d\eta\, a\phi_l . \tag{3.26}$$

Equations (3.23) and (3.26) follow immediately from (3.12) and (3.16).

4. General form of the equations for cosmological perturbations

In this section we shall derive the general from of the equations which describe small cosmological perturbations. To do that, we start with the Einstein equations which relate the general distribution of energy and momentum given by the stress tensor T^{μ}_{ν} of matter to the geometrical properties of space–time,

$$G^{\mu}{}_{\nu} = 8\pi G T^{\mu}{}_{\nu} , \qquad G^{\mu}{}_{\nu} = R^{\mu}{}_{\nu} - \tfrac{1}{2}\delta^{\mu}{}_{\nu} R , \tag{4.1a, b}$$

where $G^{\mu}{}_{\nu}$ is the Einstein tensor and $R^{\mu}{}_{\nu}$ is the Ricci tensor. $R = R^{\mu}_{\mu}$ is the scalar curvature.

For the background model with metric (2.1) describing a homogeneous and isotropic expanding universe, the Einstein tensor is

$${}^{(0)}G^{0}{}_{0} = 3a^{-2}(\mathscr{H}^2 + \mathscr{K}) , \qquad {}^{(0)}G^{0}{}_{i} = 0 , \qquad {}^{(0)}G^{i}{}_{j} = a^{-2}(2\mathscr{H}' + \mathscr{H}^2 + \mathscr{K})\delta^{i}{}_{j} , \tag{4.2}$$

where $\mathscr{H} = a'/a$ using conformal time. The form of ${}^{(0)}G^{\mu}{}_{\nu}$ using physical time can be obtained by inserting the change of variables $t = t(\eta)$ into the above tensor elements for conformal time. Then, the background equations are

$${}^{(0)}G^{\mu}{}_{\nu} = 8\pi G\, {}^{(0)}T^{\mu}{}_{\nu} , \tag{4.3}$$

where ${}^{(0)}T^{\mu}{}_{\nu}$ is the background energy–momentum tensor. In order to satisfy these equations, ${}^{(0)}T^{\mu}{}_{\nu}$ must satisfy the following symmetry properties:

$${}^{(0)}T^{i}{}_{0} = {}^{(0)}T^{0}{}_{i} = 0 , \qquad {}^{(0)}T^{i}{}_{j} \propto \delta^{i}{}_{j} . \tag{4.4}$$

For a metric with small perturbations, the Einstein tensor can be written as

$$G^{\mu}{}_{\nu} = {}^{(0)}G^{\mu}{}_{\nu} + \delta G^{\mu}{}_{\nu} + \cdots , \tag{4.5}$$

and the energy–momentum tensor can be split in a similar way. The equations of motion for small perturbations linearized about the background metric are

$$\delta G^{\mu}{}_{\nu} = 8\pi G\, \delta T^{\mu}{}_{\nu} . \tag{4.6}$$

where δ denotes the terms linear in metric and matter fluctuations.

For scalar type metric perturbations with a line element given in (2.10) (in conformal time), the perturbed Einstein equations can be obtained as a result of straightforward but rather tedious calculations,

$$\begin{aligned} \delta G^{0}{}_{0} &= 2a^{-2}\{-3\mathscr{H}(\mathscr{H}\phi + \psi') + \nabla^2[\psi - \mathscr{H}(B - E')] + 3\mathscr{K}\psi\} = 8\pi G\, \delta T^{0}{}_{0} , \\ \delta G^{0}{}_{i} &= 2a^{-2}[\mathscr{H}\phi + \psi' - \mathscr{K}(B - E')]_{|i} = 8\pi G\, \delta T^{0}{}_{i} , \end{aligned} \tag{4.7}$$

$$\begin{aligned} \delta G^{i}{}_{j} &= -2a^{-2}\{[(2\mathscr{H}' + \mathscr{H}^2)\phi + \mathscr{H}\phi' + \psi'' + 2\mathscr{H}\psi' - \mathscr{K}\psi + \tfrac{1}{2}\nabla^2 D]\delta^{i}{}_{j} - \tfrac{1}{2}D_{|ij}\} = 8\pi G\, \delta T^{i}{}_{j} , \\ D &= (\phi - \psi) + 2\mathscr{H}(B - E') + (B - E')' . \end{aligned} \tag{4.8}$$

Note that if we use physical time t instead of conformal time η, all but the δG^0_i and δG^i_0 terms can be obtained immediately by changing the time variable $t = t(\eta)$ in the above tensor elements. For δG^0_i and δG^i_0 there are extra factors of a,

$$\delta^{(t)} G^0{}_i(t) = a(t)\, \delta^{(\eta)} G^0{}_i(\eta(t)) . \tag{4.9}$$

$\delta G^\mu{}_\nu$ is not gauge invariant. Hence, the right- and left-hand sides of the perturbed Einstein equations are not separately invariant under gauge transformations. Under the transformation (3.10), we find from (3.6) that $\delta G^\mu{}_\nu$ transforms in the following manner:

$$\delta G^0_0 \to \delta G^0_0 - ({}^{(0)}G^0_0)'\xi^0 , \qquad \delta G^0_i \to \delta G^0_i - ({}^{(0)}G^0_0 - \tfrac{1}{3}{}^{(0)}G^k_k)\xi^0_{|i} , \qquad \delta G^i_j \to \delta G^i_j - ({}^{(0)}G^i_j)'\xi^0 . \tag{4.10}$$

The same kind of transformation law holds for δT^μ_ν if we take into account (4.3).

We may rewrite $\delta G^\mu{}_\nu$ in terms of the gauge-invariant variables Φ and Ψ by substituting ϕ and ψ in terms of Φ, Ψ and $(B - E')$ using (3.13) to get

$$\begin{aligned}
\delta G^0{}_0 &= 2a^{-2}[-3\mathcal{H}(\mathcal{H}\Phi + \Psi') + \nabla^2\Psi + 3\mathcal{K}\Psi + 3\mathcal{H}(-\mathcal{H}' + \mathcal{H}^2 + \mathcal{K})(B - E')] , \\
\delta G^0{}_i &= 2a^{-2}[\mathcal{H}\Phi + \Psi' + (\mathcal{H}' - \mathcal{H}^2 - \mathcal{K})(B - E')]_{,i} , \\
\delta G^i{}_j &= -2a^{-2}\{[(2\mathcal{H}' + \mathcal{H}^2)\Phi + \mathcal{H}\Phi' + \Psi'' + 2\mathcal{H}\Psi' - \mathcal{K}\Psi + \tfrac{1}{2}\nabla^2 D]\delta^i{}_j \\
&\quad + (\mathcal{H}'' - \mathcal{H}\mathcal{H}' - \mathcal{H}^3 - \mathcal{K}\mathcal{H})(B - E')\delta^i{}_j - \tfrac{1}{2}\gamma^{ik}D_{|kj}\} ,
\end{aligned} \tag{4.11}$$

where $D = \Phi - \Psi$. It is easy to construct the gauge-invariant variables $\delta G^{(\mathrm{gi})\alpha}_\beta$ and $\delta T^{(\mathrm{gi})\alpha}_\beta$ corresponding to δG^α_β and δT^α_β,

$$\begin{aligned}
&\delta G^{(\mathrm{gi})0}_0 = \delta G^0_0 + ({}^{(0)}G^0_0)'(B - E') , \qquad \delta G^{(\mathrm{gi})0}_i = \delta G^0_i + ({}^{(0)}G^0_0 - \tfrac{1}{3}{}^{(0)}G^k_k)(B - E')_{|i} , \\
&\delta G^{(\mathrm{gi})i}_j = \delta G^i_j + ({}^{(0)}G^i_j)'(B - E') ,
\end{aligned} \tag{4.12}$$

and analogously for δT^μ_ν,

$$\begin{aligned}
&\delta T^{(\mathrm{gi})0}_0 = \delta T^0_0 + ({}^{(0)}T^0_0)'(B - E') , \qquad \delta T^{(\mathrm{gi})0}_i = \delta T^0_i + ({}^{(0)}T^0_0 - \tfrac{1}{3}{}^{(0)}T^k_k)(B - E')_{|i} , \\
&\delta T^{(\mathrm{gi})i}_j = \delta T^i_j + ({}^{(0)}T^i_j)'(B - E') .
\end{aligned} \tag{4.13}$$

Using the background equations of motion, eq. (4.6) for small perturbations may be written in the following form:

$$\delta G^{(\mathrm{gi})\mu}_\nu = 8\pi G\, \delta T^{(\mathrm{gi})\mu}_\nu . \tag{4.14}$$

Both sides are now explicitly gauge-invariant.

The left-hand side of the above equation can be expressed in terms of the gauge-invariant potentials Φ and Ψ alone. Taking into account (4.11) and (4.12), we find that all terms on the left-hand side of (4.14) which include B and E cancel. Thus, from (4.14) we obtain the following general form of the gauge-invariant equations for cosmological perturbations (in conformal time):

$$-3\mathcal{H}(\mathcal{H}\Phi + \Psi') + \nabla^2\Psi + 3\mathcal{K}\Psi = 4\pi G a^2\,\delta T_0^{(\mathrm{gi})0}\,, \qquad (\mathcal{H}\Phi + \Psi')_{,i} = 4\pi G a^2\,\delta T_i^{(\mathrm{gi})0}\,,$$
$$[(2\mathcal{H}' + \mathcal{H}^2)\Phi + \mathcal{H}\Phi' + \Psi'' + 2\mathcal{H}\Psi' - \mathcal{K}\Psi + \tfrac{1}{2}\nabla^2 D]\delta^i_{\ j} - \tfrac{1}{2}\gamma^{ik}D_{|kj} = -4\pi G a^2\,\delta T_j^{(\mathrm{gi})i}\,, \tag{4.15}$$

where $D = \Phi - \Psi$ as before, and the $\delta T_j^{(\mathrm{gi})i}$ are given by (4.13).

In the above derivation, we have followed the Lifshitz [3] coordinate perturbation approach which we consider to be technically the most straightforward. However, there are other derivations which should be mentioned at this point. One alternative method is the fluid-flow approach [5, 6] in which one follows the total energy density along comoving world lines and analyzes its perturbations. This approach has been used by Lyth and collaborators [31, 32] to derive the gauge-invariant equations of motion. Very closely related to this is the covariant formulation which has recently been developed by Bruni, Ellis, Hwang and Vishniac [30, 68] and which has been applied to inflationary universe models [69] and generalized gravity theories [70]. The most geometrical method is based on the ADM formulation of general relativity and has been used by Durrer and Straumann [34, 66] to rederive the gauge-invariant equations. Recently, this formulation has been extended [67] to cover models with seed perturbations such as cosmic strings and textures.

To close the system of equations, we need equations of motion for the matter which are also formulated in a gauge-invariant way. Examples of perfect fluid and scalar field matter will be discussed in chapters 5 and 6. Another example (treated in refs. [71, 66]) is collisionless matter.

5. Hydrodynamical perturbations

According to the Big Bang theory, the universe was dominated by radiation at early times. Since the energy density in radiation decays faster than that in matter, at some time (called the time of equal matter and radiation), the universe becomes matter-dominated. In this chapter, we derive the growth rates of cosmological perturbations in models with conventional hydrodynamical matter. First, we study the evolution during the radiation- and matter-dominated phases, then we evolve the perturbations in the transition region.

First, we discuss the general form of the energy–momentum tensor and illustrate that adiabatic and entropy perturbations evolve in quite different ways. Next, we write down the equations of motion for the perturbations, focusing attention on their gauge-invariant form. The bulk of this chapter consists of discussions of solutions [29] of the perturbation equations in the case of adiabatic perturbations (section 5.3) and entropy perturbations (section 5.4). (See also ref. [72] for a related gauge-invariant approach to hydrodynamical perturbations.)

Sections 5.3 and 5.4 contain the detailed derivations of the solutions for the gauge-invariant perturbation variables. Those readers that are not interested in the technical details of the derivations but are only concerned with the asymptotic behavior or the physical interpretation of the results may refer to the summary part at the end of the section.

5.1. Energy–momentum tensor

Restricting our attention to scalar perturbations, we can express the most general first-order perturbation of the energy–momentum tensor in fluid notation in terms of four scalar functions $\delta\varepsilon$, δp, $\mathcal{V}$ and σ, all of which depend on space–time variables. Here, $\delta\varepsilon$ and δp are the perturbed energy density and pressure respectively, $\mathcal{V}$ is the potential for the three-velocity field $u^i(\boldsymbol{x}, t)$, and σ determines the anisotropic stress. The perturbed stress-energy tensor has the form [10]

$$\delta T^{\mu}{}_{\nu} = \begin{pmatrix} \delta\varepsilon & -(\varepsilon_0 + p_0)a^{-1}\mathcal{V}_{,i} \\ (\varepsilon_0 + p_0)a\mathcal{V}_{,i} & -\delta p\,\delta_{ij} + \sigma_{|ij} \end{pmatrix} . \tag{5.1}$$

In the following we will consider a perfect fluid for which the energy–momentum tensor can be described in terms of only three functions, energy density ε, pressure p and four-fluid velocity u^{α},

$$T^{\alpha}_{\beta} = (\varepsilon + p)u^{\alpha}u_{\beta} - p\delta^{\alpha}_{\beta} . \tag{5.2}$$

Thus, for a perfect fluid the anisotropic stress which leads to nondiagonal space–space components of the energy–momentum tensor vanishes. In addition, possible dissipation has been neglected in (5.2). This is justified for many problems under consideration.

In general, the pressure p depends not only on the energy density ε, but also on the entropy per baryon ratio S. Given $p(\varepsilon, S)$, the pressure fluctuation δp can be expressed in terms of the energy density and entropy perturbation $\delta\varepsilon$ and δS,

$$\delta p = (\partial p/\partial\varepsilon)|_S\,\delta\varepsilon + (\partial p/\partial S)|_{\varepsilon}\,\delta S \equiv c_s^2\,\delta\varepsilon + \tau\,\delta S . \tag{5.3}$$

For hydrodynamical matter, c_s can be interpreted as the sound velocity.

In a single component ideal gas there are no entropy perturbations. However, in the universe there are at least two components: plasma and radiation. Hence, entropy perturbations may be important. At late times, when the temperature T is low compared to the masses of the baryons, the pressure of the baryons is negligible ($p_m = 0$), and the total pressure is given by the radiation

$$p = p_r = \tfrac{1}{3}\varepsilon_r \tag{5.4}$$

where ε_r is the energy density in radiation. Hence

$$\delta p = \tfrac{1}{3}\delta\varepsilon_r . \tag{5.5}$$

Since the entropy per baryon is proportional to T^3/n_b, where n_b is the number density of baryons, T is the temperature, and $\varepsilon_r \propto T^4$, the entropy perturbation can be rewritten in the form

$$\delta S/S = \tfrac{3}{4}\,\delta\varepsilon_r/\varepsilon_r - \delta\varepsilon_m/\varepsilon_m . \tag{5.6}$$

Here, $\varepsilon_m \propto n_b$ is the energy density in baryons (matter). Taking into account that

$$\delta\varepsilon = \delta\varepsilon_r + \delta\varepsilon_m , \tag{5.7}$$

and using this relation to express $\delta\varepsilon_{\mathrm{m}}$ in terms of $\delta\varepsilon$ and $\delta\varepsilon_{\mathrm{r}}$ in (5.6), we can solve (5.6) to find $\delta\varepsilon_{\mathrm{r}}$ in terms of $\delta S/S$ and $\delta\varepsilon$. Inserting into (5.5), one obtains

$$\delta p = \tfrac{1}{3}(1 + \tfrac{3}{4}\varepsilon_{\mathrm{m}}/\varepsilon_{\mathrm{r}})^{-1}\,\delta\varepsilon + \tfrac{1}{3}\varepsilon_{\mathrm{m}}(1 + \tfrac{3}{4}\varepsilon_{\mathrm{m}}/\varepsilon_{\mathrm{r}})^{-1}\,\delta S/S\,. \tag{5.8}$$

Hence, comparing with (5.3), we can read off the expressions for c_{s}^2 and τ,

$$c_{\mathrm{s}}^2 = \tfrac{1}{3}(1 + \tfrac{3}{4}\varepsilon_{\mathrm{m}}/\varepsilon_{\mathrm{r}})^{-1}\,, \qquad \tau = c_{\mathrm{s}}^2\varepsilon_{\mathrm{m}}/S\,. \tag{5.9a, b}$$

When applied to the early universe, the above model describes the smooth transition from the radiation-dominated period ($\varepsilon_{\mathrm{r}} \gg \varepsilon_{\mathrm{m}}$) with $c_{\mathrm{s}}^2 = \frac{1}{3}$ to the matter-dominated epoch ($\varepsilon_{\mathrm{m}} \gg \varepsilon_{\mathrm{r}}$) with $c_{\mathrm{s}}^2 = 0$.

The first-order perturbation of the energy–momentum tensor, $\delta T^{\alpha}_{\beta}$, can be expressed in terms of $\delta\varepsilon, \delta p$ and δu^i, the velocity perturbation of the fluid. In conformal time we obtain,

$$\delta T^0_0 = \delta\varepsilon\,, \qquad \delta T^0_i = (\varepsilon_0 + p_0)a^{-1}\,\delta u_i\,, \qquad \delta T^i_j = -\delta p\,\delta^i_j\,, \tag{5.10}$$

where ε_0 and p_0 are the background values of ε and p. In the above equations, δp must be written in terms of $\delta\varepsilon$ and δS using (5.8).

Now we shall calculate the gauge-invariant perturbations of the energy–momentum tensor $\delta T^{(\mathrm{gi})\alpha}_{\beta}$. They can be expressed in terms of the gauge-invariant energy density, pressure and velocity perturbations. The gauge-invariant energy density and pressure perturbations $\delta\varepsilon^{(\mathrm{gi})}$ and $\delta p^{(\mathrm{gi})}$ are defined in the same way as the gauge-invariant perturbation of a general four-scalar was in (3.16),

$$\delta\varepsilon^{(\mathrm{gi})} = \delta\varepsilon + \varepsilon_0'(B - E')\,, \qquad \delta p^{(\mathrm{gi})} = \delta p + p_0'(B - E')\,. \tag{5.11}$$

The gauge-invariant three-velocity $\delta u_i^{(\mathrm{gi})}$ is given by

$$\delta u_i^{(\mathrm{gi})} = \delta u_i + a(B - E')_{|i}\,. \tag{5.12}$$

[One may easily verify that $\delta u_i^{(\mathrm{gi})}$ is in fact invariant under transformations of type (3.10).]

Now, using (3.16) we immediately obtain

$$\delta T^{(\mathrm{gi})0}_0 = \delta\varepsilon^{(\mathrm{gi})}\,, \qquad \delta T^{(\mathrm{gi})0}_i = (\varepsilon_0 + p_0)a^{-1}\,\delta u_i^{(\mathrm{gi})}\,, \qquad \delta T^{(\mathrm{gi})i}_j = -\delta p^{(\mathrm{gi})}\,\delta^i_j\,. \tag{5.13}$$

These are the quantities which enter the gauge-invariant equations. Note that the physical meaning of $\delta\varepsilon^{(\mathrm{gi})}, \delta p^{(\mathrm{gi})}$ and $\delta u_i^{(\mathrm{gi})}$ is very simple: they coincide, respectively, with the perturbations of energy density, pressure and velocity in longitudinal gauge.

5.2. Equations of motion

Based on our general considerations in chapter 4 and taking into account (5.13), we can immediately write down the gauge-invariant equations of motion for hydrodynamical perturbations,

$$-3\mathscr{H}(\mathscr{H}\Phi + \Psi') + \nabla^2\Psi + 3\mathscr{K}\Psi = 4\pi G a^2\,\delta\varepsilon^{(\mathrm{gi})}\,. \tag{5.14}$$

$$(\mathcal{H}\Phi + \Psi')_{,i} = 4\pi Ga(\varepsilon_0 + p_0)\,\delta u_i^{(\mathrm{gi})}\,, \tag{5.15}$$

$$[(2\mathcal{H}' + \mathcal{H}^2)\Phi + \mathcal{H}\Phi' + \Psi'' + 2\mathcal{H}\Psi' - \mathcal{K}\Psi + \tfrac{1}{2}\nabla^2 D]\delta^i_j - \tfrac{1}{2}\gamma^{ik}D_{|kj} = 4\pi Ga^2\,\delta p^{(\mathrm{gi})}\,\delta^i_j\,, \tag{5.16}$$

where $D = \Phi - \Psi$.

The absence of nondiagonal space–space components in the energy–momentum tensor leads to a significant simplification of this set of equations. It follows from the $i \neq j$ equation that $\Phi = \Psi$. One can show this using invariance arguments and the fact that the spatial average of a perturbation must vanish. Now, from (5.16), all mixed spatial derivatives of $f = \Phi - \Psi$ vanish. Hence, $f(x_1, x_2, x_3)$ must be a sum of three functions $f_i(x_j)$. The only way that this structure can be preserved under coordinate transformations is for f_i to be linear functions. Since the spatial average must vanish, the only possibility is for f_i to vanish. In a flat universe, there is a simpler way to reach this conclusion; we go to momentum space. The $i \neq j$ equation must hold for each mode separately, and the only way for this to happen is if all Fourier coefficients vanish. With the above identification, the perturbation equations become

$$\nabla^2\Phi - 3\mathcal{H}\Phi' - 3(\mathcal{H}^2 - \mathcal{K})\Phi = 4\pi Ga^2\,\delta\varepsilon^{(\mathrm{gi})}\,, \tag{5.17}$$

$$(a\Phi)'_{,i} = 4\pi Ga^2(\varepsilon_0 + p_0)\,\delta u_i^{(\mathrm{gi})}\,, \tag{5.18}$$

$$\Phi'' + 3\mathcal{H}\Phi' + (2\mathcal{H}' + \mathcal{H}^2 - \mathcal{K})\Phi = 4\pi Ga^2\,\delta p^{(\mathrm{gi})}\,. \tag{5.19}$$

In the Newtonian limit, (5.17) is the usual Poisson equation for the gravitational potential induced by some energy-density perturbation. This supports the interpretation of Φ as the relativistic generalization of the Newtonian gravitational potential ϕ. Equation (5.17) generalizes the Poisson equation by taking into account the expansion of the universe. Since the equation is similar to the heat equation, it is possible to find its Green function and hence also its general solution if we consider the right-hand side as a given source function. Note that (5.17) is true whenever linear perturbation theory is valid. This requires $|\Phi| \ll 1$ but not necessarily $|\delta\varepsilon/\varepsilon| \ll 1$ (see, e.g., ref. [72]). On scales larger than the Hubble radius, $|\Phi| \ll 1$ may be true even if $|\delta\varepsilon/\varepsilon| > 1$.

It is interesting to compare the gauge-invariant equations for hydrodynamical perturbations with the synchronous-gauge equations. To obtain these equations, we start from the general gauge-dependent Einstein equations (4.6) for first-order perturbations and set $\phi = B = 0$. The result is

$$\begin{gathered}\nabla^2(\psi_s + \mathcal{H}E'_s) - 3\mathcal{H}\psi'_s = 4\pi Ga^2\,\delta\varepsilon_s\,, \qquad \psi'_{s,i} = 4\pi Ga^3(\varepsilon_0 + p_0)\,\delta u_{si}\,, \\ \psi''_s + 2\mathcal{H}\psi'_s + \tfrac{1}{2}(\nabla^2 D_s - D_{s|ii}) = 4\pi G\,\delta p_s\,, \qquad D_{s|ij} = 0\,, \quad i \neq j\,,\end{gathered} \tag{5.20}$$

where the subscripts s stand for synchronous gauge. In this case, D_s can be read off from (4.7)

$$D_s = -\psi_s - E''_s - 2\mathcal{H}E'_s\,. \tag{5.21}$$

Thus, the $i \neq j$ equation allows us to express ψ_s in terms of E_s. After this simplification, the perturbation equations can be expressed in terms of a single metric perturbation $E_s(\boldsymbol{x}, \eta)$ and its derivatives. Unfortunately, the equation contains [36] up to third time derivatives of E_s, which renders it hard to solve. This mathematical problem is a reflection of the fact that synchronous gauge has a residual gauge freedom.

Thus, the equations for cosmological perturbations in synchronous gauge are manifestly more complicated than the gauge-invariant equations. This gives an additional practical motivation for choosing the gauge-invariant approach besides its intrinsic advantages related to the physical interpretation.

Returning to the general development of the theory of gauge-invariant hydrodynamical perturbations, we see that using relation (5.3) which expresses δp in terms of energy density and entropy perturbations, (5.17) and (5.19) can be combined to give

$$\Phi'' + 3\mathcal{H}(1 + c_s^2)\Phi' - c_s^2 \nabla^2\Phi + [2\mathcal{H}' + (1 + 3c_s^2)(\mathcal{H}^2 - \mathcal{K})]\Phi = 4\pi G a^2 \tau\, \delta S\,. \tag{5.22}$$

In particular, in the case of pure adiabatic perturbations, the source term in the above second-order partial differential equation vanishes. For readers familiar with Bardeen's article [10] we mention that the above equation (5.22) agrees with eq. (5.30) in ref. [10].

The equation of motion (5.22) can be recast into a very convenient form by introducing the gauge-invariant function

$$\zeta = \tfrac{2}{3}(H^{-1}\dot{\Phi} + \Phi)/(1 + w) + \Phi\,. \tag{5.23}$$

It can be verified that the equation $\dot{\zeta} = 0$ is equivalent to (5.22) as long as only adiabatic perturbations ($\delta S = 0$) are considered on scales larger than the Hubble radius (when $c_s^2 \nabla^2\Phi$ is negligible). Thus, under these conditions, the equation of motion for Φ becomes a very useful conservation law. A variable which differs from ζ only by terms of the order $(k/aH)^2$ was first introduced in ref. [16] and was discussed in detail in refs. [55, 56]. The above form of ζ was given by Lyth in ref. [18], who also pointed out that in comoving gauge ζ takes on the physical meaning of a curvature perturbation.

The friction term proportional to Φ' can be eliminated by the following change of variables:

$$\Phi = 4\pi G(\varepsilon_0 + p_0)^{1/2} u = (4\pi G)^{1/2}[(\mathcal{H}^2 - \mathcal{H}' + \mathcal{K})/a^2]^{1/2} u \tag{5.24}$$

(where the second equality sign follows from the FRW background equations). After some tedious calculations using (5.9a) and the background equations for hydrodynamical perturbations, the equation of motion for u can be obtained in the form

$$u'' - c_s^2 \nabla^2 u - (\theta''/\theta) u = \mathcal{N}\,, \tag{5.25}$$

$$\theta = (\mathcal{H}/a)[\tfrac{2}{3}(\mathcal{H}^2 - \mathcal{H}' + \mathcal{K})]^{-1/2} = \frac{1}{a}\left(\frac{\varepsilon_0}{\varepsilon_0 + p_0}\right)^{1/2}\left(1 - \frac{3\mathcal{K}}{8\pi G a^2 \varepsilon_0}\right)^{1/2}, \tag{5.26}$$

$$\mathcal{N} = (4\pi G)^{1/2} a^3 (\mathcal{H}^2 - \mathcal{H}' + \mathcal{K})^{-1/2} \tau\, \delta S = a^2(\varepsilon_0 + p_0)^{-1/2} \tau\, \delta S\,. \tag{5.27}$$

Equations (5.17) and (5.18) give

$$\delta\varepsilon^{(\mathrm{gi})}/\varepsilon_0 = 2[3(\mathcal{H}^2 + \mathcal{K})]^{-1}[\nabla^2\Phi - 3\mathcal{H}\Phi' - 3(\mathcal{H}^2 - \mathcal{K})\Phi]\,, \tag{5.28}$$

$$\delta u_i^{(\mathrm{gi})} = -a^{-2}(\mathcal{H}^2 - \mathcal{H}' + \mathcal{K})^{-1}(a\Phi)'_{,i} . \tag{5.29}$$

Equation (5.22) or equivalently (5.25) determine the evolution of perturbations in a hydrodynamical universe. They are the main result of this section.

In the following we will separately investigate the evolution of adiabatic and entropy perturbations in a hydrodynamical universe.

5.3. Adiabatic perturbations

For adiabatic perturbations, the source term in (5.22) or (5.25) vanishes and the equations become homogeneous. First, the solutions for a matter-dominated universe ($p = 0$) will be considered. Then, the equations in a radiation-dominated universe ($p = \frac{1}{3}$) and in a universe containing both matter and radiation will be analyzed. Finally, a summary of the results will be presented. Those readers who are only interested in the asymptotic behavior and the physical interpretation of the results may skip the following and go directly to the summary at the end of the section.

Adiabatic perturbations in a universe with hydrodynamical matter have been studied by many authors using both the synchronous-gauge framework (see, e.g., refs. [3, 5–7, 2]) and the gauge-invariant formulation (see, e.g., ref. [10]).

We first consider a cold *matter-dominated universe* ($p = 0$). The background equations of motion for $a(\eta)$ are given in eqs. (2.4a) and (2.5a). Their solution is

$$a(\eta) = a_{\mathrm{m}} \begin{cases} \cosh\eta - 1 & \mathcal{K} = -1 , \\ \eta^2/2 & \mathcal{K} = 0 , \\ 1 - \cos\eta & \mathcal{K} = 1 . \end{cases} \tag{5.30}$$

Since in a flat universe $\eta(t) \sim t^{1/3}$, (5.30) in this case implies $a(t) \sim t^{2/3}$, the well known scaling for $p = 0$. Here, a_{m} is a constant.

Since $p = 0$, $c_s^2 = 0$ and (5.25) becomes a differential equation which does not include space derivatives. The most general exact solution of this equation is

$$u(\boldsymbol{x}, \eta) = C_1'(\boldsymbol{x})\theta(\eta) + C_2'(\boldsymbol{x})\theta(\eta) \int \frac{\mathrm{d}\eta'}{\theta^2} , \tag{5.31}$$

where $C_1'(\boldsymbol{x})$ and $C_2'(\boldsymbol{x})$ are arbitrary functions of the spatial coordinates. Evaluating (5.31) for a flat universe we obtain

$$\Phi(\boldsymbol{x}, \eta) = C_1(\boldsymbol{x}) + C_2(\boldsymbol{x})\eta^{-5} , \tag{5.32}$$

where $C_1(\boldsymbol{x})$ and $C_2(\boldsymbol{x})$ are arbitrary functions of the spatial coordinates proportional to $C_2'(\boldsymbol{x})$ and $C_1'(\boldsymbol{x})$ respectively. From (5.28) and (5.29) we can find the gauge-invariant density and velocity perturbations,

$$\begin{aligned} \delta\varepsilon^{(\mathrm{gi})}/\varepsilon_0 &= \tfrac{1}{6}[(\nabla^2 C_1 \eta^2 - 12 C_1) + (\nabla^2 C_2 \eta^2 + 18 C_2)\eta^{-5}] , \\ \delta u_i^{(\mathrm{gi})} &= (1/a_{\mathrm{m}})(2C_{2,i}/\eta^6 - \tfrac{4}{3} C_{1,i}/\eta) . \end{aligned} \tag{5.33}$$

The important lesson to draw from these solutions is that (neglecting the decaying mode) the gravitational potential Φ remains constant. On length scales smaller than the Hubble radius, the energy-density perturbation increases as η^2 or $t^{2/3}$, whereas for scales larger than the Hubble radius, also the energy-density perturbation remains constant [since the spatial derivative terms in (5.33) are suppressed]. The velocity perturbation decreases as η^{-1} or $t^{-1/3}$.

The above results for the time evolution of adiabatic perturbations on scales smaller than the Hubble radius are well known and agree with the conclusions of previous analyses. A nice feature of the gauge-invariant approach considered here is that the results on both large and small scales emerge in a unified and fairly simple treatment.

In the case of an open universe ($\mathcal{K} = -1$), (5.33) can also be explicitly integrated to yield

$$\Phi(x, \eta) = C_1(x)\, \frac{2 \sinh^2\eta - 6\eta \sinh \eta + 8 \cosh \eta - 8}{(\cosh \eta - 1)^3} + C_2(x)\, \frac{\sinh \eta}{(\cosh \eta - 1)^3}\,, \tag{5.34}$$

where $C_1(x)$ and $C_2(x)$ are arbitrary functions of the spatial coordinates satisfying the constraints

$$\int C_i(x) \sqrt{\gamma}\, d^3x = 0\,, \quad i = 1, 2\,. \tag{5.35}$$

The expressions for gauge-invariant energy-density and velocity perturbations again follow from (5.28) and (5.29),

$$\begin{aligned} \delta\varepsilon^{(\mathrm{gi})}/\varepsilon_0 &= \tfrac{1}{3}[(\cosh \eta - 1)\nabla^2\Phi + 9\Phi - 6C_1]\,, \\ \delta u_i^{(\mathrm{gi})} &= -\frac{2}{3a_\mathrm{m}}\left(\coth \eta\, \Phi_{,i} + \frac{2}{\sinh \eta}\, C_{1,i}\right). \end{aligned} \tag{5.36}$$

The solutions for a closed universe ($\mathcal{K} = 1$) can be obtained by replacing η with $\mathrm{i}\eta$ in (5.34) and (5.36).

From the above solutions we can see an important advantage of the gauge-invariant formalism over the synchronous-gauge approach. In synchronous gauge the metric perturbations h_{ij} are of order 1 when $\delta\varepsilon/\varepsilon \sim 1$ whereas the gauge-invariant variable

$$\Phi \sim (\lambda/\eta)^2 \ll 1 \quad \text{when} \quad \delta\varepsilon/\varepsilon \sim 1\,, \tag{5.37}$$

on length scales λ much smaller than the Hubble radius. Thus, the linearized 0–0 Einstein equation can be extended further in time than it can in synchronous gauge.

From (5.34) and (5.36) we conclude that for an open universe the density perturbations are frozen in for $\eta \gg 1$, namely

$$\delta\varepsilon^{(\mathrm{gi})}/\varepsilon_0 \sim \tfrac{2}{3}(\nabla^2 C_1 - 3C_1) \quad \text{for} \quad \eta \gg 1\,, \tag{5.38}$$

and the gravitational potential Φ decays exponentially, $\Phi \sim 4C_1\, \mathrm{e}^{-\eta}$.

Next, we consider a *radiation-dominated universe* ($p = \frac{1}{3}\varepsilon$). The background equations of motion yield the following time evolution of the scale factor:

$$a(\eta) = a_\mathrm{r} \begin{cases} \sinh \eta & \mathcal{K} = -1\,, \\ \eta & \mathcal{K} = 0\,, \\ \sin \eta & \mathcal{K} = 1\,, \end{cases}\,, \tag{5.39}$$

where a_r is a constant. In terms of physical time, (5.39) implies $a(t) \sim t^{1/2}$ in a spatially flat universe. It is easy to verify that in the case of ultrarelativistic matter with an equation of state $p = \varepsilon/3$ ($c_s^2 = 1/3$), $\theta''/\theta = 2a_r^2/a^2$, and, accordingly, eq. (5.25) for the rescaled gauge-invariant potential u becomes

$$u'' - \tfrac{1}{3}\nabla^2 u - (2a_r^2/a^2)u = 0 . \tag{5.40}$$

The general solution of this equation can be obtained as follows: we expand the function u in terms of eigenfunctions u_k of the operator ∇^2 ($-k^2$ denotes the eigenvalue of this operator) and solve eq. (5.40) for each mode separately. Resumming the terms, we find that the general solution of (5.40) can be expressed in terms of a function $\mathcal{D}(\eta, \boldsymbol{x})$ which satisfies the wave equation

$$(\partial^2/\partial^2\eta - \tfrac{1}{3}\nabla^2)\mathcal{D}(\boldsymbol{x}, \eta) = 0 . \tag{5.41}$$

The functions u and $\mathcal{D}$ are related as follows:

$$u(\boldsymbol{x}, \eta) \propto (\mathcal{H}^2 - \mathcal{H}' + \mathcal{K})^{-1/2}(\partial/\partial\eta)[\mathcal{D}(\boldsymbol{x}, \eta)/a] . \tag{5.42}$$

Hence, we can write the gauge-invariant gravitational potential Φ, energy-density perturbation $\delta\varepsilon^{(\mathrm{gi})}/\varepsilon_0$ and velocity perturbation $\delta u^{(\mathrm{gi})}$ in terms of $\mathcal{D}$ in the following form:

$$\begin{aligned} \Phi &= a^{-1}(\partial/\partial\eta)[a^{-1}\mathcal{D}(\boldsymbol{x}, \eta)] , \\ \frac{\partial\varepsilon^{(\mathrm{gi})}}{\varepsilon_0} &= 2\left[\left(\frac{a}{a_r}\right)^2\left(\frac{1}{a}\frac{\partial}{\partial\eta}\right)^2 a^2 - 1\right]\left(a^{-1}\frac{\partial}{\partial\eta}(a^{-1}\mathcal{D})\right) , \\ \delta u_i^{(\mathrm{gi})} &= -(1/2a_r^2)(\partial^2/\partial\eta^2)(a^{-1}\mathcal{D}_{,i}) . \end{aligned} \tag{5.43}$$

Let us analyze the implications of this solution in the case of a spatially flat universe. In this case, the general solution of the wave equation (5.41) can be expanded in plane waves. For fixed wave vector $\boldsymbol{k}$, the solution is

$$\mathcal{D}(\boldsymbol{x}, \eta) = [C_1 \sin(\omega\eta) + C_2 \cos(\omega\eta)]\, \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}} , \quad \omega = k/\sqrt{3} . \tag{5.44}$$

Hence, from (5.43),

$$\Phi(\boldsymbol{x}, \eta) = \eta^{-3}\{[\omega\eta \cos(\omega\eta) - \sin(\omega\eta)]C_1 + [\omega\eta \sin(\omega\eta) + \cos(\omega\eta)]C_2\}\, \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}} , \tag{5.45}$$

$$\begin{aligned} \delta\varepsilon^{(\mathrm{gi})}/\varepsilon_0 = (4/\eta^3)(&\{[(\omega\eta)^2 - 1]\sin(\omega\eta) + \omega\eta[1 - \tfrac{1}{2}(\omega\eta)^2]\cos(\omega\eta)\}C_1 \\ &+ \{[1 - (\omega\eta)^2]\cos(\omega\eta) + \omega\eta[1 - \tfrac{1}{2}(\omega\eta)^2]\sin(\omega\eta)\}C_2)\, \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}} . \end{aligned} \tag{5.46}$$

In the long-wavelength limit ($\omega\eta \ll 1$) when the scale of the perturbation is larger than the Hubble radius, the nondecaying mode has constant amplitude,

$$\Phi \sim \mathrm{const} , \quad \delta\varepsilon^{(\mathrm{gi})}/\varepsilon_0 \sim \mathrm{const} . \tag{5.47}$$

Since both potential Φ and $\delta\varepsilon^{(\mathrm{gi})}/\varepsilon_0$ are constant for long wavelengths, it is possible to give a self-consistent definition of statistical fluctuations. If we denote by N the number of particles in a box of length l, then statistical fluctuations in energy density on scale l can be defined as [3, 6, 73]

$$\delta\varepsilon^{(\mathrm{gi})}/\varepsilon_0 \sim N^{-1/2} . \tag{5.48}$$

In a radiation-dominated universe, neglecting the decaying mode, it does not matter at which time the fluctuations are defined, provided we consider times when l is longer than the Hubble radius. This is a further advantage of a gauge-invariant formalism.

In contrast to this situation, if we work in synchronous gauge, then in the corresponding coordinates the density fluctuations increase in time on scales larger than the Hubble radius. Hence, it is unclear how to define statistical fluctuations in terms of energy-density perturbations. A definition such as (5.48) would no longer be time independent. In chapter 20 we give a definition of statistical fluctuations based on the quantum theory which will be developed in part II. It will be shown that this definition is in agreement with (5.48).

Finally, we consider a universe filled with *cold matter and radiation* (in particular, the discussion here applies to the transition region between the radiation- and matter-dominated phases of the evolution of the universe). In this case, energy density and pressure are given by

$$\varepsilon = \varepsilon_{\mathrm{m}} + \varepsilon_{\mathrm{r}} , \qquad p = p_{\mathrm{r}} = \tfrac{1}{3}\varepsilon_{\mathrm{r}} , \tag{5.49}$$

where the subscripts m and r stand for matter and radiation, respectively. A careful study of this case is important for many applications, e.g., in calculating the spectrum of density perturbations on galactic scales in all cosmological models.

It is possible to solve the background equations exactly, and to find the following time dependence of the scale factor:

$$a(\eta) = \begin{cases} (2a_{\mathrm{eq}}/\eta_{\mathrm{eq}}^2)[\eta_{\mathrm{eq}} \sinh\eta + \cosh\eta - 1] & \mathcal{K} = -1 , \\ a_{\mathrm{eq}}[2(\eta/\eta_{\mathrm{eq}}) + (\eta/\eta_{\mathrm{eq}})^2] & \mathcal{K} = 0 , \\ (2a_{\mathrm{eq}}/\eta_{\mathrm{eq}}^2)(1 - \cos\eta + \eta_{\mathrm{eq}} \sin\eta) & \mathcal{K} = 1 , \end{cases} \tag{5.50}$$

$$\eta_{\mathrm{eq}} = [3a_{\mathrm{eq}}^2/(2\pi G\varepsilon_{\mathrm{r}} a^4)]^{1/2} , \tag{5.51}$$

and a_{eq} are conformal time and scale factor evaluated at the time t_{eq} of equal matter and radiation, i.e., when $\varepsilon_{\mathrm{r}} = \varepsilon_{\mathrm{m}}$. Unfortunately, in this case it is only possible to find the solutions of the equations for perturbations in the asymptotic limits. We will consider a single Fourier-mode perturbation or, more generally (for $\mathcal{K} \neq 0$), a perturbation whose space dependence is an eigenfunction of the operator ∇^2 with eigenvalue $-k^2$. In the long-wavelength limit $k\eta \ll c_s^{-1}$ we can neglect the spatial gradient term in the equation of motion (5.25) for u. As can be verified using eq. (5.9a) for c_s^2 and eq. (5.50) for $a(\eta)$, the above condition holds during the entire evolution of the universe for inhomogeneities with

$$k\eta_{\mathrm{eq}} \ll 1 \tag{5.52}$$

in the model under consideration. Note that this model breaks down at recombination at which point c_s^2

rapidly drops to 0. It is important to stress that the long-wavelength limit considered here is different from demanding that the wavelength be larger than the Hubble radius. The asymptotic behavior discussed here is valid also on smaller scales, provided (5.52) is satisfied; (5.52) means that the scale of the perturbation enters the Hubble radius after η_{eq}, the time of equal matter and radiation.

For $\mathcal{K}=0$, calculating θ for $a(\eta)$ given by (5.50), substituting the result into (5.31) and integrating, one finds u. Then, using (5.23) and (5.28) we obtain the following results for long-wavelength perturbations:

$$\Phi = \frac{\xi+1}{(\xi+2)^3}\left[C_{1k}\left(\tfrac{3}{5}\xi^2 + 3\xi + \frac{1}{\xi+1} + \tfrac{13}{3}\right) + C_{2k}\frac{1}{\xi^3}\right], \tag{5.53}$$

$$\frac{\delta\varepsilon^{(\mathrm{gi})}}{\varepsilon_0} = -2\left[\left(\frac{(k\eta_{eq})^2}{12}\frac{\xi^2(\xi+2)^2}{(\xi+1)^2} - \frac{3\xi^2+6\xi+4}{2(\xi+1)^2}\right)\Phi + C_{1k}\frac{3\xi^4+4\xi^3+18\xi^2+30\xi+16}{2(\xi+1)^2(\xi+2)^2}\right], \tag{5.54}$$

where $\xi = \eta/\eta_{eq}$ and where the subscript k refers to the eigenvalue of ∇^2.

In spite of the apparent complexity of the above equations, the main result is easy to read off. Considering the nondecaying mode only, we see that Φ and $\delta\varepsilon^{(\mathrm{gi})}/\varepsilon_0$ are constant at times both early and late compared to η_{eq}. During the transition, the amplitude of both Φ and $\delta\varepsilon^{(\mathrm{gi})}/\varepsilon_0$ decreases by a factor 9/10. Also, the amplitude of $\delta\varepsilon^{(\mathrm{gi})}/\varepsilon_0$ is twice that of Φ. Note that the decrease in Φ by a factor of 9/10 during the transition between the radiation- and matter-dominated epochs can easily be derived by applying the conservation law (5.23).

The formulas for an open universe are more complicated. Again, they follow from integrating (5.25) explicitly,

$$\begin{aligned}\Phi = {} & (\eta_{eq}\sinh\eta + \cosh\eta - 1)^{-3}[C_{1k}(3(\eta_{eq}\cosh\eta + \sinh\eta)^2 \\ & + (2\eta_{eq}^2 - 9)\eta(\eta_{eq}\cosh\eta + \sinh\eta) + (3 - 2\eta_{eq}^2)\eta_{eq}\sinh\eta \\ & + (12 - 10\eta_{eq}^2)\cosh\eta + 7\eta_{eq}^2 - 12) + C_{2k}(\eta_{eq}\cosh\eta + \sinh\eta)],\end{aligned} \tag{5.55}$$

$$\begin{aligned}\delta\varepsilon^{(\mathrm{gi})}/\varepsilon_0 = {} & -2\left(\frac{\tfrac{1}{3}k^2(\eta_{eq}\sinh\eta + \cosh\eta - 1)^2 - 3(\eta_{eq}\sinh\eta + \cosh\eta - 1) - 2\eta_{eq}^2}{\eta_{eq}^2 + 2(\eta_{eq}\sinh\eta + \cosh\eta - 1)}\Phi\right. \\ & \left. + C_{1k}\frac{3(\eta_{eq}\sinh\eta + \cosh\eta - 1) + 2\eta_{eq}^2}{\eta_{eq}^2 + 2(\eta_{eq}\sinh\eta + \cosh\eta - 1)}\right).\end{aligned} \tag{5.56}$$

The corresponding formulas for a closed universe can be obtained from the above by substituting $\eta \to \mathrm{i}\eta$. The solutions given by (5.53) and (5.55) are exact in the limit $k \ll aH$. For $\eta \ll 1$, the solutions for $\mathcal{K}=0$ and $\mathcal{K}=\pm 1$ coincide.

Summary. Let us now summarize the evolution of adiabatic perturbations in a universe dominated by hydrodynamical matter. We shall focus on the nondecaying mode. For inhomogeneities with wavelength larger than the Hubble radius, the evolution of Φ and $\delta\varepsilon^{(\mathrm{gi})}/\varepsilon_0$ is quite simple. When the equation of state is constant, i.e., for early times when $p \simeq \varepsilon/3$ and at late times when $p = 0$, the amplitudes of Φ and $\delta\varepsilon^{(gi)}/\varepsilon_0$ are constant. The values of Φ and $\delta\varepsilon^{(gi)}/\varepsilon_0$ are related, $\delta\varepsilon^{(gi)}/\varepsilon_0 \simeq -2\Phi$. This result is true for any arbitrary, time independent, equation of state $p = w\varepsilon$ with constant w. For such an equation of state, the solution in a flat universe ($\mathcal{K} = 0$) for a general k is

$$\Phi = \eta^{-\nu}[C_1 J_\nu(\sqrt{w}\, k\eta) + C_2 Y_\nu(\sqrt{w}\, k\eta)]\,, \quad \nu = \tfrac{1}{2}(5 + 3w)/(1 + 3w)\,, \tag{5.57}$$

where J_ν and Y_ν are the Bessel functions of order ν. Considering the small-argument expansion of the Bessel functions, we see that for $k\eta \ll 1$ the nondecaying mode of Φ is constant.

When the equation of state changes from $w = 1/3$ to $w = 0$ at the time of equal matter and radiation ($\eta \simeq \eta_{eq}$), the amplitude of the long-wavelength perturbation changes by a factor 9/10. If such perturbations ($k\eta_{eq} \ll 1$) enter the Hubble radius at $k^{-1}\mathcal{H} \sim 1$, then, during the period of matter domination, Φ remains constant while $\delta\varepsilon^{(gi)}/\varepsilon_0$ begins to increase, $\delta\varepsilon^{(gi)}/\varepsilon_0 \sim \eta^2$.

In an open universe ($\mathcal{K} = -1$), for a linear perturbation that enters the horizon at $\eta \ll 1$, the amplitude of $\delta\varepsilon^{(gi)}/\varepsilon_0$ once again freezes out when $\eta > 1$, whereas the amplitude of Φ starts to decrease as $a^{-1}(\eta)$ [see (5.55) and (5.56)]. Figure 5.1 shows the evolution of the amplitude of Φ and $\delta\varepsilon^{(gi)}/\varepsilon_0$ in an open universe.

Short-wavelength perturbations ($k\eta_{eq} \gg 1$) enter the Hubble radius before the time of equal matter and radiation η_{eq}. For $\eta_{eq} < \eta_r$ (recombination time), there will be a time interval given by $\eta_r \gg \eta \gg 1/k$ when the spatial gradient term in the equation of motion for the perturbations (5.25) dominates (i.e., $k\eta \gg c_s^{-1}$). In that case, we can use the WKB approximation to find the solutions in this time interval. While $\eta \ll \eta_{eq}$ we can neglect the cold matter to describe the evolution of perturbations [(5.45) and (5.46)]. A comparison of the WKB solution with the radiation era, once in their common interval of applicability, $\eta_{eq} \gg \eta \gg 1/k$, gives the initial conditions for the WKB solution which then can be used in its entire range of validity (i.e., up to η_r).

These perturbations become sound waves when they enter the Hubble radius (strictly speaking when the scale becomes smaller than the Jeans length, $k\eta \gg c_s^{-1}$). The amplitude of $\delta\varepsilon^{(gi)}/\varepsilon_0$ for these sound waves is approximately constant while the potential Φ decays as η^{-2}. At the time of recombination, the

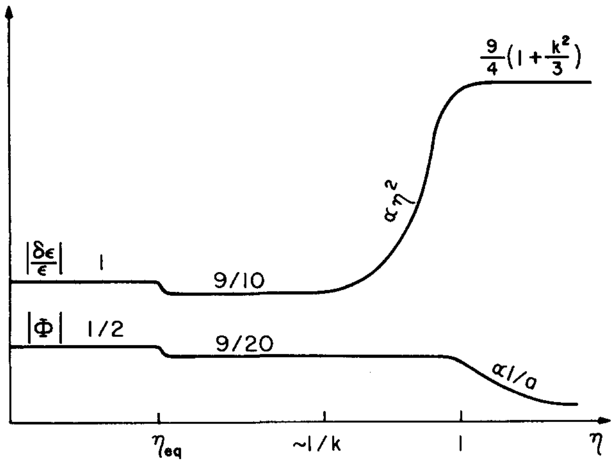

Fig. 5.1. Time evolution of adiabatic perturbations of wavenumber k in an open universe.

velocity of sound drops abruptly to a negligible value. For many scales with $k\eta_{eq} \gg 1$, the condition of applicability of the long-wave approximation ($k\eta \ll c_s^{-1}$) is restored. Since we are in the matter-dominated period, we can use the appropriate equations [(5.32) and (5.33)] to describe the behavior of these perturbations.

5.4. Entropy perturbations

Entropy perturbations can and generically will arise in all multi-component systems. They also generate scalar-type metric perturbations. The basic equations for Φ or equivalently u are (5.22) and (5.25). For entropy perturbations, the source terms do not vanish. We shall assume that at the initial time there are no adiabatic perturbations. Then, to define the entropy perturbations, we impose the initial condition

$$\Phi \rightarrow 0 \quad \text{as} \quad t \rightarrow 0 . \qquad (5.58)$$

Instead of distinguishing between adiabatic and entropy perturbations, fluctuations are often divided into isocurvature and adiabatic ones. Isocurvature perturbations are defined by the initial condition that the initial curvature perturbation vanishes, i.e., that the gauge-invariant curvature perturbation ζ vanishes at the initial time [35]. If the initial time is taken to be 0, then this definition coincides with the definition (5.58) of entropy perturbations. If the initial time is finite, then the two definitions might differ by a term proportional to the decaying mode of Φ.

Entropy perturbations can be generated if the different matter components are distributed nonuniformly in space but with uniform total energy density and hence uniform curvature at the beginning. Such perturbations are hence often called isocurvature perturbations. An example of entropy perturbations is an inhomogeneous distribution of baryons in a radiation background, the energy density excess in baryons being initially compensated by a deficit in radiation energy.

Entropy perturbations may well be important for galaxy formation. In particular, they can be generated in axion models (refs. [74, 75]) and in nonsimple inflationary models [76–81]. They also arise in phase transitions which produce topological defects [82]. For example, in phase transitions which produce cosmic strings, strings stretching from one side of the universe to the other will be produced [83]. Causality constraints [84] forbid the formation of adiabatic perturbations on scales larger than the Hubble radius. Hence, the only perturbations which can be formed on these scales are isocurvature (entropy) perturbations. In the case of cosmic strings, this implies that the energy overdensity in scalar-field matter along the string must be compensated by an energy deficit in radiation [85].

In the following, we shall consider various effects connected with entropy perturbations. First, we shall investigate the behavior of long-wavelength inhomogeneities with $k\eta \ll c_s^{-1}$ in a universe filled with cold plasma and radiation. This case is sufficiently simple to analyze but demonstrates the specific features of the evolution of entropy perturbations. In the second part of this section, we will study the generation of sound waves by short-wavelength entropy perturbations near η_{eq}.

As before, we shall expand the fluctuation into eigenmodes of the operator ∇^2 and consider perturbations with a space dependence corresponding to a solution of the mode equation,

$$\nabla^2 \Phi_k = -k^2 \Phi_k , \qquad (5.59)$$

with fixed wavenumber k. We will first consider *long-wave perturbations*. The formulas obtained below [until eq. (5.67)] are applicable for inhomogeneities with $k\eta \ll c_s^{-1}$ and for times smaller than the time of recombination. They are also applicable on scales larger than the Hubble radius ($k\eta \ll 1$) for later times, since on these scales for causality reasons $\delta S/S = \text{const}$ (on scales smaller than the Hubble radius, in general $\delta S/S \neq \text{const}$ for $\eta > \eta_r$).

We start with eq. (5.22) and omit the term $c_s^2 \nabla^2 \Phi$ since $k\eta \ll c_s^{-1}$. Expressing $\mathcal{H}$ and $\mathcal{H}'$ in terms of ε_0 and p_0, using the background equations of motion, and taking into account (5.9a, b), we can rewrite this equation in the form

$$\Phi'' - \frac{(4\varepsilon_r + 3\varepsilon_m)'}{4\varepsilon_r + 3\varepsilon_m}\,\Phi' + \left(\frac{8\pi G a^2}{3}\,\frac{\varepsilon_m \varepsilon_r}{4\varepsilon_r + 3\varepsilon_m} - 2\mathcal{K}\,\frac{8\varepsilon_r + 3\varepsilon_m}{4\varepsilon_r + 3\varepsilon_m}\right)\Phi = \frac{16\pi G a^2}{3}\,\frac{\varepsilon_m \varepsilon_r}{4\varepsilon_r + 3\varepsilon_m}\,\frac{\delta S}{S}\,, \tag{5.60}$$

where $\delta S/S$ is constant. Hence, for a flat universe ($\mathcal{K} = 0$) we immediately can write down a particular solution for Φ,

$$\Phi = 2\,\delta S/S\,. \tag{5.61}$$

Obviously, the above particular solution does not satisfy the required initial conditions (5.58). The solution which does is obtained by adding to (5.61) a general solution of the homogeneous equation [for $\mathcal{K} = 0$ see (5.53)] and choosing the coefficients such that the initial conditions are satisfied. In a flat universe, the result is

$$\Phi = \tfrac{1}{5}\xi\,\frac{\xi^2 + 6\xi + 10}{(\xi + 2)^3}\,\frac{\delta S}{S}\,, \qquad \xi = \eta/\eta_{eq}\,, \tag{5.62}$$

$$\frac{\delta\varepsilon^{(gi)}}{\varepsilon_0} = -\left[\frac{(k\eta_{eq})^2}{30}\,\frac{\xi^3(\xi^2 + 6\xi + 10)}{(\xi + 1)^2(\xi + 2)} + \frac{2}{5}\,\frac{\xi}{(\xi + 2)^3}\left(\frac{\xi^3 + 7\xi^2 + 18\xi + 20}{\xi + 1}\right)\right]\frac{\delta S}{S}\,. \tag{5.63}$$

The important conclusion (see also ref. [86]) is that the gauge-invariant amplitude for this type of entropy perturbations increases linearly in conformal time until η_{eq}, whereas it is constant for adiabatic perturbations.

The calculations are significantly more involved in the case of a nonflat universe ($\mathcal{K} = \pm 1$). Substituting

$$\Phi = 2\,\delta S/S + \Phi_1 \tag{5.64}$$

into (5.60), we obtain an equation for Φ_1 which can be solved by standard methods [87]. Taking into account the initial conditions (5.58) and using (5.55) to satisfy them, one finally gets (for $\mathcal{K} = -1$)

$$\begin{aligned}\Phi = {} & (n_{eq}\sinh\eta + \cosh\eta - 1)^{-3} \\ & \times [(\eta_{eq}\cosh\eta + \sinh\eta)^2 - 3\eta(\eta_{eq}\cosh\eta + \sinh\eta) \\ & + 2(2 - \eta_{eq}^2)\cosh\eta + \eta_{eq}\sinh\eta + \eta_{eq}^2 - 4]\,\delta S/S\,. \end{aligned} \tag{5.65}$$

The corresponding expression for $\delta\varepsilon^{(gi)}/\varepsilon_0$ follows immediately from (5.28). The formulas for a closed universe ($\mathcal{K}=1$) can be obtained by substituting $i\eta$ for η in the above equation. Using (5.6) and $\delta\varepsilon = \delta\varepsilon_m + \delta\varepsilon_r$ we can express $\delta\varepsilon_m/\varepsilon_m$ and $\delta\varepsilon_r/\varepsilon_r$ in terms of $\delta\varepsilon/\varepsilon$ and $\delta S/S$.

In figure 5.2 we plot the time dependence of the amplitudes of Φ, $\delta\varepsilon_m/\varepsilon_m$ and $\delta\varepsilon/\varepsilon$ for entropy perturbations in an open universe. (The behavior of entropy perturbations in a flat universe is exactly the same as here for times $\eta \ll 1$.) The amplitudes of Φ, $\delta\varepsilon/\varepsilon$ and $\delta\varepsilon_r/\varepsilon_r$ increase in time until $\eta=\eta_{eq}$. Between η_{eq} and $\eta_k \simeq 1/k$ they are constant (we are assuming $\eta_{eq} \ll \eta_k \ll 1/k$). The fluctuations in the cold-matter density, $\delta\varepsilon_m/\varepsilon_m$ decrease to 2/5 of their initial values by the time $\eta=\eta_{eq}$. For $\eta>\eta_{eq}$, the entropy perturbations evolve like the nondecaying mode of the adiabatic perturbations. There is a difference, however: for adiabatic perturbations

$$\delta\varepsilon_m/\varepsilon_m = \tfrac{3}{4}\,\delta\varepsilon_r/\varepsilon_r\,, \tag{5.66}$$

whereas for entropy perturbations

$$\delta\varepsilon_m/\varepsilon_m = -\tfrac{1}{2}\,\delta\varepsilon_r/\varepsilon_r\,. \tag{5.67}$$

Note that entropy perturbations do not turn into decaying adiabatic perturbations when they enter the Hubble radius. This point was incorrectly treated in some of the literature [88].

Let us finally investigate the generation of sound waves by *short-wavelength* ($k\eta_{eq} \gg 1$) entropy perturbations in a bath of cold plasma and radiation. To simplify the calculations we will consider only the case of a flat universe ($\mathcal{K}=0$). At first, let us investigate times η before the time of equal matter and radiation η_{eq}. In this case $c_s^2 = 1/3$ and the scale factor $a(\eta)$ evolves according to (5.39). Hence, the equation of motion (5.25) for u takes the form

$$u'' - \tfrac{1}{3}\nabla^2 u - (2a_{eq}^2/a^2)u = (a_{eq}/\sqrt{2\pi G}\,\eta_{eq}^2)(\delta S/S)\eta\,, \tag{5.68}$$

where we took into account (5.9b) when evaluating the right-hand side of (5.25). For $\eta \ll \eta_{eq}$, this equation is correct for perturbations with arbitrary k. The solution satisfying the initial conditions (5.58)

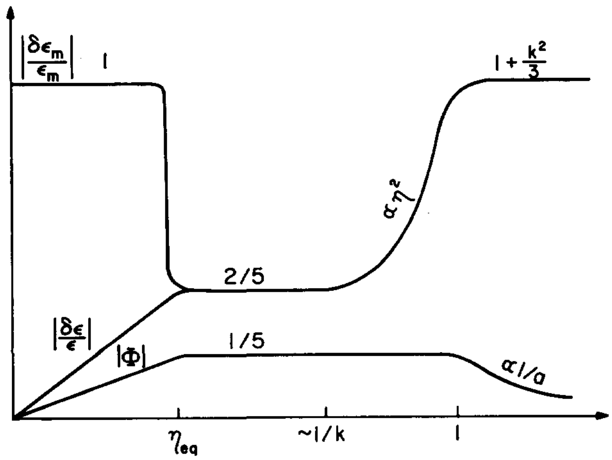

Fig. 5.2. Time evolution of entropy perturbations of wavenumber k in an open universe.

recalculated for Φ by means of (5.24) is

$$\Phi_k = \frac{\delta S}{S} \frac{\eta}{\eta_{eq}} \frac{1}{(\omega\eta_{eq})^3} \left[\omega\eta + \frac{2}{\omega\eta} + 2\eta \frac{d}{d\eta}\left(\frac{\cos\omega\eta}{\omega\eta}\right)\right], \quad \omega = \frac{k}{\sqrt{3}} . \tag{5.69}$$

It is important to stress that solution (5.69) is valid both for short-wavelength and for long-wavelength perturbations at early times ($\eta < \eta_{eq}$). Next, let us consider short-wavelength fluctuations ($k\eta \gg c_s^{-1}$) at times in the interval $\eta_r > \eta \gg 1/k$. For them, the last term in (5.25) is negligible and we can use the WKB approximation to find the solution. The result is

$$\Phi_k = F(\eta) + c_s^{-3/2}(a_{eq}/a)^2\left[A_k \sin\left(\int c_s k\, d\eta\right) + B_k \cos\left(\int c_s k\, d\eta\right)\right], \tag{5.70}$$

where F does not contain oscillating terms to leading order in $(k\eta)^{-1}$. The coefficients A_k and B_k can be obtained by matching (5.70) with (5.69) in the interval $\eta_{eq} \gg \eta \gg 1/k$ when both of the solutions are valid. The matching conditions have been discussed in detail in ref. [89]. The final result is

$$\frac{\delta\varepsilon^{(gi)}}{\varepsilon_0} \simeq \frac{3^{7/4} c_s^{1/2}}{k\eta_{eq}} \frac{\delta S}{S} \sin\left(\int_0^\eta c_s k\, d\eta\right) - \frac{\delta S}{S} . \tag{5.71}$$

This formula is true for $\eta_r \gg \eta \gg 1/k$. Besides the constant mode coming from $F(\eta)$, there is an oscillating part of the matter-density perturbation which corresponds to sound waves which were generated by the entropy perturbations [90]. These sound waves can be significant for anisotropies in the microwave background.

Note that for a collisionless gas, the above formulation of hydrodynamical perturbations should be extended to allow for a gauge-invariant analysis of the phase-space density of the gas. This has been done in refs. [71, 66].

6. Perturbations in the universe filled by a scalar field

At very high energies, it is no longer reasonable to believe that a hydrodynamical description of matter will be valid. Rather, we expect that matter will be described in terms of fields. In many particle-physics models, scalar fields are introduced. Their role is to break high-energy symmetries and to give fermions masses by the mechanism of spontaneous symmetry breaking [91]. For scalar fields, it is possible to construct nontrivial potential terms (we call all terms in the Lagrangian not containing derivatives potential terms). Nontrivial means that there are higher than quadratic terms. These terms correspond to self-interactions of the scalar field φ [92].

It was soon realized that due to the potential $V(\varphi)$, scalar fields can play a particular role in cosmology [93]. In particular, an equation of state unfamiliar from classical statistical mechanics may result; for example, if φ is homogeneous in space and trapped in a local minimum of the potential. In this case, the scale factor of the universe will expand exponentially; this is the de Sitter phase [94].

In this chapter we study perturbations in a universe filled with scalar-field matter. In the first section, the energy–momentum tensor for a scalar field will be given. Then, the evolution of the background in the presence of a homogeneous scalar field will be studied. With only mild conditions on $V(\varphi)$ (the self-coupling constants must all be somewhat smaller than 1) most energetically allowed (to justify a classical consideration of the gravitational field) initial conditions lead to chaotic type inflation [95]. In the third section, the gauge-invariant equations for cosmological perturbations in a model with scalar-field matter are studied. After reducing the equations to a single second-order differential equation for perturbations, its solutions in various interesting cases can be obtained.

6.1. Energy–momentum tensor

In this chapter we take the matter in the universe to be a scalar field minimally coupled to gravity. Such a theory is given by the action

$$S = \int [\tfrac{1}{2}\varphi^{;\alpha}\varphi_{;\alpha} - V(\varphi)]\sqrt{-g}\,\mathrm{d}^4x\,, \tag{6.1}$$

where $V(\varphi)$ is the potential of scalar field. The corresponding energy–momentum tensor is

$$T^{\mu}_{\nu} = \varphi^{;\mu}\varphi_{;\nu} - [\tfrac{1}{2}\varphi^{;\alpha}\varphi_{;\alpha} - V(\varphi)]\delta^{\mu}_{\nu}\,, \quad \varphi^{;\mu} = g^{\mu\nu}\varphi_{;\nu}\,. \tag{6.2}$$

We shall consider a homogeneous and isotropic universe with small scalar-type metric perturbations (2.9). For consistency the scalar field must also be approximately homogeneous. In this case the field $\varphi(\boldsymbol{x}, t)$ can be decomposed into two parts,

$$\varphi(\boldsymbol{x}, t) = \varphi_0(t) + \delta\varphi(\boldsymbol{x}, t)\,, \tag{6.3}$$

where $\varphi_0(t)$ is the homogeneous part of the scalar field which drives the background isotropic model and $|\delta\varphi| \ll \varphi_0$ is a small perturbation. The energy–momentum tensor can be also decomposed into "background" and "perturbed" parts

$$T^{\mu}_{\nu} = {}^{(0)}T^{\mu}_{\nu} + \delta T^{\mu}_{\nu}\,, \tag{6.4}$$

where δT^{μ}_{ν} is linear in matter and metric perturbations $\delta\varphi$ and $\delta g_{\alpha\beta}$.

Substituting (6.3) and (2.9) into (6.2) we obtain the background energy–momentum tensor (in conformal time)

$${}^{(0)}T^0_0 = \frac{1}{2a^2}\,\varphi_0'^2 + V(\varphi_0) = \varepsilon\,, \qquad {}^{(0)}T^0_i = 0\,, \qquad {}^{(0)}T^i_j = [-(1/2a^2)\varphi_0'^2 + V(\varphi)]\delta^i_j = -p\delta^i_j\,, \tag{6.5}$$

and the first-order perturbation

$$\begin{aligned} &\delta T^0_0 = a^{-2}[-\varphi_0'^2\phi + \varphi_0'\,\delta\varphi' + V_{,\varphi}a^2\,\delta\varphi]\,, \qquad \delta T^0_i = a^{-2}\varphi_0'\,\delta\varphi_{,i}\,, \\ &\delta T^i_j = [\varphi_0'^2\phi - \varphi_0'\,\delta\varphi' + V_{,\varphi}a^2\,\delta\varphi]\delta^i_j\,. \end{aligned} \tag{6.6}$$

Here a prime denotes differentiation with respect to conformal time η, $V_{,\varphi} = \mathrm{d}V/\mathrm{d}\varphi$, and the comma with space index means differentiation with respect to the corresponding coordinate.

For the sake of simplicity, we shall in this section only consider a flat universe ($\mathcal{K}=0$). Most of the results which will be obtained can be generalized without difficulties to the case of a space with arbitrary curvature ($\mathcal{K}=\pm 1$). To write down the equations for cosmological perturbations we will need the gauge-invariant perturbations of the energy–momentum tensor $\delta T_{\beta}^{(\mathrm{gi})\alpha}$ [see eq. (4.13)]. Using (6.6) and (6.5) we find

$$\delta T_{0}^{(\mathrm{gi})0} = a^{-2}[-\varphi_0'^2\Phi + \varphi_0'\,\delta\varphi^{(\mathrm{gi})\prime} + V_{,\varphi}a^2\,\delta\varphi^{(\mathrm{gi})}]\,, \qquad \delta T_{i}^{(\mathrm{gi})0} = a^{-2}\varphi_0'\,\delta\varphi_{,i}^{(\mathrm{gi})}\,,$$
$$\delta T_{j}^{(\mathrm{gi})i} = a^{-2}[+\varphi_0'^2\Phi - \varphi_0'\,\delta\varphi^{(\mathrm{gi})\prime} + V_{,\varphi}a^2\,\delta\varphi^{(\mathrm{gi})}]\delta_j^i\,, \tag{6.7}$$

where Φ is the gauge-invariant potential (3.13), and

$$\delta\varphi^{(\mathrm{gi})} = \delta\varphi + \varphi_0'(B-E') \tag{6.8}$$

is the gauge-invariant perturbation of the scalar field.

6.2. Background

In the case of hydrodynamical matter, the time evolution of the background model was simple. For a time-independent equation of state, the scale factor $a(t)$ increases as a fixed power of t. For scalar-field matter, the background evolution is significantly more complicated. The homogeneous background part $\varphi_0(t)$ of the scalar field has nontrivial dynamics. This leads to a time-dependent equation of state, and to a more complicated time evolution of $a(t)$.

In this section, we shall analyze the background model in some detail. The time evolution of $\varphi_0(t)$ depends on the scale factor $a(t)$, which in turn depends on the equation of state of matter. Thus, we are dealing with a system of coupled differential equations. We shall present an analysis of the background evolution which does not depend on the particular scalar-field potential $V(\varphi)$. The reader who is only interested in perturbations may skip this section.

The main ideas of our method are the following. We first eliminate the explicit time dependence from the problem. This allows an analysis which is independent of the particular potential. The next step is to consider a generalized phase diagram for the resulting background equations. We investigate the time evolution which yields curves in this phase diagram, and find attractors for the dynamical evolution. This step in the analysis generalizes a method used previously [95] to study the time evolution of a homogeneous scalar field in an expanding universe. Finally, we calculate $a(t)$ for attractor curves. In this section, physical time t and not conformal time η will be used. Also, the zero subscript for background fields will be omitted.

The starting point consists of the time–time Einstein equation, which, using (6.5), gives

$$H^2 = l^2[\tfrac{1}{2}\dot{\varphi}^2 + V(\varphi)]\,, \tag{6.9}$$

where the derivative with respect to physical time t is denoted by a dot, $l^2 = 8\pi G/3$ and $H \equiv \dot{a}/a$ is the Hubble parameter. The second starting equation is the scalar-field equation of motion,

$$\ddot{\varphi} + 3H\dot{\varphi} + V_{,\varphi} = 0\,. \tag{6.10}$$

This equation can be derived by varying the action (6.1) with respect to φ, or from the conservation law (2.7) for the energy–momentum tensor (6.5). Taking the derivative of eq. (6.9) with respect to time and subtracting $\dot{\varphi}$ times eq. (6.10) we also obtain the following useful relationship:

$$\dot{H} = -\tfrac{3}{2} l^2 \dot{\varphi}^2 . \tag{6.11}$$

Before starting the technical analysis, we mention some important points. Using the mild constraints on $V(\varphi)$ mentioned at the beginning of this chapter, fairly general initial conditions for a homogeneous scalar field will lead to a period of inflation, i.e., exponential expansion of the scale factor. To see heuristically how this might occur, consider a situation when φ is static ($\dot{\varphi}^2 \simeq 0$) but $V(\varphi)$ does not vanish but takes on a large positive value. In this case, the right-hand side of (6.9) is approximately constant and hence

$$a(t) \propto \exp(Ht) , \quad H \simeq lV(\varphi)^{1/2} . \tag{6.12a, b}$$

Note that exponential inflation corresponds to an equation of state $p \simeq -\varepsilon$, as can be seen from (6.5). In the following, we shall see that inflation will arise for a wide region of initial data. It is important to mention that initial large spatial inhomogeneities can prevent the onset of inflation. For a discussion of this issue see e.g. refs. [24, 96–98].

There are different approaches to investigate the field equations in the model given by eqs. (6.9) and (6.10). For example, using (6.9) we may express H in terms of φ and $\dot{\varphi}$ and substitute it in (6.10). Then we obtain a differential equation for the variable φ only, which can be studied by means of a phase diagram method. In the following discussion we follow a more general approach.

As a first step, let us introduce the new variable $x = \ln a$ instead of t and transform the equations. We have

$$\dot{x} = H , \qquad \mathrm{d}/\mathrm{d}t = H\, \mathrm{d}/\mathrm{d}x . \tag{6.13}$$

Using (6.13) to express the time derivative in terms of derivatives with respect to x, the expression (6.9) for the Hubble parameter can be recast as

$$H = lV^{1/2}[1 - \tfrac{1}{6} l^2 (\mathrm{d}\varphi/\mathrm{d}x)^2]^{-1/2} . \tag{6.14}$$

Applying eqs. (6.13) and (6.14), the equation of motion (6.10) for φ takes the form

$$\left\{ \frac{\mathrm{d}^2\varphi}{\mathrm{d}^2 x} + 3\left[1 - \tfrac{1}{6} l^2 \left(\frac{\mathrm{d}\varphi}{\mathrm{d}x}\right)^2\right] \frac{\mathrm{d}\varphi}{\mathrm{d}x} \right\} \frac{l^2 V}{1 - \tfrac{1}{6} l^2 (\mathrm{d}\varphi/\mathrm{d}x)^2} + V_{,\varphi} = 0 . \tag{6.15}$$

This equation can be reduced to a first-order equation if we introduce the new variable $y = \mathrm{d}\varphi/\mathrm{d}x$, and consider it as a function of φ,

$$\mathrm{d}y/\mathrm{d}\varphi = -(1 - \tfrac{1}{6} l^2 y^2)[3 + (V_{,\varphi}/V)/l^2 y] . \tag{6.16}$$

From (6.14) it follows that for physical real-time solutions (called Lorentzian solutions) there is a constraint on the magnitude of y, $|y| \le \sqrt{2}/l$. This follows since for Lorentzian solutions H must be real. Points on the φ–y plane where this condition is not satisfied correspond to Euclidean solutions (solutions defined in imaginary time).

Equation (6.16) can be analyzed by means of the phase-diagram method [99]. The general structure of this diagram does not depend on the details of the potential $V(\varphi)$. Only two assumptions are important: first we consider potentials $V(\varphi)$ which are symmetric functions of φ. Secondly, we assume that as $|\varphi|\to\infty$, $V(\varphi)$ increases less fast than an exponential, i.e.,

$$V_{,\varphi}/V\to 0\,, \tag{6.17}$$

as $|\varphi|\to\infty$. In this case, the phase diagram of eq. (6.16) is depicted schematically in fig. 6.1.

Let us start with solutions which are close to the lines $|y| = \sqrt{2}/l$. For such solutions $V \ll \dot{\varphi}^2$, and in this case the scale factor evolves as in a universe filled with hydrodynamical matter which obeys the equation of state $p = +\varepsilon$ $[a(t) \propto t^{1/3}]$.

To investigate the solutions near $y = \sqrt{2}/l$ [which is itself a solution of (6.16)], we write y in the form $y = \pm(\sqrt{2}/l)(1 - \Delta y)$, where $0 < \Delta y \ll 1$ for Lorentzian solutions. The linearized equation for Δy immediately follows from (6.16),

$$\mathrm{d}\,\Delta y/\mathrm{d}\varphi = \pm(3\sqrt{2}l \pm V_{,\varphi}/V)\,\Delta y\,, \tag{6.18}$$

and has a solution of the form

$$\Delta y = \Delta y_i[V(\varphi)/V(\varphi_i)]\exp[\pm 3\sqrt{2}l(\varphi - \varphi_i)]\,, \tag{6.19}$$

where φ_i and Δy_i are the initial conditions for (6.16): $\Delta y_i = \Delta y(\varphi_i)$.

Analyzing eqs. (6.18) and (6.19), it is easy to understand the flow lines near $|y| = \sqrt{2}/l$ in fig. 6.1. (Note that the arrows in the figure indicate evolution in time.) Consider as the simplest example the

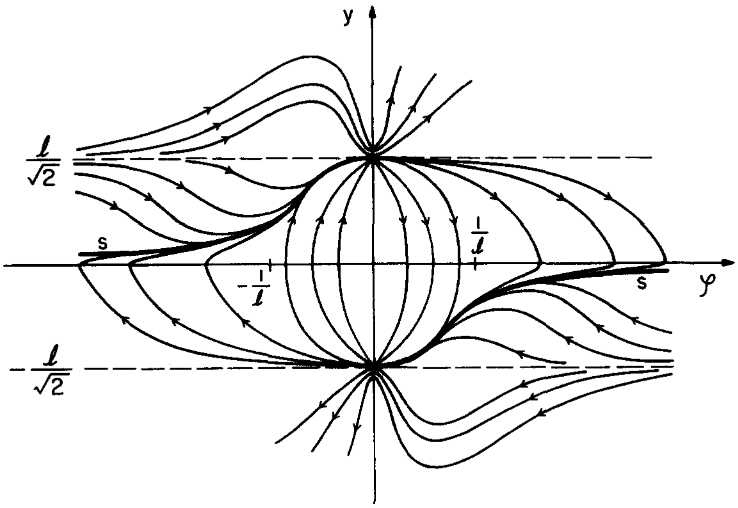

Fig. 6.1. The phase diagram for the background equations describing a scalar-field dominated universe. The bold lines labelled s are separatrices since they separate trajectories of different type from one another [101].

theory of a free real scalar field with mass m, for which the potential is

$$V(\varphi) = \tfrac{1}{2}m^2\varphi^2 . \tag{6.20}$$

This is one of the models used to realize chaotic inflation [100]. In this case, (6.19) reads

$$\mathrm{d}\,\Delta y/\mathrm{d}\varphi = \pm(3\sqrt{2}l \pm 2/\varphi)\,\Delta y . \tag{6.21}$$

For large positive and negative values of φ, the bracket is positive. Hence, as φ increases starting at large negative values, the trajectory moves away from the line $y = \sqrt{2}/l$. For small negative values of φ, however, trajectories converge back towards the line. For small positive values of φ, trajectories are again repelled.

Since the phase diagram is symmetric about the y-axis, we shall restrict our attention to the behavior of trajectories on the left-hand side of the diagram. Moving away from the line $y = \sqrt{2}/l$, we eventually reach another curve in the phase diagram which is an attractor for solutions, a separatrix [101] marked s in fig. 6.1. At large values of φ ($|\varphi| \gg 1/l$) this separatrix is close to the line where $\mathrm{d}y/\mathrm{d}\varphi = 0$. Hence, we can obtain the equation for this separatrix in the form

$$y = -(1/3l^2)V_{,\varphi}/V + \mathrm{O}[(V_{,\varphi}/V)^2, V_{,\varphi\varphi}/V, \ldots)] . \tag{6.22}$$

For $|\varphi| \gg 1/l$ we can neglect the terms of order $\sim 1/l^2\varphi^2$ and higher.

The ratio $\dot{H}/H^2$ can be calculated from (6.11) and (6.13). Inspecting the result, it follows that for solutions in the vicinity of this separatrix, for $|\varphi| \gg 1/l$

$$|\dot{H}|/H^2 = \tfrac{3}{2}l^2y^2 \ll 1 , \tag{6.23}$$

provided that the potential increases less fast than exponential [see (6.17)]. This implies that these solutions describe an inflationary period. We call them quasi de Sitter solutions for which the effective equation of state is $p \approx -\varepsilon$. For an exponential potential, the background evolution leads to power-law inflation [102].

Since $y^{-1} = \mathrm{d}x/\mathrm{d}\varphi$, we can integrate (6.18) to obtain

$$a(\varphi) \sim \exp\left(-3l^2 \int (V/V_{,\varphi})\,\mathrm{d}\varphi\right) \tag{6.24}$$

during the inflationary period. For a polynomial potential,

$$V(\varphi) = (\lambda_n/n)\varphi^n , \tag{6.25}$$

one gets

$$a(\varphi) = a_i \exp[-(3l^2/2n)(\varphi^2 - \varphi_i^2)] , \tag{6.26}$$

To find the dependence of the scale factor on time we need to obtain the time dependence of the scalar field φ. During the quasi de Sitter stage we may use $H \simeq lV^{1/2}$ and (6.22) to obtain the equation

$$\dot{\varphi} = Hy \simeq -(1/3l)V_{,\varphi}/V^{1/2}\,. \tag{6.27}$$

Integrating this equation gives the time dependence of φ. For polynomial potentials (6.25) with $n \neq 4$ the result is

$$\varphi = [(4-n)(\lambda_n n)^{1/2}(6l)^{-1}(t_i - t)]^{2/(4-n)}\,, \tag{6.28}$$

where t_i is an integration constant. For $n = 4$ we get

$$\varphi = \varphi_i \exp[-(2/3l)\lambda_4^{1/2}(t - t_i)]\,. \tag{6.29}$$

A quasi de Sitter period will arise if the phase trajectory approaches the stationary line at $|\varphi| > 1/l$. This will happen if Δy reaches O(1) when $|\varphi| > 1/l$. Using eq. (6.19), we may rewrite this condition in terms of bounds on the initial data φ_i and $\dot{\varphi}_i$. The result is

$$\dot{\varphi}_i^2 \leq V(\varphi_i)\exp(3\sqrt{2}|l\varphi_i|)\,. \tag{6.30}$$

This bound can be understood as a constraint on the initial conditions which have to be satisfied in order to realize inflation. For inflation to correctly model the observed universe we need inflation to last more than 65 Hubble expansion times [41]. The constraint on the initial conditions required to achieve this is similar to eq. (6.30).

When the value of the scalar field drops below the Planck scale ($\varphi < 1/l$) inflation ceases. The scalar field begins to oscillate. In fig. 6.1 this corresponds to the ellipsoidal curves close to the origin. However, for this part of the phase diagram the effective equation of state depends sensitively on the potential $V(\varphi)$. To describe a concrete example we return to the model of a free scalar field with mass m given by the potential (6.20). For this choice of potential it is conventient to introduce radial and angle variables r and θ instead of y and φ in the following manner:

$$y = (\sqrt{2}/l)\sin\theta\,, \qquad \varphi = r\cos\theta\,. \tag{6.31a, b}$$

These variables satisfy equations which follow immediately from eq. (6.16)

$$\dot{\theta} = -(3l/2\sqrt{2})mr\sin 2\theta - m\,, \qquad \dot{r} = -(3l/\sqrt{2})mr^2\sin^2\theta\,. \tag{6.32a, b}$$

Eliminating r from the above system we get a closed form equation for θ,

$$\sin 2\theta\,\frac{\mathrm{d}^2\theta}{\mathrm{d}\tau^2} - 2\left(\frac{\mathrm{d}\theta}{\mathrm{d}\tau} + 1\right)\left(\frac{\mathrm{d}\theta}{\mathrm{d}\tau}\cos^2\theta + \sin^2\theta\right) = 0\,, \quad \tau = mt. \tag{6.33}$$

If we introduce the new variable $u \equiv \mathrm{d}\theta/\mathrm{d}\tau$ and regard it as a function of θ, then (6.33) reduces to

$$\mathrm{d}u/\mathrm{d}\theta = (u+1)(u\cot\theta + \tan\theta)/u\,. \tag{6.34}$$

A particular solution for this equation is $u = -1$. The general solution during the period of oscillation is close to $u=-1$. To see this more clearly let us introduce the new variable $\tilde{u}=u+1\ll 1$ and set out to prove that $\tilde{u}\to 0$. Keeping in (6.34) only terms up to second order in $\tilde{u}$ we get

$$\mathrm{d}\tilde{u}/\mathrm{d}\theta = 2\tilde{u}\cot 2\theta - \tilde{u}^2\tan\theta\,. \tag{6.35}$$

This is the Bernoulli equation whose solution is [87]

$$\tilde{u} = \sin 2\theta/(\theta + c - \tfrac{1}{2}\sin 2\theta)\,, \tag{6.36}$$

where c is an integration constant. From the above solution (6.36), it follows that as $|\theta|\to\infty$, $\tilde{u}\to 0$. Thus, as a general solution in the regime under consideration we use the solution $u=-1$ for which $\theta=\theta_0 - mt$ where θ_0 is an arbitrary phase. Substituting this into eq. (6.38) we can solve for r,

$$\begin{aligned} r &\simeq \frac{2\sqrt{2}}{3lm(t-t_0)}\left(1+\frac{1}{2m(t-t_0)}\sin 2m(t-t_0)\right)^{-1} \\ &\simeq \frac{2\sqrt{2}}{3lm(t-t_0)}\left[1-\frac{1}{2m(t-t_0)}\sin 2m(t-t_0) + \mathrm{O}\left(\frac{1}{(t-t_0)^2}\right)\right]. \end{aligned} \tag{6.37}$$

Using (6.14) and (6.31a, b) to express H in terms of r it follows that $H=(l/\sqrt{2})mr$. It is possible to integrate this expression and find the scale factor $a(t)$. The result is

$$a \propto (t-t_0)^{2/3}\left(1+\frac{\cos 2m(t-t_0)}{6m^2(t-t_0)^2} - \frac{1}{24m^2(t-t_0)^2} + \mathrm{O}(t-t_0)^{-3}]\right). \tag{6.38}$$

Thus we see that the evolution of the scale factor in a universe dominated by a massive scalar field during the period of oscillation is nearly the same (up to oscillating corrections) as if the universe were dominated by pressureless matter. However, the oscillating correction terms in (6.38) are very important since they determine the time evolution of geometrical invariants which are constructed from higher-order time derivatives of the scale factor. For example, the Hubble parameter has oscillating correction terms (compared to the evolution for a dust universe) which decay only as $(t-t_0)^{-1}$. For the scalar curvature the result is

$$R = -[4/3(t-t_0)]\{1 - 3\cos 2m(t-t_0) + \mathrm{O}[(t-t_0)^{-1}]\}\,, \tag{6.39}$$

compared to $R=-4/3t$ for a dust universe.

6.3. Perturbations

To obtain the gauge-invariant equations of motion for cosmological perturbations in the universe dominated by scalar-field matter, we insert into the general equations (4.15) the gauge-invariant energy–momentum tensor $\delta T^{(\mathrm{gi})\alpha}_{\beta}$ [see eq. (6.7)]. First of all, from the i–j $(i\neq j)$ equation it follows that we can set (as in the perfect fluid example of chapter 5) $\Phi=\Psi$, since $\delta T^{(\mathrm{gi})i}_{j}=0$ $(i\neq j)$. Substituting the energy–momentum tensor $\delta T^{(\mathrm{gi})\alpha}_{\beta}$ from (6.7) into eqs. (4.15) and setting $\Psi=\Phi$ we find (in conformal time)

$$\nabla^2\Phi - 3\mathcal{H}\Phi' - (\mathcal{H}' + 2\mathcal{H}^2)\Phi = \tfrac{3}{2}l^2(\varphi_0'\,\delta\varphi^{(\mathrm{gi})\prime} + V_{,\varphi}a^2\,\delta\varphi^{(\mathrm{gi})}) , \tag{6.40}$$

$$\Phi' + \mathcal{H}\Phi = \tfrac{3}{2}l^2\varphi_0'\,\delta\varphi^{(\mathrm{gi})} , \tag{6.41}$$

$$\Phi'' + 3\mathcal{H}\Phi' + (\mathcal{H}' + 2\mathcal{H}^2)\Phi = \tfrac{3}{2}l^2(\varphi_0'\,\delta\varphi^{(\mathrm{gi})\prime} - V_{,\varphi}a^2\,\delta\varphi^{(\mathrm{gi})}) . \tag{6.42}$$

Note that only two of these equations are independent. In obtaining this form of the equations, we have used the background relation [see (6.11)]

$$\tfrac{3}{2}l^2\varphi_0'^2 = \mathcal{H}^2 - \mathcal{H}' . \tag{6.43}$$

Equations (6.40)–(6.42) are the basic perturbation equations in the case of scalar-field matter. They were first derived in this form in ref. [103] (see also ref. [63]).

We can obtain the equation of motion for the gauge-invariant scalar-field perturbation $\delta\varphi^{(\mathrm{gi})}$ by combining eqs. (6.40)–(6.42). However, it is easier to derive this equation starting from the Klein–Gordon equation for φ,

$$\varphi^{;\alpha}_{;\alpha} + V_{,\varphi} = 0 . \tag{6.44}$$

which can be obtained by varying the action (6.1). For the background part φ_0, the equation becomes [see also eq. (6.10)]

$$\varphi_0'' + 2\mathcal{H}\varphi_0' + V_{,\varphi}(\varphi_0)a^2 = 0 . \tag{6.45}$$

The equation of motion for the scalar-field perturbation $\delta\varphi^{(\mathrm{gi})}$ can be obtained by linearizing the Klein–Gordon equation about the background solution. The result is

$$\delta\varphi^{(\mathrm{gi})\prime\prime} + 2\mathcal{H}\,\delta\varphi^{(\mathrm{gi})\prime} - \nabla^2\,\delta\varphi^{(\mathrm{gi})} + V_{,\varphi\varphi}a^2\,\delta\varphi^{(\mathrm{gi})} - 4\varphi_0'\Phi' + 2V_{,\varphi}a^2\Phi = 0 . \tag{6.46}$$

Often, the terms due to gravitational fluctuations are left out in the equation for the scalar-field perturbation, which is then written as

$$\delta\varphi'' + 2\mathcal{H}\,\delta\varphi' - \nabla^2\,\delta\varphi + V_{,\varphi\varphi}a^2\,\delta\varphi = 0 . \tag{6.47}$$

However, the terms due to metric perturbations are also linear in fluctuations and are, even during inflation, of the same order as the term $V_{,\varphi\varphi}a^2\,\delta\varphi$ which has been kept in (6.47). Thus, strictly speaking eq. (6.47) is incorrect. In many cases, the difference between (6.46) and (6.47) is significant, e.g., in eternal inflation [104]. It is very difficult to investigate (6.46) by itself. It is more convenient to return to the system of eq. (6.40)–(6.42) in order to calculate the time evolution of the fluctuations.

Subtracting (6.40) from (6.42), using (6.41) to express $\delta\varphi^{(\mathrm{gi})}$ in terms of Φ' and Φ, and taking into account (6.43) and (6.45), we obtain a second-order partial differential equation for Φ which can be written either as

$$\Phi'' + 2(\mathcal{H} - \varphi_0''/\varphi_0')\Phi' - \nabla^2\Phi + 2(\mathcal{H}' - \mathcal{H}\varphi_0''/\varphi_0')\Phi = 0 , \tag{6.48}$$

or as

$$\Phi'' + 2(a/\varphi_0')'(a/\varphi_0')^{-1}\Phi' - \nabla^2\Phi + 2\varphi_0'(\mathcal{H}/\varphi_0')'\Phi = 0\,. \tag{6.49}$$

Introducing the new variable

$$u = (a/\varphi_0')\Phi \tag{6.50}$$

instead of Φ and using the equations for the background model, we reduce (6.49) to the following equation for u:

$$u'' - \nabla^2 u - (\theta''/\theta)u = 0\,, \quad \theta = \mathcal{H}/a\varphi_0'\,. \tag{6.51a, b}$$

The solutions of eq. (6.51a) can easily be found in the asymptotic limits. Let us consider plane-wave perturbation with wavenumber k, i.e. Φ, $\delta\varphi^{(\mathrm{gi})}$, $u \propto \exp(i\boldsymbol{k}\cdot\boldsymbol{x})$. For short-wavelength perturbations which satisfy $k^2 \gg (\theta)''/\theta$ it is possible to neglect the last term in eq. (6.51a) and one obtains

$$u \propto \mathrm{e}^{\pm ik\eta}\,. \tag{6.52}$$

For long-wavelength perturbations with $k^2 \ll (\theta)''/\theta$ the second term in eq. (6.51a) can be neglected and we get

$$u \simeq C_1\theta + C_2\theta \int \frac{\mathrm{d}\eta}{\theta^2} = \frac{A}{\varphi_0'}\left(\frac{1}{a}\int^{\eta} \mathrm{d}\eta\, a^2(\eta)\right)'\,, \tag{6.53}$$

where C_1, C_2 and A are integration constants. The second integration constant in the final expression has been absorbed in the integral. The corresponding expressions for Φ and $\delta\varphi^{(\mathrm{gi})}$ following immediately from (6.50) and (6.41).

For short-wavelength perturbations we have

$$\Phi \simeq \dot{\varphi}_0\left[C_1 \sin\left(k\int a^{-1}\,\mathrm{d}t\right) + C_2 \cos\left(k\int a^{-1}\,\mathrm{d}t\right)\right]\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}}\,, \tag{6.54}$$

$$\delta\varphi^{(\mathrm{gi})} \simeq \frac{2}{3l^2}\,\frac{k}{a}\left[C_1 \cos\left(k\int a^{-1}\,\mathrm{d}t\right) - C_2 \sin\left(k\int a^{-1}\,\mathrm{d}t\right)\right]\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}}\,, \tag{6.55}$$

where the dot means the derivative with respect to physical time $t = \int a\,\mathrm{d}\eta$.

For long-wavelength perturbations one gets

$$\Phi \simeq A\left(\frac{1}{a}\int a\,\mathrm{d}t\right)^{\cdot} = A\left(1 - \frac{H}{a}\int a\,\mathrm{d}t\right)\,, \tag{6.56}$$

$$\delta\varphi^{(\mathrm{gi})} \simeq A\dot{\varphi}_0\left(a^{-1}\int a\,\mathrm{d}t\right)\,. \tag{6.57}$$

Before applying the above analysis to fluctuations in inflationary universe models, we will rewrite our main equation (6.48) to make contact with other work. For long-wavelength perturbations satisfying $k^2 \ll \theta''/\theta$, when the spatial derivative terms can be neglected, (6.48) can be recast as a "constant of motion" or "conservation law". Let us introduce the quantity ζ defined by

$$\zeta = \tfrac{2}{3}(H^{-1}\dot{\Phi} + \Phi)/(1+w) + \Phi\,, \quad w = p/\varepsilon\,. \tag{6.58}$$

Then, it is easy to verify that

$$\tfrac{3}{2}\dot{\zeta}H(1+w) = \ddot{\Phi} + (H - 2\ddot{\varphi}_0/\dot{\varphi}_0)\dot{\Phi} + 2(\dot{H} - H\ddot{\varphi}_0/\dot{\varphi}_0)\Phi\,, \tag{6.59}$$

which is (up to the term $\nabla^2\Phi$) the left-hand side of (6.48) expressed in terms of physical time rather than conformal time. According to eq. (6.48), this expression vanishes when considering wavelengths far outside the Hubble radius for which $\nabla^2\Phi$ can be neglected. Thus, on these scales ζ is conserved. This conservation law was first derived by Bardeen et al. [53], using rather different methods. It has been used in many brief reviews [105] to estimate the spectrum of density perturbations in inflationary universe models.

6.4. *Application to inflationary universe models*

One of the main applications of the theory of cosmological perturbations is to inflationary universe models. In this section we shall discuss the evolution of classical perturbations in inflationary universe models. The quantum generation of fluctuations will be discussed in the second part of this review.

First, let us recall the reason why it is possible in inflationary universe models to have a causal generation mechanism for cosmological perturbations. Figure 6.2 is a sketch of physical distance versus scale factor a. We imagine that there was a period of exponential expansion during the very early universe, for $a < a_R$. During this period, the equation of state of matter is approximately $p = -\varepsilon$, and the Hubble radius H^{-1} is constant. At a_R, the time of reheating, the vacuum energy density is assumed to convert into usual matter (massive particles and radiation) in a time interval smaller than the Hubble expansion time H^{-1}. Thereafter, the universe evolves as in the standard Big Bang model, which is to say that the universe is radiation-dominated until the time of equal matter and radiation, and matter-dominated thereafter. In the matter- and radiation-dominated periods, the Hubble radius increases more rapidly than a fixed comoving scale. In fig. 6.2 we also show a line corresponding to a fixed comoving scale on which perturbations are to be generated. The success of inflationary universe

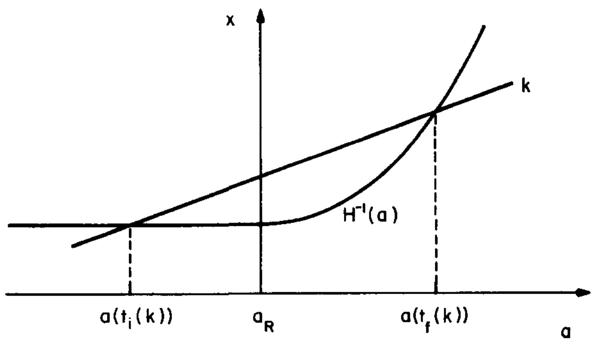

Fig. 6.2. Sketch of the evolution of length scales in inflationary universe models in terms of physical distance x versus scale factor a. The line k denotes the wavelength of a perturbation with fixed comoving wavenumber. H^{-1} is the Hubble radius. The scale of the fluctuation leaves the Hubble radius at time $t = t_i(k)$ and reenters at $t = t_f(k)$.

models consists of allowing perturbations to be generated inside the Hubble radius during the period of inflation, and then to grow to become perturbations on the scale of galaxies and clusters of galaxies today. This is possible provided that the period of inflation is longer than about $65H^{-1}$.

Models with a single period of inflation during which the Hubble parameter is constant predict a scale-invariant Harrison–Zel'dovich [47] spectrum. This is easy to see heuristically, but a rigorous derivation is not possible without including quantum considerations. The heuristic argument goes as follows. Consider a fixed comoving wavenumber k. Let $t_i(k)$ be the time when the wavelength becomes larger than the Hubble radius during the period of inflation, and let $t_f(k)$ be the time when it reenters the Hubble radius at late times. By time-translation invariance of the de Sitter phase, the time evolution of a perturbation should be independent of k when considered up to the time $t_i(k)$. This implies that the amplitude of the perturbation should be independent of k when measured at $t_i(k)$. As we shall see below, when we follow the evolution of the perturbation while outside the Hubble radius, its amplitude increases by a factor that depends only on the net change in the equation of state. Hence, the amplitude of the perturbation is practically independent of k when measured at the time $t_f(k)$. This is what is meant by a scale-invariant spectrum. In the following, we shall analyze the classical evolution of the perturbations and verify the above claim.

Applying the techniques of eqs. (6.48)–(6.57), we see that for short-wavelength fluctuations $[k^2 \gg a^2 \max(V_{,\varphi\varphi}, V_{,\varphi}^2/V, \ldots)]$, the amplitude of the metric perturbation Φ is proportional to $\dot\varphi_0$. During inflation when

$$\dot\varphi_0 \simeq -V_{,\varphi}/3H \simeq -V_{,\varphi}/3lV^{1/2}\,, \tag{6.60}$$

the time change in Φ is negligible. In particular, for the quadratic potential of eq. (6.20) the amplitude $\dot\varphi_0$ is exactly constant during inflation and hence $\Phi \propto \dot\varphi_0 \propto m$.

The amplitude of the scalar-field perturbations decreases as a^{-1}. For perturbations with a given comoving wavenumber k, this amplitude was large in the past and at some point in time linear theory breaks down since the condition $\delta\varphi/\varphi_0 < 1$ is violated. Note that the metric perturbations were always small when calculated by linear theory. However, when the wavelength of the perturbation is of the order of the Planck scale, nonlinear terms due to scalar-field perturbations in the energy–momentum tensor drive the metric perturbations to values that are larger than ones determined by linear theory. It is possible to estimate [106] this effect and show that the metric perturbations on Planckian scales become of the order unity. Thus, linear perturbation theory can be applied only when the scale of the perturbations is larger than the Planck scale.

When the universe has expanded sufficiently, the perturbations with a given wavenumber k reach the long-wave regime, $k^2 \ll a^2 \max(V_{,\varphi\varphi}/V,\ V_{\varphi}^2/V, \ldots)$, where according to (6.56),

$$\begin{aligned}\Phi &\simeq A\Big(a^{-1}\int \mathrm{d}t\, a\Big)^{\cdot} = A\Big(a^{-1}\int \mathrm{d}a\, H^{-1}\Big)^{\cdot} = A\Big(H^{-1} - \int \mathrm{d}t\, a(H^{-1})^{\cdot}\Big)^{\cdot} \\ &= A([H^{-1}]^{\cdot} - [H^{-1}[H^{-1}]^{\cdot}]^{\cdot} + [H^{-1}[H^{-1}[H^{-1}]^{\cdot}]^{\cdot}]^{\cdot} - \cdots)\,,\end{aligned} \tag{6.61}$$

$$\delta\varphi^{(\mathrm{gi})} = A\dot\varphi_0(H^{-1} - H^{-1}[H^{-1}]^{\cdot} + H^{-1}[H^{-1}[H^{-1}]^{\cdot}]^{\cdot} - \cdots)\,. \tag{6.62}$$

The expansions above were obtained as a result of integration by parts and they are asymptotic series. During inflation, when $|\dot H| \ll H^2$, we can neglect all but the first terms in eqs. (6.61) and (6.62).

Further, taking into account that during inflation $\dot{\varphi}_0^2 \ll V(\varphi)$, and $|\ddot{\varphi}_0| \ll |V_{,\varphi}(\varphi)|$ and using the background equations of motion, we find

$$\Phi \simeq -A\dot{H}/H^2 \simeq A V_{,\varphi}^2 (6l^2V^2)^{-1} , \tag{6.63}$$

$$\delta\varphi^{(\mathrm{gi})} \simeq A\, \dot{\varphi}_0/H \simeq -A\, V_{,\varphi} (3l^2V)^{-1} . \tag{6.64}$$

After the end of the period of inflation, the scalar field oscillates and decays into ultra-relativistic particles [107]. Afterwards, the scale factor grows as a power of time, i.e., $a \propto t^m$. It follows from (6.56) that the amplitude of the perturbation is

$$\Phi \simeq A/(m+1) , \tag{6.65}$$

and practically does not change in time. Here A is the same as in (6.63) and (6.64). Expressing A in terms of H, $\dot{\varphi}_0$ and $\delta\varphi^{(\mathrm{gi})}$ using (6.64), we obtain

$$\Phi \simeq (m+1)^{-1} H\, \delta\varphi^{(\mathrm{gi})}/\dot{\varphi}_0 , \tag{6.66}$$

in agreement with the results of previous investigations [50–53, 105]. Here, $\delta\varphi^{(\mathrm{gi})}/\dot{\varphi}_0$ and H have to be evaluated at the time when $k^2 = \theta''/\theta$ (which in many cases occurs when the wavelength of the perturbation crosses the Hubble radius). Note that in this formalism we do not need to make any assumptions about the duration and detailed mechanism of reheating, except to assume that we are considering scales which leave the Hubble radius before reheating and enter afterwards, and that the length of the reheating period is small compared to 50 Hubble expansion times. In synchronous gauge, tracking the perturbations through the reheating period is quite difficult, and the independence on the reheating mechanism is obscured.

For the potential $V = \frac{1}{2}m^2\varphi^2$ (eq. 6.20) the oscillating corrections to the evolution of the scale factor a [see (6.38)] lead to oscillating corrections in (6.65). The time evolution of the amplitude of the metric perturbations Φ for inhomogeneities with a given wavenumber k is depicted in fig. 6.3. We shall briefly add how the final amplitude of perturbations can be derived using the conservation law (6.58). Since at very early and very late times $\dot{\Phi}$ vanishes, (6.58) simplifies to yield

$$\Phi(t_{\mathrm{f}}) = \frac{1 + \frac{2}{3}[1 + w(t_{\mathrm{f}})]^{-1}}{1 + \frac{2}{3}[1 + w(t_{\mathrm{i}})]^{-1}}\, \Phi(t_{\mathrm{i}}) , \tag{6.67}$$

where t_{i} and t_{f} are the k-dependent times of initial and final Hubble radius crossing and $w = p/\varepsilon$. From this equation it is also clear that the amplitude of the final metric perturbations is independent of the details of reheating, since it is given in terms of the initial amplitude multiplied by a factor which depends only on the net change in the equation of state. Using (6.63) and (6.64) to express $1 + w(t_i)$ and $\Phi(t_i)$ in terms of $\dot{\varphi}_0$ and $\delta\varphi$, and applying (6.11) to replace $\dot{H}$, we obtain

$$\Phi(t_{\mathrm{f}}) \sim H\, \delta\varphi^{(\mathrm{gi})}/\dot{\varphi}_0 , \tag{6.68}$$

in agreement with (6.66).

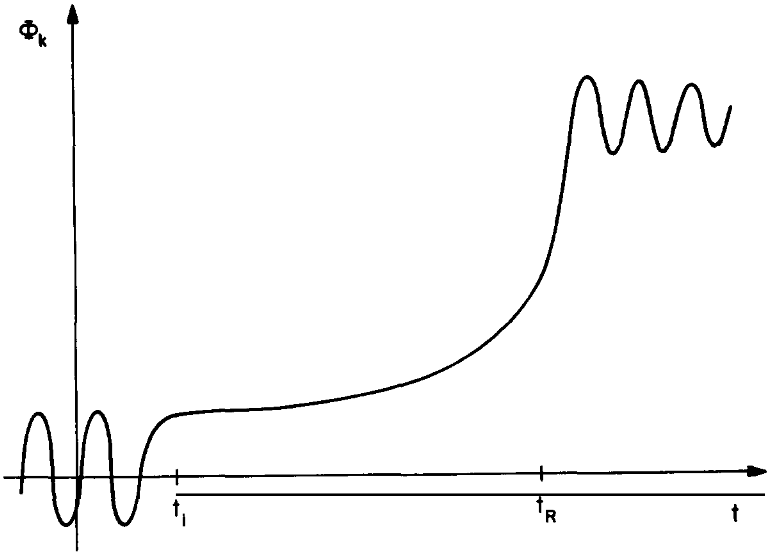

Fig. 6.3. Evolution of the gauge-invariant gravitational potential Φ for a plane-wave perturbation with wavenumber k as a function of time in an inflationary universe. t_R is the time of reheating.

Note that for long-wavelength fluctuations, both H and $\delta\varphi^{(gi)}/\dot{\varphi}_0$ depend on time, but the product does not. However, if we consider the product $H(\delta\varphi/\dot{\varphi})$ and work in a particular coordinate system, then also the product is time-dependent. Thus, when using a particular gauge to compute the initial fluctuations, care must be taken to evaluate the product at the correct time (when $k^2 = \theta''/\theta$), which often occurs at the time $t_i(k)$ when the wavelength of the perturbation leaves the Hubble radius.

7. Perturbations in higher-derivative gravity theories

We know that Einstein's general relativity theory based on an action with integrand R is a good description of space and time. However, this theory is also known to be non-renormalizable. The simplest way to render gravity renormalizable is to change the Einstein action at high energies [108]. Any change in the gravitational action will generically lead to higher derivative terms in the equations of motion. A second reason for considering higher-derivative gravity theories is related to vacuum polarization effects. Whenever coupling Einstein gravity to quantum matter fields and calculating in the semiclassical limit in which the expectation value of the energy–momentum tensor operator of the field theory is coupled to gravity, higher-derivative terms arise when calculating the expectation values. We will not specifically investigate these models in this review, and refer the reader to the literature [109].

Higher-derivative gravity has recently arisen in several different areas. It results in the low-energy limit of many superstring theories [110]. It is also used in a new realization of inflation, extended inflation [111]. The evolution of a homogeneous background in higher-order theories of gravity was studied by many authors (e.g., see ref. [112]). In higher-order derivative gravity theories, the evolution of the early universe can have some very interesting features even without matter. In particular, inflation arises in such models [49, 113] without fine tuning of initial conditions [25]. This is a major reason why higher-order gravity models are of interest in cosmology.

To simplify the analysis we use the conformal equivalence between a higher-derivative theory without matter and the usual Einstein theory with scalar-field matter. (This equivalence was first proved for R^2-gravity by Whitt [114] and for a general $F(R)$ action in refs. [115, 116].) We shall consider theories with Lagrangian densities which are a function of the scalar curvature R. This covers a lot of models which are of interest. In the first section of this chapter, we demonstrate that a theory with Lagrangian density $L = f(R)$ is conformally equivalent to Einstein's theory with a scalar field. Then, we briefly discuss the background evolution before analyzing the perturbations in the third section.

7.1. Conformal equivalence

Let us consider a gravity theory with metric $g_{\mu\nu}$ and action

$$S = (1/6l^2) \int f(R)\sqrt{-g}\, d^4x \,, \tag{7.1}$$

where $f(R)$ is an arbitrary function of the scalar curvature R and $l^2 = 8\pi G/3$. The field equations are obtained by varying the action (7.1). The result is

$$(\partial f/\partial R)R^\mu_\nu - \tfrac{1}{2}\delta^\mu_\nu f(R) + (\partial f/\partial R)^{;\alpha}_{;\alpha}\delta^\mu_\nu - (\partial f/\partial R)^{;\mu}_{;\nu} = 0 \,. \tag{7.2}$$

Our goal is to show that these vacuum gravity equations coincide with the usual Einstein equations,

$$\tilde{R}^\mu_\nu - \tfrac{1}{2}\delta^\mu_\nu \tilde{R} = 3l^2 \tilde{T}^\mu_\nu(\varphi) \,, \tag{7.3}$$

in a theory with scalar-field matter and a new metric which is conformally related to the old one, i.e.,

$$\tilde{g}_{\mu\nu} = F g_{\mu\nu} \,. \tag{7.4}$$

The correspondence will hold for appropriate choice of the conformal factor $F \equiv F(R)$ and of the potential V of the scalar field $\varphi(R)$, $V(\varphi) \equiv V(\varphi(R))$.

We now determine F, φ and $V(\varphi)$ as functions of R. Here and in the next formulas the tilde means that the corresponding variables are calculated for the conformal metric $\tilde{g}_{\mu\nu}$. Under the conformal transformation $g_{\mu\nu} \to \tilde{g}_{\mu\nu} = F g_{\mu\nu}$ the Ricci tensor and scalar curvature transform in the following manner [114]:

$$\begin{aligned} R^\mu_\nu \to \tilde{R}^\mu_\nu &= F^{-1}R^\mu_\nu - F^{-2}F_{;\nu\alpha}g^{\alpha\mu} + \tfrac{3}{2}F^{-3}F_{;\nu}F_{;\alpha}g^{\alpha\mu} - \tfrac{1}{2}F^{-2}F_{;\alpha\beta}g^{\alpha\beta}\delta^\mu_\nu \,, \\ R \to \tilde{R} &= F^{-1}R - 3F^{-2}F_{;\alpha\beta}g^{\alpha\beta} + \tfrac{3}{2}F^{-3}F_{;\alpha}F_{;\beta}g^{\alpha\beta} \,. \end{aligned} \tag{7.5}$$

Correspondingly, the Einstein equations (7.3) rewritten in terms of the metric $g_{\mu\nu}$ by substituting (7.5) and the energy–momentum tensor of a scalar field with potential $V(\varphi)$ take the form

$$\begin{aligned} &F^{-1}(R^\mu_\nu - \tfrac{1}{2}\delta^\mu_\nu R) + \tfrac{3}{2}F^{-1}((\ln F)_{;\nu}(\ln F)_{;\alpha}g^{\alpha\mu} - \tfrac{1}{2}(\ln F)_{;\alpha}(\ln F)_{;\beta}g^{\alpha\beta}\delta^\mu_\nu) \\ &\quad + F^{-2}(F_{;\alpha\beta}g^{\alpha\beta}\,\delta^\mu_\nu - F_{;\nu\alpha}g^{\alpha\mu}) \\ &= 3l^2F^{-1}[\varphi_{;\nu}\varphi_{;\alpha}g^{\alpha\mu} - \tfrac{1}{2}\delta^\mu_\nu\varphi_{;\alpha}\varphi_{;\beta}g^{\alpha\beta} + \delta^\mu_\nu F V(\varphi)] \,. \end{aligned} \tag{7.6}$$

If we set

$$\varphi = (1/\sqrt{2}l) \ln F\,, \tag{7.7}$$

then the terms in (7.6) with a tensor structure different from that of terms in eq. (7.2) cancel. Hence, eq. (7.6) can be rewritten in the form

$$F(R^\mu_\nu - \tfrac{1}{2}\delta^\mu_\nu R) + F^{;\alpha}_{;\alpha}\delta^\mu_\nu - F^{;\mu}_{;\nu} - 3l^2F^2V(\varphi)\delta^\mu_\nu = 0\,. \tag{7.8}$$

This equation coincides with (7.2) if we choose

$$F = \partial f/\partial R\,, \tag{7.9}$$

and the scalar-field potential

$$V(\varphi) = \frac{1}{6l^2}\,\frac{f(R) - R\,\partial f/\partial R}{(\partial f/\partial R)^2}\,. \tag{7.10}$$

The φ-dependence of V is obtained from (7.10) if we express R in (7.10) in terms of φ using the formulas (7.7) and (7.9).

7.2. The background model

As an example we shall describe the evolution of the background model in R^2-gravity, a model with action

$$S = -\frac{1}{6l^2}\int\left(R - \frac{1}{6M^2}R^2\right)\sqrt{-g}\,\mathrm{d}^2x\,. \tag{7.11}$$

In this model

$$f(R) = R - (1/6M^2)R^2\,, \tag{7.12}$$

and, correspondingly, from (7.9), (7.7) and (7.10) one obtains

$$F = 1 - R/3M^2\,, \qquad \varphi = (1/\sqrt{2}\,l)\ln\left(1 - \frac{R}{3M^2}\right), \tag{7.13a, b}$$

$$V(\varphi) = (M^2/4l^2)[1 - \exp(-\sqrt{2}l\varphi)]^2\,. \tag{7.14}$$

Note that when $\varphi \ll 1/\sqrt{2}l$, then $V(\varphi) \approx M^2\varphi^2/2$.

Now we shall use the conformal equivalence of R^2-gravity and ordinary gravity theory in the presence of scalar-field matter to study the evolution of the background. The results of section 6.2 for a scalar-field matter-dominated universe can then be applied. For the sake of simplicity we will consider only the case of a flat universe. The results of this and the following section can in principle easily be generalized to nonflat universes. The background metric is

$$ds^2 = a^2(\eta)(d\eta^2 - \delta_{\alpha\beta}\, dx^\alpha\, dx^\beta)\,, \tag{7.15}$$

and the corresponding conformal metric becomes

$$ds^2 = \tilde{a}^2(\eta)(d\eta^2 - \delta_{\alpha\beta}\, dx^\alpha\, dx^\beta)\,. \tag{7.16}$$

The conformal scale factor $\tilde{a}$, related to a by

$$\tilde{a}(\eta) = F_0^{1/2} a(\eta) = (1 - R/3M^2)^{1/2} a(\eta)\,, \tag{7.17}$$

satisfies the Einstein equations for the conformal theory [see (6.9)]

$$\tilde{H}^2 = l^2[\tfrac{1}{2}\dot{\varphi}^2 + V(\varphi)]\,. \tag{7.18}$$

The scalar field φ defined in (7.13a) obeys the equation [see (6.10)]

$$\ddot{\varphi} + 3\tilde{H}\dot{\varphi} + V_{,\varphi} = 0\,. \tag{7.19}$$

Here, $\tilde{H} = \dot{\tilde{a}}/\tilde{a}$ and – as we stress at this point – the dot denotes the derivative not with respect to the original physical time variable $t = \int a(\eta)\, d\eta$, but with respect to a new time variable

$$\tilde{t} = \int \tilde{a}(\eta)\, d\eta = \int F^{1/2}\, dt\,. \tag{7.20}$$

The system of background equations (7.18) and (7.19) was investigated in detail in section 6.2, and in the following we will use the results obtained there. The potential $V(\varphi)$ is not symmetric [see (7.14)]. Hence, the phase diagram (which can be constructed as in section 6.2) will not be symmetric about the y axis. Also, the condition $V_{,\varphi}/V \ll l$ will not be satisfied for any negative values of φ. Thus, the phase diagram for higher-derivative gravity will look very different for $\varphi < -1/l$. We will not study this part of the phase diagram since in this region solutions do not lead to inflation. However, for $|\varphi| \ll 1/l$ and for all positive values of φ, the phase diagram will be similar to that depicted in fig. 6.1, and thus we can use the results obtained in section 6.2. In particular, we shall use eqs. (6.24), (6.27) and (6.38) which give the conformal scale factor in terms of the scalar field φ and in terms of the time $\tilde{t}$. Note that when applied in this chapter, also the differentiation in eq. (6.27) is with respect to $\tilde{t}$.

First consider trajectories starting at a value $\varphi \gg 1/\sqrt{2}\, l$. As can be seen from the phase diagram, these trajectories rapidly approach the region in which inflation is realized. Once in this region, eq. (6.24) applies and we can integrate it for the potential of (7.14) to obtain

$$\tilde{a}(\varphi) = \tilde{a}_0 \exp[-\tfrac{3}{4}\exp(\sqrt{2}\, l\varphi) + (3l/2\sqrt{2})\varphi]\,, \tag{7.21}$$

where $\tilde{a}_0$, like a_0 in the following equation, is a constant. Using (7.17) and (7.7), we can find $a(\varphi)$,

$$a(\varphi) = a_0 \exp[-\tfrac{3}{4}\exp(\sqrt{2}\, l\varphi) + (l/2\sqrt{2})\varphi]\,. \tag{7.22}$$

To determine $a(t)$ we need to calculate the time dependence of $\varphi(t)$. To do this one can use eq. (6.27).

Rewritten in terms of physical time t instead of $\tilde{t}$ one obtains

$$\frac{dt}{d\tilde{t}}\frac{d\varphi}{dt} = -\frac{1}{3l}\frac{V_{,\varphi}}{V^{1/2}} = \frac{\sqrt{2}}{3}\frac{M}{l}\exp(-\sqrt{2}\,l\varphi)\,. \tag{7.23}$$

Taking into account that

$$dt/d\tilde{t} = F^{-1/2} = \exp(-\tfrac{1}{2}\sqrt{2}l\varphi)\,, \tag{7.24}$$

and integrating (7.23) one gets

$$\exp[(l/\sqrt{2})\varphi] = \tfrac{1}{3}M(t_s - t)\,, \tag{7.25}$$

where t_s is some integration constant with the approximate meaning of the time when inflation ceases. Substituting (7.25) into (7.22) we conclude that

$$a(t) \propto (t_s - t)^{1/2}\exp[-\tfrac{1}{12}M^2(t_s - t)^2]\,. \tag{7.26}$$

Thus, we see that for $\varphi \gg 1/\sqrt{2}\,l$ quasi-exponential expansion is realized. While $t \ll t_s$, the scale factor $a(t)$ grows like a slightly modified exponential. Inflation ends at $t \sim t_s$. In terms of physical variables, the condition $\varphi \gg 1/\sqrt{2}\,l$ necessary to obtain a quasi-de Sitter period means $|R| \gg M^2$. The limitations on the initial derivative of R required to realize inflation can be read off from (6.30).

When the field φ drops below the Planck value, the potential $V(\varphi)$ can be approximated by the quadratic potential $V(\varphi) = (1/2)M^2\varphi^2$. Thus, to analyze the background evolution in this period we can apply the results from the second part of section 6.2. The scalar field φ is oscillating, and the corresponding solution for the conformal scale factor can be obtained from (6.38),

$$\tilde{a}(\tilde{t}) \propto (\tilde{t} - \tilde{t}_0)^{2/3}\left[1 + \frac{\sin[2M(\tilde{t} - \tilde{t}_0)] - \frac{1}{4}}{6M^2(\tilde{t} - \tilde{t}_0)^2} + O\left(\frac{\sin(M\tilde{t})}{M^3(\tilde{t} - \tilde{t}_0)^3}\right)\right]. \tag{7.27}$$

For small values of φ ($|\varphi| \ll 1/l$), the time variable $\tilde{t} = \int F^{1/2}\,dt$ coincides with t up to oscillating correction terms which decay as $1/t$. This follows from Taylor-expanding F in terms of φ about $\varphi = 0$,

$$F = 1 + \sqrt{2}\,l\varphi + O(l^2\varphi^2) \simeq 1\,. \tag{7.28}$$

Recalling from (7.17) that physical and conformal scale factors are related by

$$a(t) = F^{-1/2}\tilde{a}(t) = [1 - l\varphi/\sqrt{2} + O(l^2\varphi^2)]\tilde{a}\,, \tag{7.29}$$

and using eq. (6.37) for $\varphi(\tilde{t})$, (7.27) gives

$$a(t) \propto (t - t_0)^{2/3}\left[1 - \frac{2\cos[M(t - t_0)]}{3M(t - t_0)} + O\left(\frac{\sin[M(t - t_0)]}{m^2(t - t_0)^2}\right)\right]. \tag{7.30}$$

We conclude that in contrast to the expression for $\tilde{a}(\tilde{t})$, the physical scale factor $a(t)$ in the case of higher-derivative gravity theories contains oscillating correcting terms decaying as $1/t$ and not only as t^{-2}. This is due to the factor $F^{1/2}$ connecting a and $\tilde{a}$. The corresponding Hubble parameter is

$$H(t) = a^{-1}\,\mathrm{d}a/\mathrm{d}t = \tfrac{2}{3}(t-t_0)^{-1}[1+\sin[M(t-t_0)] + \mathrm{O}(t-t_0)^{-1}]\,. \tag{7.31}$$

Let us compare the dust model, the theory with an oscillating scalar field, and the higher-derivative gravity theory discussed in this chapter. In all three models, the scale factor $a(t)$ increases as $(t-t_0)^{2/3}$. In the scalar-field theory, there is a periodic modulation with amplitude decreasing as $(t-t_0)^{-2}$. In the case analyzed above, the modulation has an amplitude which decreases only as $(t-t_0)^{-1}$. In all three models, $H(t)$ has the same overall decay rate proportional to $(t-t_0)^{-1}$. In R^2-gravity, the oscillating correction terms to $a(t)$ lead to a complete modulation of $H(t)$ which varies periodically from 0 to 1. In the oscillating scalar-field model, there is only a small modulation of the amplitude of $H(t)$, and the relative amplitude of the modulation decreases as $(t-t_0)^{-1}$. An even larger difference occurs for the Ricci curvature R. For a dust model, $R(t)$ scales as $(t-t_0)^{-2}$, for an oscillating scalar field this behavior is modified by an amplitude which oscillates in time [see (6.39)]. However, in R^2-gravity,

$$R = 4M\cos(Mt)/(t-t_0) + \mathrm{O}(t-t_0)^{-2}\,. \tag{7.32}$$

Thus, the dominant term decays only as $(t-t_0)^{-1}$. Hence, the oscillating corrections to the scale factor are very important since they determine the overall decay rates of physical invariants which are constructed from higher-order derivatives of the scale factor with respect to time.

7.3. The perturbations

Let us consider the cosmological perturbations in a higher-derivative theory of gravity with action (7.1) [115]. The full metric of the perturbed universe is [see (2.10)]

$$\mathrm{d}s^2 = a^2(\eta)\{(1+2\phi)\,\mathrm{d}\eta^2 - 2B_{,i}\,\mathrm{d}\eta\,\mathrm{d}x^i - [(1-2\psi)\delta_{ij} + 2E_{,ij}]\,\mathrm{d}x^i\,\mathrm{d}x^j\}\,. \tag{7.33}$$

To analyze the perturbations in this case, it is more convenient to work with the corresponding conformal theory based on the Einstein action and including scalar-field matter. We have already considered models with a scalar field and thus we can use the results which were obtained in the previous section. (Basically the same method has also recently been used in ref. [70] to study fluctuations in a wide class of generalized gravity theories including scalar–tensor theory, nonminimally coupled scalar-field theory and induced-gravity theory by gauge-invariant methods.)

For the perturbed conformal metric $\tilde{g}_{\mu\nu} = Fg_{\mu\nu}$ we can write down an ansatz equivalent to (7.33),

$$\mathrm{d}\tilde{s}^2 = F\,\mathrm{d}s^2 = \tilde{a}^2(\eta)\{(1+2\tilde{\phi})\,\mathrm{d}\eta^2 - 2\tilde{B}_{,i}\,\mathrm{d}x^i\,\mathrm{d}\eta - [(1-2\tilde{\psi})\delta_{ij} + 2\tilde{E}_{,ij}]\,\mathrm{d}x^i\,\mathrm{d}x^j\}\,, \tag{7.34}$$

where the conformal scale factor is $\tilde{a} = F_0^{1/2}a$. F_0 is the background value of the function $F = \partial f/\partial R$. The conformal scale factor, as was shown in the previous part of this section, must satisfy the same background equations as in the case of the scalar-field model of chapter 6. For convenience, we will write down these equations again, using conformal time,

$$\tilde{\mathcal{H}}^2 = l^2[\varphi_0'^2 + V(\varphi_0)\tilde{a}^2]\,, \qquad \tilde{\mathcal{H}}^2 - \tilde{\mathcal{H}}' = \tfrac{3}{2}l^2\varphi_0'^2\,, \tag{7.35a, b}$$

which are the exact analogs of (6.9) and (6.11). Here $\tilde{\mathscr{H}} = \tilde{a}'/a$ and φ and $V(\varphi)$ are defined in formulas (7.7) and (7.10). Comparing (7.33) and (7.34) and taking into account that in linear approximation $F = F_0[1 + (\partial \ln F_0/\partial R)\, \delta R]$ we find that the conformal potentials $\tilde{\phi}$, $\tilde{\psi}$, $\tilde{B}$, $\tilde{E}$ are connected with the potentials ϕ, ψ, B, E of the original metric (7.33) in the following manner:

$$\tilde{\phi} = \phi + (\partial \ln F_0^{1/2}/\partial R)\, \delta R\,, \quad \tilde{B} = B\,, \quad \tilde{\psi} = \psi - (\partial \ln F_0^{1/2}/\partial R)\, \delta R\,, \quad \tilde{E} = E\,, \tag{7.36}$$

where δR is the perturbation of the scalar curvature.

The conformal metric perturbations are not gauge-invariant. The corresponding gauge-invariant variables can be constructed in the standard manner [see (3.19)],

$$\begin{aligned} \tilde{\Phi} &= \tilde{\phi} + (1/\tilde{a})[(\tilde{B} - \tilde{E}')\tilde{a}]' = \Phi + (\partial \ln F_0^{1/2}/\partial R)\, \delta R^{(\mathrm{gi})}\,, \\ \tilde{\Psi} &= \tilde{\psi} - (\tilde{a}'/\hat{a})(\tilde{B} - \tilde{E}') = \Psi - (\partial \ln F_0^{1/2}/\partial R)\, \delta R^{(\mathrm{gi})}\,, \end{aligned} \tag{7.37}$$

where Φ, Ψ are the gauge-invariant potentials for the original metric (7.33) and

$$\delta R^{(\mathrm{gi})} = \delta R + R_0'(B - E') \tag{7.38}$$

is the gauge-invariant measure of the scalar curvature perturbation [see also (3.16)].

The conformal gauge-invariant potentials $\tilde{\Phi}$ and $\tilde{\Psi}$ satisfy equations which are exactly analogous to the equations for a scalar field and gravitation [see (4.15), (6.40) and (6.41)]. Of course, we need to substitute in (6.40) and (6.41) the corresponding conformal variables to obtain these equations,

$$\Delta\tilde{\Psi} - 3\tilde{\mathscr{H}}\tilde{\Psi}' - (\tilde{\mathscr{H}}' + 2\tilde{\mathscr{H}}^2)\Phi = \tfrac{3}{2} l^2 (\varphi_0'\, \delta\varphi^{(\mathrm{gi})\prime} + V_{,\varphi}\tilde{a}^2\, \delta\varphi^{(\mathrm{gi})})\,, \tag{7.39}$$

$$\tilde{\Psi}' + \tilde{\mathscr{H}}\tilde{\Phi} = \tfrac{3}{2} l^2 \varphi_0'\, \delta\varphi^{(\mathrm{gi})}\,. \tag{7.40}$$

From the i–j $(i \neq j)$ equation in (4.15) it follows that, as for ordinary scalar-field matter, $\tilde{\Phi} = \tilde{\Psi}$.

Before discussing the solutions of the equation of motion for $\tilde{\Phi}$, we need to find the explicit formulas relating the gauge-invariant potentials Φ and Ψ of the original metric to $\tilde{\Phi}$. To find these relations, we take into account that

$$\begin{aligned} \delta\varphi^{(\mathrm{gi})} = \delta\varphi + \varphi_0'(\tilde{B} - \tilde{E}') &= \frac{1}{\sqrt{2}l}\left(\frac{\partial \ln F}{\partial R}\,\delta R + \frac{\partial \ln F}{\partial R}\, R_0'(B - E')\right) \\ &= \frac{1}{\sqrt{2}l}\, \frac{\partial \ln F_0}{\partial R}\, \delta R^{(\mathrm{gi})}\,. \end{aligned} \tag{7.41}$$

Expressing $\delta R^{(\mathrm{gi})}$ in terms of $\tilde{\Phi} - \Phi$ and $\tilde{\Psi} - \Psi$ by means of (7.37) where we set $\tilde{\Phi} = \tilde{\Psi}$, and substituting the resulting expression for $\delta\varphi^{(\mathrm{gi})}$ from (7.41) into (7.40) we find the following results for Φ and Ψ:

$$\Phi = -\tfrac{2}{3}(F^2/F'a)[(a/F)\tilde{\Phi}]'\,, \qquad \Psi = \tfrac{2}{3}(1/FF'a)(aF^2\tilde{\Phi})'\,. \tag{7.42}$$

Thus, we can find Φ and Ψ if we know the solutions for $\tilde{\Phi}$. Note that the two gauge-invariant potentials for the original metric are not equal. This is a special feature of higher-derivative theories of gravity.

Returning to the equations for the gauge-invariant potential $\tilde{\Phi}$ of the conformal metric, we can, as in the case of scalar-field matter in ordinary gravity (chapter 6), obtain a single second-order differential equation for $\tilde{\Phi}$ if we substitute $\delta\varphi^{(\mathrm{gi})}$ from (7.40) into (7.39). Then, taking into account that $\tilde{\Psi} = \tilde{\Phi}$ and using the background equations (7.35a, b) we obtain the equation

$$\tilde{\Phi}'' + 2(\tilde{a}/\varphi_0')'(\tilde{a}/\varphi_0')^{-1}\tilde{\Phi}' - \nabla^2\tilde{\Phi} + 2\varphi_0'(\tilde{\mathcal{H}}/\varphi_0')'\tilde{\Phi} = 0\,, \tag{7.43}$$

which has exactly the same form as (6.48). The solutions of (6.48) were investigated in detail in chapter 6 and we shall use the results which were obtained there. Thus, if we introduce the new variable

$$\tilde{u} = (\tilde{a}/\varphi_0')\tilde{\Phi} = \sqrt{2}\,l(aF^{3/2}/F')\tilde{\Phi}\,, \tag{7.44}$$

then $\tilde{u}$ satisfies the equation

$$\tilde{u}'' - \Delta\tilde{u} - (\theta''/\theta)\tilde{u} = 0\,, \quad \theta = \tilde{\mathcal{H}}/\tilde{a}\varphi_0' = \sqrt{2}\,l(aF_0^{1/2})'/a^2F'\,. \tag{7.45a, b}$$

The solutions of (7.45a) can be easily found in the asymptotic short- and long-wavelength limits. For a plane-wave perturbation with $\tilde{u} \propto \mathrm{e}^{ikx}$, the short-wave limit (for $k^2 \gg \theta''/\theta$) gives

$$\tilde{u} \propto C\,\mathrm{e}^{ik\eta} + \mathrm{c.c.} = C\exp\Big(ik\int a^{-1}\,\mathrm{d}t\Big) + \mathrm{c.c.} \tag{7.46}$$

where C is a constant and c.c. stands for the complex conjugate. In the long-wave limit ($k^2 \ll \theta''/\theta$) the result is

$$\tilde{u} \propto \tilde{A}\theta\int\frac{\mathrm{d}\eta}{\theta^2} + \tilde{B}\theta = A\Big(\frac{aF^{3/2}}{F'} - \frac{(aF^{1/2})'}{a^2F'}\int a^2F\,\mathrm{d}\eta\Big) = A\Big(\frac{F^{3/2}}{\dot{F}} - \frac{(aF^{1/2})^{\cdot}}{a^2\dot{F}}\int aF\,\mathrm{d}t\Big)\,. \tag{7.47}$$

Here, the physical time $t = \int a\,\mathrm{d}\eta$ has been introduced; and dot means the differentiation with respect to t. To simplify the integral in (7.47) eq. (7.35b) was used in the form

$$F'^2 = -\tfrac{4}{3}aF^{5/2}[(aF^{1/2})'/a^2F]'\,. \tag{7.48}$$

Let us now express the potentials Φ and Ψ in terms $\tilde{u}$. Using the relation (7.44) between $\tilde{\Phi}$ and $\tilde{u}$, (7.42) becomes

$$\Phi = -\frac{1}{3\sqrt{2}lF^{1/2}}\Big[\Big(\frac{\ddot{F}}{\dot{F}} - \frac{5}{2}\frac{\dot{F}}{F} + H\Big)\tilde{u} + \dot{\tilde{u}}\Big]\,, \qquad \Psi = \frac{1}{3\sqrt{2}lF^{1/2}}\Big[\Big(\frac{\ddot{F}}{\dot{F}} + \frac{1}{2}\frac{\dot{F}}{F} + H\Big)\tilde{u} + \dot{\tilde{u}}\Big]\,, \tag{7.49}$$

where $H = \dot{a}/a$. Taking into account (7.37) and using $\tilde{\Phi} = \tilde{\Psi}$, one can express $\delta R^{(\mathrm{gi})}$ in terms of Φ and Ψ,

$$\delta R^{(\mathrm{gi})} = \tfrac{1}{2}(\partial\ln F^{1/2}/\partial R)^{-1}(\Psi - \Phi)\,. \tag{7.50}$$

Now, substituting $\tilde{u}$ from (7.47) and (7.48) in (7.49), it follows that for the short-wavelength perturbations ($k^2 \gg \theta''/\theta$)

$$\Phi = -\frac{1}{F^{1/2}}\left[\left(\frac{\ddot F}{\dot F} - \frac{5}{2}\frac{\dot F}{F} + H + \frac{ik}{a}\right) c \exp\left(ik \int a^{-1}\,dt\right) + \text{c.c.}\right],$$
$$\Psi = +\frac{1}{F^{1/2}}\left[\left(\frac{\ddot F}{\dot F} - \frac{1}{2}\frac{\dot F}{F} + H + \frac{ik}{a}\right) c \exp\left(ik \int a^{-1}\,dt\right) + \text{c.c.}\right], \tag{7.51}$$

and, correspondingly, for long-wavelength perturbations ($k^2 \ll \theta''/\theta$),

$$\Phi = A\left(\frac{1}{aF}\int aF\,dt\right)^{\cdot}, \qquad \Psi = \Phi + A\,\frac{\dot F}{aF^2}\int aF\,dt\,. \tag{7.52}$$

The solutions (7.51) and (7.52) can be applied to describe the evolution of cosmological perturbations in theories of gravity with action $S = (1/6l^2)\int f(R)\sqrt{-g}\,d^4x$, where f is an arbitrary function of the scalar curvature R and determines F via $F = \partial f/\partial R$.

Let us analyze in detail the behavior of perturbations during the quasi-de Sitter stage (7.21) in R^2 gravity. As in the case of scalar-field perturbations, there are two distinguished physical length scales. The first is the Hubble radius H^{-1}, the second (and larger) one is given by the inverse mass M^{-1}. Taking into account that during this stage

$$\dot H \approx -\tfrac{1}{6}M^2 \ll H^2\,, \tag{7.53}$$

we obtain (after integration by parts) the following for the long-wavelength limit, in which the length scale exceeds the larger of the above two distinguished lengths in this model ($k^2 \ll \frac{2}{3}M^2a^2$),

$$\Phi \simeq A\left[\frac{1}{6}\,\frac{M^2}{H^2(t)} + O\left(\frac{M^4}{H^4}\right)\right], \qquad \Psi \simeq -A\left[\frac{1}{6}\,\frac{M^2}{H^2(t)} + O\left(\frac{M^4}{H^4}\right)\right],$$
$$\frac{\delta R^{(\text{gi})}}{R} \simeq -A\left[\frac{1}{3}\,\frac{M^2}{H^2(t)} + O\left(\frac{M^4}{H^4}\right)\right]. \tag{7.54}$$

For the case of short-wavelength perturbations ($k^2 \gg \frac{2}{3}M^2a^2$), the terms $\ddot F/\dot F$ and $\dot F/F$ in (7.51) can be neglected during the quasi-de Sitter period, and we find

$$\Phi \simeq -\Psi \propto \left(1 + \frac{ik}{Ha}\right)\exp\left(i\int \frac{k}{a}\,dt\right) + \text{c.c.}\,, \qquad \frac{\delta R^{(\text{gi})}}{R} \simeq -2\Phi\,. \tag{7.55}$$

As shown in ref. [70], the difference between Φ and Ψ in higher-derivative gravity theories is due to anisotropic pressure terms which arise in these models.

Let us now follow the amplitude of a perturbation whose wavelength starts out much smaller than the Hubble radius during the quasi-de Sitter period. According to (7.55), the amplitude of the metric perturbations decays as $1/a(t)$ up to when the length scale crosses the Hubble radius ($k \sim Ha$), since the decaying mode in (7.55) has a larger amplitude. According to the same equation, the amplitude then freezes at a value given by $\Phi \simeq -\Psi = \text{constant}$ until the physical length scale reaches the value M^{-1} ($k \sim Ma$). From that point on, the amplitude of the long-wavelength metric perturbations ($k \ll Ma$) increases as $H^{-2}(t)$ [$H(t)$ decreases during the quasi-de Sitter period]. The increase is cut off when the quasi-de Sitter period ends and the period of oscillation of the curvature begins.

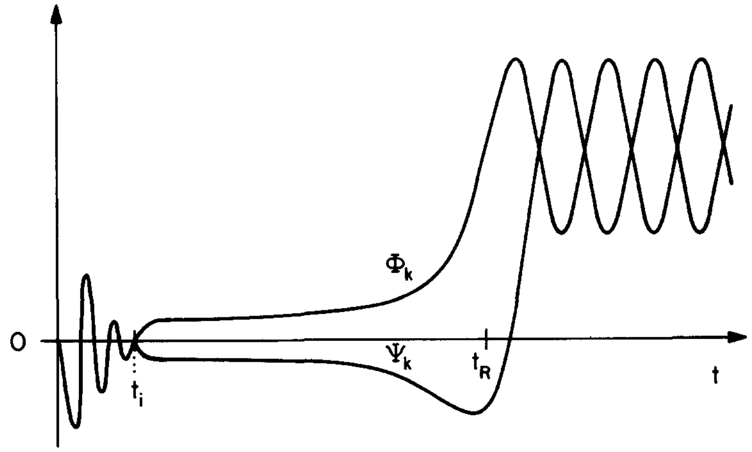

Fig. 7.1. Evolution of the two gauge-invariant gravitational potentials Φ and Ψ for a plane-wave perturbation with wavenumber k as a function of time in a higher-derivative theory of gravity during and after inflation.

To analyze the time evolution of the perturbations during the period of oscillation, we substitute the result for the scale factor from eq. (6.38) into (7.52) and take into account that during this period $F \approx 1$. Then we obtain for the long-wavelength perturbations,

$$\Phi \simeq A\{\tfrac{3}{5} - \tfrac{2}{5}\cos[M(t-t_1)]\}\,, \qquad \Psi \simeq A\{\tfrac{3}{5} + \tfrac{2}{5}\cos[M(t-t_1)]\}\,, \tag{7.56}$$

where the amplitude A is the same as in (7.54). We see that the increase in the amplitude of the metric perturbations stops. The amplitude contains, in addition to a constant term for which $\Phi = \Psi$, a piece oscillating with a constant amplitude. The oscillating part of the metric perturbations will presumably decay only if we take into account the fact that reheating damps out the oscillations [107]. The oscillating terms in the Φ and Ψ equations are shifted after half the oscillation period. This effect may be significant for reheating. The time dependence of Φ and Ψ is represented schematically in fig. 7.1.

Let us stress the interesting features of the perturbations in R^2-gravity. During the quasi-exponential period of expansion, both the short-wavelength and long-wavelength fluctuations are conformally flat, i.e., $\Phi = -\Psi$, in leading order. This is to be distinguished from scalar-field models where $\Phi = \Psi$. When the period of oscillation starts, the nonoscillating part of the long-wavelength perturbations becomes conformal–Newtonian, i.e. $\Phi = \Psi$, as in the hydrodynamical and scalar-field models. However, the short-wavelength fluctuations always remain conformally flat to leading order, as can be seen from (7.49).

8. Conclusions

In the first part of this review article we have presented the theory of classical cosmological perturbations. It is important to distinguish physical perturbations from gauge-mode changes in the metric which are due to a change in the background coordinates of space–time. We have argued that the best way to achieve this separation is by eliminating the gauge modes altogether by adopting a gauge-invariant formalism.

In chapter 3 we defined simple gauge-invariant variables Φ and Ψ and discussed their physical meaning. Φ is the generalization of the Newtonian gravitational potential and is proportional to δg_{00} in longitudinal gauge. Chapter 4 contains a new derivation of a closed set of equations for the gauge-invariant measures of cosmological perturbations.

In chapter 5, we applied this set of equations to perfect fluid perturbations. In this case, the two gauge-invariant metric perturbations coincide. The perturbation equations can be combined to give a single second-order partial differential equation for Φ. In the case of adiabatic perturbations, this equation is homogeneous; for entropy perturbations, there is a source term proportional to $\delta S/S$. We derived the solutions for Φ for a universe containing both cold matter and radiation. For adiabatic perturbations, both Φ and the gauge-invariant energy density fluctuation $\delta\varepsilon^{(gi)}/\varepsilon$ remain constant if we consider inhomogeneities with wavelength larger than the Hubble radius. Once the wavelength drops below $H^{-1}(t)$, the density contrast $\delta\varepsilon^{(gi)}/\varepsilon$ in the matter-dominated period starts to increase as η^2, with η being the conformal time. In an open universe ($\mathcal{K} = -1$), $\delta\varepsilon^{(gi)}/\varepsilon$ freezes out when $\eta > 1$. In the case of entropy perturbations, it was shown that on length scales which enter the Hubble radius after the time of equal matter and radiation η_{eq}, both Φ and $\delta\varepsilon^{(gi)}/\varepsilon$ increase before η_{eq}, whereas the magnitude of the matter energy-density fluctuation (a measure of the entropy perturbation) is frozen in. $\delta\varepsilon^{(gi)}/\varepsilon$ catches up to $\delta\varepsilon_m/\varepsilon_m$ by the time η_{eq}. Thereafter, the evolution is as for adiabatic perturbations.

The gauge-invariant equations for cosmological perturbations were applied in chapter 6 to a universe filled with scalar-field matter. First, we presented a new analysis of the background model which shows that for a large range of initial data and for suitable potentials, the background will go through a period of inflation. Explicit equations for the scale factor $a(t)$ during inflation and during the subsequent period when the scalar field oscillates were derived. For long-wavelength perturbations, analytical formulas can be obtained for the non-decaying mode of the fluctuations. These results were applied to the study of the evolution of classical perturbations in inflationary universe models. The amplitudes of Φ after inflation and during inflation are related by a factor which depends only on the change in the equation of state.

Finally, in chapter 7 we discussed a class of higher-derivative gravity models. By a conformal transformation, the system can be recast as a model with Einstein gravity plus scalar-field matter. This enables a simple discussion of the background evolution and perturbation equations based on the analysis of chapter 6. In particular, the gauge-invariant approach to cosmological perturbations can immediately be extended to higher-derivative theories. However, after the conformal transformation back to the original variables, there are some important differences. For the background model, small oscillating correction terms in the scale factor $a(t)$ (compared to a dust model) crucially alter the time evolution of invariants, like the Ricci scalar R, which depend on higher derivatives of $a(t)$. The perturbations are no longer conformally–Newtonian ($\Phi = \Psi$), in contrast to the theories discussed in chapters 5 and 6.

We hope to have shown that the gauge-invariant approach to cosmological perturbations is both physically more appealing and mathematically simpler than analyses in a particular gauge, e.g., synchronous gauge.

PART II. QUANTUM THEORY OF PERTURBATIONS

9. Introduction

In this part of the review article we develop a consistent quantum theory of cosmological perturbations. This involves the simultaneous quantization of metric and matter fluctuations. Since this procedure requires a nonvanishing matter component, it is only possible in an expanding universe, not in Minkowski space, in contrast to the quantum theory of gravitational waves. Quantizing matter fields in a nontrivial background leads in general to particle production. This process is the basis for structure formation in inflationary universe models and thus provides the initial conditions for the classical evolution of perturbations discussed in part I. Recently, there has been a lot of progress in understanding initial conditions. Successful cosmological models have been developed based on the premise that quantum fluctuations are responsible for the observed large-scale structure of the universe today. The theory of inflation [41] is one of the most promising of these models.

The basis for quantization is the canonical commutation relations. In order to define them, we require canonical momenta and hence the action. It is incorrect to start simply with a classical equation of motion and try to interpret it quantum mechanically; this would in general lead to an incorrect normalization of the modes (see also ref. [117]). It is important to quantize only the physical degrees of freedom. Hence, the gauge-invariant formalism is very useful.

In the following, both the generation and evolution of the fluctuations will be discussed in a unified treatment. Our aim is to be able to calculate the spectrum of metric and density perturbations starting with initial quantum fluctuations in general cosmological models, and in particular in the cases analyzed in part I.

Since the small fluctuations are Gaussian, the computation of metric and density perturbations can be reduced to the determination of two-point correlation functions and power spectra [42, 52, 50, 35, 118]. Since this fact is of central importance, we shall discuss it at this point following the treatment in ref. [119]. The main physical observable connected with density perturbations (scalar metric perturbations) is the root-mean-square relative mass fluctuation $(\overline{\delta M/M})(r)$ inside a sphere of radius r. This observable can be expressed in terms of the relative density perturbation $\delta\tilde{\varepsilon}(k)/\varepsilon$ in Fourier space, which in turn can be expressed in terms of the two-point correlation function.

To derive the relation, let us express the mass fluctuation inside a sphere S_{r,x_0} of radius r centered at the point $\boldsymbol{x}_0$,

$$\delta M_{x_0}(r) = \int_{S_{r,x_0}} \mathrm{d}^3x\, \delta\varepsilon(\boldsymbol{x}) = \int \frac{\mathrm{d}^3k'}{(2\pi)^{3/2}} V^{1/2}\, \delta\tilde{\varepsilon}(k') \int_{S_{r,x_0}} \mathrm{d}^3x \exp[\mathrm{i}\boldsymbol{k}'\cdot(\boldsymbol{x}-\boldsymbol{x}_0)]\,, \tag{9.1}$$

where V is the total volume of the system and where for simplicity $\mathcal{K}=0$ has been assumed. Taking the square of (9.1), averaging over $\boldsymbol{x}_0$ and dividing by the square of the average mass inside the sphere S_r gives [119]

$$\left|\frac{\overline{\delta M}}{M}\right|^2(r) = \int \mathrm{d}^3k' \left|\frac{\delta\tilde{\varepsilon}(k')}{\varepsilon}\right|^2 |W_r(\boldsymbol{k}')|^2\,, \quad W_r(k') = \frac{\int_{S_r} \mathrm{d}^3x\, \mathrm{e}^{ik'\cdot x}}{\int_{S_r} \mathrm{d}^3x}\,, \tag{9.2a, b}$$

where $W_r(k')$ is a window function which vanishes for $k'r \gg 1$ and which is 1 for $k'r \ll 1$. Provided that $|\delta\tilde{\varepsilon}(k')/\varepsilon|^2$ does not increase faster than k'^{-3} as $k' \to 0$, the integral on the right-hand side of (9.2a) is dominated by $k \sim r^{-1}$ and hence

$$|\overline{\delta M}/M|^2(k^{-1}) \sim k^3|\delta\tilde{\varepsilon}(k)/\varepsilon|^2 \equiv |\delta_\varepsilon(k)|^2\,, \tag{9.3}$$

where $|\delta_\varepsilon(k)|$ characterizes the amplitude of the perturbations. In the following we will refer to this as the power spectrum. However, in much of the literature on large-scale structure, $k^{-3/2}\delta_\varepsilon(k)$ is defined as the power spectrum. On the other hand, the two-point energy-density correlation function

$$\xi_\varepsilon(r) = \overline{\frac{\delta\varepsilon}{\varepsilon}(x)\,\frac{\delta\varepsilon}{\varepsilon}(x+r)} \tag{9.4}$$

can be expanded in terms of the Fourier space quantity $\delta\tilde{\varepsilon}(\boldsymbol{k})$. After integrating over the angular part of k we obtain

$$\xi_\varepsilon(r) = 4\pi \int_0^\infty \frac{\mathrm{d}k}{k}\,\frac{\sin kr}{kr}\,|\delta_\varepsilon(k)|^2 \,. \tag{9.5}$$

In a quantum theory, the classical variables become Heisenberg operators, and averaging means taking the expectation values of the operators in the quantum state of the system. Thus it follows that in a quantum theory of cosmological perturbations, the relevant quantities to calculate are the two-point correlation functions of the operators corresponding to the classical physical observables. In this part of our review article, the focus will be on evaluating these two-point correlation functions.

The quantization principle used will be quite conventional. After deriving the relevant action for cosmological perturbations, the resulting theory will be quantized using canonical quantization. The way in which this standard quantization scheme will be applied, however, is new, quite simple and does eliminate ambiguities that exist in other approaches.

As a first step, the action for cosmological perturbations will be written in terms of a single gauge-invariant variable v which satisfies a particularly simple equation (a wave equation with an added potential term). Next, the physical observables (the gauge-invariant metric perturbation variable Φ and the energy density ε) are expressed in terms of v. The functions which relate Φ (ε) and v depend on the background cosmological model. The third step in our method is standard canonical quantization of the action written in terms of v. The operator $\hat{v}$ corresponding to v is expanded in creation and annihilation operators of the modes of the equation for v. The coefficient functions in this expansion will satisfy the same equations as the classical perturbation variables. This fact establishes a deep connection between parts I and II. The calculation of the correlation functions in order to determine the power spectra thus reduces to the evaluation of expectation values of products of creation and annihilation operators in the quantum state of the system.

Our approach has several important advantages. It is simple (since it only involves one quantization variable), physically unambiguous (since it is a gauge-invariant formalism), and allows a unified picture of cosmological perturbations. The analysis gives both the generation and evolution of cosmological perturbations. The approach applies to a wide variety of cosmological models: to standard Friedmann cosmologies with hydrodynamical perturbations, to scalar-field driven inflationary models, and to higher-derivative gravity models.

The original idea that quantum fluctuations in an expanding universe could lead to classical density perturbations can be traced back to Sakharov [44]. The first concrete formalism was developed independently by Chibisov and Mukhanov [42] and by Lukash [43]. It was in particular realized that in inflationary universe models quantum fluctuations during the phase of exponential expansion would reenter the Hubble radius at later times with a scale-invariant spectrum [45, 46, 42]. Mukhanov and Chibisov [48] gave the first quantitative calculation of the spectrum of density perturbations. They considered a model [49] in which higher-derivative gravity actions lead to inflation. The spectrum of density perturbations in New Inflationary Universe models was first estimated in refs. [54, 50–53].

The approach outlined in the following chapters has a major advantage in that generation and evolution of the perturbations are discussed in a unified way. This approach was developed in refs. [42, 54] (see also ref. [43]) and recast in gauge-invariant form in refs. [20, 120]. For quite a similar analysis see refs. [21, 121] (a rather different treatment was presented in ref. [122]). In much of the earlier work [53, 52, 118, 58] (see, e.g., reviews in refs. [119, 105]), an artificial separation between the period of quantum generation and classical evolution of perturbations was used. For a similar unified analysis (albeit in synchronous gauge) see ref. [123]. For an analysis of metric fluctuations in the context of quantum cosmology see ref. [124].

The outline of part II of this review is as follows. In chapter 10, the action for cosmological perturbations is reduced to an action of a single gauge-invariant variable v. The analysis is done for hydrodynamical, scalar field, and higher-derivative gravity perturbations. This chapter is in principle straightforward but technically quite tedious. The reader interested in the physical results is advised to skip this chapter and to refer back to the crucial results as required. Chapter 11 contains the application of the standard canonical quantization method to the reduced actions of chapter 10 for cosmological perturbations. Chapters 12–14 contain the key results. We calculate the power spectra in models with hydrodynamical perturbations (chapter 12), in theories of inflation driven by a scalar field (chapter 13) and in higher-derivative gravity models (chapter 14).

10. Variational principles

The quantum theory of cosmological perturbations discussed here is the quantization of the first-order metric and matter perturbations about a homogeneous and isotropic background. Thus, the original configuration space variables are the metric and matter perturbation variables discussed in the first part of this article. In order to obtain the action for them, we begin with the initial action

$$S = -\frac{1}{16\pi G}\int R\sqrt{-g}\,\mathrm{d}^4x + \int \mathscr{L}_{\mathrm{m}}(g)\sqrt{-g}\,\mathrm{d}^4x \tag{10.1}$$

for gravity and matter, and expand it up to second order in the perturbation variables since the first-order perturbation equations of motion are given by the second-order action. Note that the first-order terms vanish when expanding about a background solution which satisfies the equations of motion.

In this chapter, a derivation of the full action of matter and gravity to second order in the perturbation variables will be given [63]. $\delta_n Q$ will denote the terms in some variable Q of nth order in fluctuation variables. In section 10.1, the pure gravitational contribution $\delta_2 S_{\mathrm{gr}}$ to the action S will be derived. In the following sections, the contributions to $\delta_2 S$ which include matter perturbations are analyzed for hydrodynamical matter (section 10.2) [42] and for scalar-field matter (section 10.3) [20]. Section 10.4 is an extension of the formalism to higher-derivative theories of gravity [120].

In all cases we shall reduce the action for perturbations to the simplest form in which it is described in terms of a single gauge-invariant variable characterizing both metric and matter perturbations. The reduction of the action utilizes the constraint equations. Based on the reduced action, the canonical variables for quantization will be found.

10.1. The gravitational part of the action

To derive the purely gravitational part $\delta_2 S_{gr}$ of the action $\delta_2 S$ it proves convenient to use the ADM formalism [125] since this significantly simplifies the rather lengthy calculations. In this formalism the metric is written in the form

$$ds^2 = (\mathcal{N}^2 - \mathcal{N}_i \mathcal{N}^i)\,d\eta^2 - 2\mathcal{N}_i\,dx^i\,d\eta - \gamma_{ij}\,dx^i\,dx^j\,. \tag{10.2}$$

Here, $\mathcal{N}$ is the lapse and $\mathcal{N}_i$ is the shift vector and γ_{ij} is the metric on the constant-η hypersurface.

In terms of the ADM metric (10.2), the Einstein action takes the form (see ref. [125] for a derivation)

$$S_{gr} = -\frac{1}{16\pi G}\int R\sqrt{-g}\,d^4x$$

$$= \frac{1}{16\pi G}\int [\mathcal{N}\gamma^{1/2}(K^i_j K^j_i - K^2 + {}^{(3)}R) - 2(\gamma^{1/2}K)' + 2(\gamma^{1/2}K\mathcal{N}^i - \gamma^{1/2}\gamma^{ij}\mathcal{N}_{,j})_{,i}]\,d^4x\,, \tag{10.3}$$

$$K_{ij} = \tfrac{1}{2}\mathcal{N}^{-1}(\mathcal{N}_{i|j} + \mathcal{N}_{j|i} - \gamma'_{ij})\,, \tag{10.4}$$

where K_{ij} is the extrinsic curvature tensor of the $\Sigma: \eta = \text{constant}$ hypersurface. A prime denotes derivative with respect to conformal time η, a comma represents a partial derivative and the vertical bar stands for the covariant derivative with respect to the spatial metric γ_{ij}. Also, $K \equiv K^i_i$ and $\gamma = \det(\gamma_{ij})$. We use γ_{ij} and its inverse γ^{ij} to raise and lower spatial indices. Finally, ${}^{(3)}R$ is the scalar curvature of the hypersurface Σ. The reduced scalar curvature ${}^{(3)}R$ can be written in terms of the metric γ_{ij} and its first and second derivatives. As derived, e.g., by Fock (eqs. (B.49) and (B.50) in ref. [126]), the second derivatives of γ_{ij} only enter as a total derivative term and (10.3) can be written as

$$S_{gr} = \frac{1}{16\pi G}\int [\mathcal{N}\gamma^{1/2}(K^i_j K^j_i - K^2) + \tfrac{1}{2}(\gamma^{1/2}\gamma^{ij}\mathcal{N})_{,i}(\ln\gamma)_{,j}$$

$$+ \mathcal{N}_{,i}(\gamma^{1/2}\gamma^{ij})_{,j} - \tfrac{1}{2}\mathcal{N}\gamma^{1/2\,(3)}\Gamma^l_{ij}\gamma^{ij}_{,l} + \mathcal{D}^{gr}_1]\,d^4x\,, \tag{10.5}$$

$$\mathcal{D}^{gr}_1 = -2(\gamma^{1/2}K)' + 2(\gamma^{1/2}K\mathcal{N}^i - \gamma^{1/2}\gamma^{ij}\mathcal{N}_{,j})_{,i} - [\mathcal{N}\gamma^{ij}(\gamma^{1/2})_{,j} + \mathcal{N}(\gamma^{1/2}\gamma^{ij})_{,j}]_{,i} \tag{10.6}$$

where $\mathcal{D}^{gr}_1$ is a total derivative term which does not affect the equations of motion.

Now we will expand this action in terms of the metric variables used in part I. For simplicity, the case of a flat universe ($\mathcal{K} = 0$) will be considered. It is not difficult to generalize the results to the cases $\mathcal{K} = \pm 1$. Comparing the metrics (10.2) and (2.10) we find

$$\mathcal{N}_i = a^2 B_{,i}\,, \qquad \gamma_{ij} = a^2(1 - 2\psi)\delta_{ij} + 2a^2 E_{,ij}\,, \qquad \mathcal{N} = a(1 + \phi - \tfrac{1}{2}\phi^2 + \tfrac{1}{2}B_{,i}B_{,i})\,. \tag{10.7}$$

(In all expressions, only terms up to second order in the perturbation variables will be kept.) The inverse metric γ^{ij} is

$$\gamma^{ij} = a^{-2}(\delta_{ij} + 2\psi\delta_{ij} - 2E_{,ij} + 4\psi^2\delta_{ij} + 4E_{,il}E_{,lj} - 8E_{,ij}\psi)\,, \tag{10.8}$$

where summation over repeated lower indices is implicit. From these equations, follows

$$\gamma^{1/2} = a^3(1 - 3\psi + \tfrac{3}{2}\psi^2 + E_{,ii} + \tfrac{1}{2}E_{,ii}E_{,jj} - E_{,ij}E_{,ji} - E_{,ii}\psi) ,$$
$$^{(3)}\Gamma^l_{ij} = \psi_{,l}\delta_{ij} - \psi_{,i}\delta_{lj} - \psi_{,j}\delta_{li} + E_{,lij} + t_2 . \tag{10.9}$$

Here, t_2 denotes terms of second order in the perturbation variables which we will not need since they lead to third-order terms in the action. We also get

$$K^i_j = -a^{-1}[\delta_{ij}(\mathcal{H} - \mathcal{H}\phi - \psi') - (B - E')_{,ij} + \delta_{ij}(\tfrac{3}{2}\mathcal{H}\phi^2 + \phi\psi' - 2\psi\psi') + 2(E_{,ij}\psi)' + \phi B_{,ij} - \phi E'_{,ij}$$
$$- 2\psi B_{,ij} + 2E_{,il}B_{,lj} - 2E_{,il}E'_{,lj} + \delta_{ij}\psi_{,l}B_{,l} - \psi_{,i}B_{,j} - \psi_{,j}B_{,i} + E_{,ijl}B_{,l} - \tfrac{1}{2}\mathcal{H}\delta_{ij}B_{,l}B_{,l}] . \tag{10.10}$$

Inserting (10.7)–(10.10) into (10.5), we can determine the terms in the Einstein action which are quadratic in the perturbation variables. The calculations are straightforward but very long. The final result is

$$\delta_2 S_{\rm gr} = \frac{1}{16\pi G}\int \{a^2[-6\psi'^2 - 12\mathcal{H}(\phi + \psi)\psi' - 9\mathcal{H}^2(\phi + \psi)^2$$
$$- 2\psi_{,i}(2\phi_{,i} - \psi_{,i}) - 4\mathcal{H}(\phi + \psi)(B - E')_{,ii}$$
$$+ 4\mathcal{H}\psi' E_{,ii} - 4\psi'(B - E')_{,ii} - 4\mathcal{H}\psi_{,i}B_{,i} + 6\mathcal{H}^2(\phi + \psi)E_{,ii} - 4\mathcal{H}E_{,ii}(B - E')_{,jj}$$
$$+ 4\mathcal{H}E_{,ii}B_{,jj} + 3\mathcal{H}^2E^2_{,ii} + 3\mathcal{H}^2B_{,i}B_{,i}] + \mathcal{D}^{\rm gr}_1 + \mathcal{D}^{\rm gr}_2\}\,{\rm d}^4x , \tag{10.11}$$

where $\mathcal{D}^{\rm gr}_2$ is another total derivative term,

$$\mathcal{D}^{\rm gr}_2 = \{a^2[-8\mathcal{H}(E_{,ij}(B - E')_{,j} - E_{,jj}(B - E')_{,i} + \tfrac{1}{2}E_{,jj}B_{,i})$$
$$+ 6\mathcal{H}^2(E_{,ij}E_{,j} - E_{,jj}E_{,i}) + (B - E')_{,ij}(B - E')_{,j}$$
$$- (B - E')_{,jj}(B - E')_{,i} + E_{,ijl}E_{,jl} - E_{,jjl}E_{,li}]\}_{,i} . \tag{10.12}$$

It will be convenient to combine (10.11) with the matter part of the action. Thus, we will proceed to the discussion of the action for matter.

10.2. The action for hydrodynamical matter

For hydrodynamical matter, the action $S_{\rm m}$ is given by [126]

$$S_{\rm m} = -\int \varepsilon\sqrt{-g}\,{\rm d}^4x , \tag{10.13}$$

where ε is the energy density. Unlike in the case of scalar-field matter, it is not completely straightforward to extract the contribution $\delta_2 S_m$ which is second order in perturbation variables. We begin this section by bringing S_m into a form in which the perturbation expansion is manifest.

To start out let us recall some basic expressions for the background (see part I, chapters 2 and 5). The Friedmann–Robertson–Walker equations for hydrodynamical matter can be written as

$$\varepsilon_0 = a^{-2} l^{-2} (\mathcal{H}^2 + \mathcal{K}) , \qquad p_0 = a^{-2} l^{-2} (\tfrac{2}{3}\beta - \mathcal{H}^2 - \mathcal{K}) , \tag{10.14a, b}$$

where in order to simplify future equations the following notation has been introduced:

$$\beta = \mathcal{K} + \mathcal{H}^2 - \mathcal{H}' . \tag{10.15}$$

l is the Planck length

$$l = (\tfrac{8}{3}\pi G)^{1/2} = 4.7 \times 10^{-33}\ \text{cm} . \tag{10.16}$$

Equations (10.14a, b) can be combined to give

$$\varepsilon_0 + p_0 = 2\beta / 3 l^2 a^2 . \tag{10.17}$$

Subscripts 0 denote background values. Using these equations, we can express c_s^2, the squared adiabatic velocity of sound, in the following equivalent forms:

$$c_s^2 = \left.\frac{\delta p}{\delta \varepsilon}\right|_s = \frac{p'}{\varepsilon'} = -\frac{(a\beta)'}{3a\mathcal{H}\beta} = -\frac{1}{3}\left(1 + \frac{\beta'}{\mathcal{H}\beta}\right) = -\frac{\mathrm{d}\ln(p_0 + \varepsilon_0)a^3}{\mathrm{d}\ln a^3} . \tag{10.18}$$

These relations for background variables will be extensively used in the further derivation.

For computational ease, the derivation of the expression for $\delta_2 S_m$ will be given in the case of a flat universe ($\mathcal{K} = 0$). The final results will be generalized to spatially open and closed universes ($\mathcal{K} = \pm 1$).

The first step in obtaining the action for hydrodynamical perturbations is to rewrite the initial action in terms of the dynamical degrees of freedom. The energy density ε is not one of these basic dynamical degrees of freedom, and hence it is not possible to immediately expand ε in the action (10.13). The basic dynamical variable characterizes the fluid flow. Thus, to introduce the dynamical degrees of freedom, we take test particles with space- and time-dependent number density ρ. The energy per particle consists of the rest mass m_0 and of the "potential energy" $\pi(\rho)$ which in turn depends on the pressure p. Following Fock [126],

$$m_0 + \pi(\rho) = m_0 + \int^{\rho} \frac{\mathrm{d}p}{\mathrm{d}\rho}\frac{\mathrm{d}\rho}{\rho} - \frac{p(\rho)}{\rho} . \tag{10.19}$$

Thus, the total energy density ε is

$$\varepsilon = \rho[m_0 + \pi(\rho)] . \tag{10.20}$$

The number density ρ satisfies the usual continuity equation,

$$(\rho u^\alpha)_{;\alpha} = 0\,, \tag{10.21}$$

where u^α is the four-velocity of the fluid, and a semicolon stands for the covariant derivative.

Next, we introduce Lagrange coordinates a^i which label the particles (the Lagrange coordinates of a given particle do not change in time) and an affine parameter λ which fixes the point along a particle trajectory. The Euclidean (comoving) coordinates x^α are then functions of a^i and λ

$$x^\alpha = f^\alpha(a^i, \lambda)\,. \tag{10.22}$$

We write the background flow as

$$x_0^\alpha = f_0^\alpha(a^i, \lambda)\,. \tag{10.23}$$

The perturbation of the fluid flow can then be described by a shift vector $\xi^\alpha = \xi^\alpha(x_0^\beta)$ which shifts the position of the test particle from x_0^β where it would be in an unperturbed universe. Then the full flow is

$$x^\alpha = f^\alpha(a^i, \lambda) = f_0^\alpha(a^i, \lambda) + \xi^\alpha(x^\beta)\,. \tag{10.24}$$

In terms of f^α, the four-velocity u^α is

$$u^\alpha = (\delta f^\alpha/\delta\lambda)[g_{\beta\gamma}(\delta f^\beta/\delta\lambda)\,\delta f^\gamma/\delta\lambda]^{-1/2}\,. \tag{10.25}$$

The number density ρ can be described by an arbitrary function $F(a^i)$ of the Lagrange coordinates and evolves in time according to the Jacobean J of the transformation between Euclidean and Lagrange coordinates,

$$J = \mathscr{D}(x^\alpha)/\mathscr{D}(a^i, \lambda)\,, \tag{10.26}$$

in the following way:

$$\rho(x^\alpha) = F(a^i)[g_{\alpha\beta}(\partial f^\alpha/\partial\lambda)\,\partial f^\beta/\partial\lambda]^{1/2}(\sqrt{-g}\,J)^{-1}\,. \tag{10.27}$$

It is not hard to verify that (10.25) and (10.27) solve the continuity equation (10.21).

Inserting (10.20) and (10.27) into the equation (10.13) for the matter action, we are finally in a position to work out the terms in the action of second order in the perturbation variables. There are contributions involving metric perturbations alone (from expanding $\sqrt{-g}$), terms with metric and matter perturbations, and expressions containing only matter variations. We obtain

$$\delta_2 S_{\mathrm{m}} = -\int \left[\varepsilon_0 \frac{\delta_2\sqrt{-g}}{\sqrt{-g_0}} + (\varepsilon_0 + p_0)\left(\frac{\delta_1\rho}{\rho_0}\,\frac{\delta_1\sqrt{-g}}{\sqrt{-g_0}} + \frac{\delta_2\rho}{\rho_0}\right) + \tfrac{1}{2}c_s^2(\varepsilon_0 + p_0)\,\frac{(\delta_1\rho)^2}{\rho_0^2}\right]\sqrt{-g_0}\,\mathrm{d}^4x\,. \tag{10.28}$$

The next task is to evaluate the individual terms. To do that, we first express the perturbations $\delta_1\rho(x^\alpha)$ and $\delta_2\rho(x^\alpha)$ of the number density at the space–time point x^α in terms of $\delta_1\rho(x^\alpha + \xi^\alpha)$, $\delta_2\rho(x^\alpha + \xi^\alpha)$ and ξ^α by means of (10.24)–(10.26). Applying the Taylor expansion and keeping only terms up to second order in ξ^α we obtain

$$(\rho_0 + \delta_1\rho + \delta_2\rho)(x^\alpha) = (\rho_0 + \delta_1\rho + \delta_2\rho)(x^\alpha + \xi^\alpha) - \frac{\partial \rho_0}{\partial x^\beta}\,\xi^\beta - \frac{\partial}{\partial x^\gamma}\left(\delta_1\rho(x^\alpha + \xi^\alpha) - \frac{\partial \rho_0}{\partial x^\beta}\,\xi^\beta\right)\xi^\gamma - \frac{1}{2}\,\frac{\partial^2 \rho_0}{\partial x^\beta\,\partial x^\gamma}\,\xi^\beta\xi^\gamma\,. \tag{10.29}$$

In our case, ρ_0 depends only on time, so only the time derivative and ξ^0 survive in all terms involving derivatives of ρ_0. The unperturbed number density of particles ρ_0 is proportional to a^{-3}. We can evaluate $\rho(x^\alpha + \xi^\alpha)$ on the right-hand side of (10.29) by using (10.27) and calculating separately the individual terms in this formula up to second order in the perturbations. Lengthy but straightforward calculations yield the following expressions for individual terms in (10.29):

$$(J_0 + \delta_1 J + \delta_2 J)(x^\alpha + \xi^\alpha) = 1 + \frac{\partial \xi^\alpha}{\partial x^\alpha} + \frac{1}{2}\left(\frac{\partial \xi^\alpha}{\partial x^\alpha}\right)^2 - \frac{1}{2}\,\frac{\partial \xi^\alpha}{\partial x^\beta}\,\frac{\partial \xi^\beta}{\partial x^\alpha}\,, \tag{10.30}$$

$$\sqrt{-g}(x^\alpha + \xi^\alpha) = \sqrt{-g}(x^\alpha)[1 + 4\mathcal{H}\xi^0 + 2(\mathcal{H}' + 4\mathcal{H}^2)(\xi^0)^2 + (\phi - 3\psi + E_{,ii})'\xi^0 + (\phi - 3\psi + E_{,ii})_{,j}\xi^j]\,. \tag{10.31}$$

Thus we get

$$\left(g_{\alpha\beta}(x^\gamma + \xi^\gamma)\,\frac{\partial f^\alpha}{\partial \lambda}\,\frac{\partial f^\beta}{\partial \lambda}\right)^{1/2} = \left({}^{(0)}g_{\alpha\beta}\,\frac{\partial f_0^\alpha}{\partial \lambda}\,\frac{\partial f_0^\beta}{\partial \lambda}\right)^{1/2}[1 + \phi + \mathcal{H}\xi^0 + \xi^{0\prime} + \mathcal{H}\phi\xi^0 + \phi\xi^{0\prime} - \tfrac{1}{2}\phi^2 + \phi'\xi^0 + \phi_{,i}\xi^i - B_{,i}\xi^{i\prime} + \tfrac{1}{2}(\mathcal{H}' + \mathcal{H}^2)(\xi^0)^2 - \tfrac{1}{2}\xi^{i\prime}\xi^{i\prime} + \mathcal{H}\xi^0\xi^{0\prime}]\,. \tag{10.32}$$

The value of $\sqrt{-g}(x^\alpha + \xi^\alpha)$ in (10.31) was expressed by expanding about the term $\sqrt{-g}(x^\alpha)$ which in turn is equal to

$$\sqrt{-g}(x^\alpha) = (\mathcal{N}\gamma^{1/2})(x^\alpha)\,, \tag{10.33}$$

where the individual terms were calculated in terms of our metric variables ϕ, ψ, B and E in the previous section [see (10.7) and (10.9)]. Substituting (10.30)–(10.33) into (10.27), and then into eq. (10.29) we find

$$(\rho_0 + \delta_1\rho + \delta_2\rho)(x^\alpha) = \rho_0[1 + 3\psi - E_{,ii} - \xi^i_{,i} - B_{,i}\xi^{i\prime} - \tfrac{1}{2}\xi^{i\prime}\xi^{i\prime} - \tfrac{1}{2}B_{,i}B_{,i} + \tfrac{15}{2}\psi^2 + \tfrac{1}{2}E_{,ii}E_{,jj} + E_{,ij}E_{,ji} - 5\psi E_{,ii} - (3\psi - E_{,ii})\xi^j_{,j} + (\xi^0\xi^{i\prime} + \tfrac{1}{2}\xi^i_{,j}\xi^j + \tfrac{1}{2}\xi^j_{,j}\xi^i)_{,i}]\,. \tag{10.34}$$

Finally, reading off the individual first- and second-order terms in (10.28) from (10.33) and (10.34), we obtain the following form for the perturbed matter action:

$$\delta_2 S_m = \int \Big([\tfrac{1}{2}\varepsilon_0 \phi^2 + p_0(\tfrac{3}{2}\psi^2 - 3\phi\psi + \phi E_{,ii} - \psi E_{,ii} + \tfrac{1}{2}E_{,ii}E_{,jj} - E_{,ij}E_{,ji} + \tfrac{1}{2}B_{,i}B_{,i})$$

$$+ (\varepsilon_0 + p_0)(\tfrac{1}{2}\xi^{i\prime}\xi^{i\prime} + B_{,i}\xi^{i\prime} + \phi \xi^i_{,i}) - \tfrac{1}{2}c_s^2(\varepsilon_0 + p_0)(3\psi - E_{,ii} - \xi^i_{,i})^2]a^4 + \frac{1}{6l^2}\mathscr{D}_1^{\mathrm{m}} \Big)\, \mathrm{d}^4x\,, \tag{10.35}$$

where $\mathscr{D}_1^{\mathrm{m}}$ is the total derivative term

$$\mathscr{D}_1^{\mathrm{m}} = -6l^2[(\varepsilon_0 + p_0)a^4(\xi^0\xi^{i\prime} + \tfrac{1}{2}\xi^i_{,j}\xi^j + \tfrac{1}{2}\xi^j_{,j}\xi^i)]_{,i}\,. \tag{10.36}$$

Now, it is possible to combine the gravitational part of the action (10.11) with the matter part (10.35) to obtain the total action to second order in the perturbations,

$$\delta_2 S = \delta_2 S_{\mathrm{gr}} + \delta_2 S_{\mathrm{m}} = \frac{1}{6l^2}\int a^2\Big\{-6\Big[\psi'^2 + 2\mathscr{H}\phi\psi' + \Big(\mathscr{H}^2 - \frac{\beta}{3c_s^2}\Big)\phi^2\Big]$$

$$- 4(\psi' + \mathscr{H}\phi)(B - E')_{,ii} - 2\psi_{,i}(2\phi_{,i} - \psi_{,i})$$

$$+ 2\beta(\xi^{i\prime} + B_{,i})(\xi^{i\prime} + B_{,i}) - 2\beta c_s^2\Big(3\psi - E_{,ii} - \xi^i_{,i} + \frac{1}{c_s^2}\phi\Big)^2\Big\}\,\mathrm{d}^4x$$

$$+ \frac{1}{6l^2}\int (\mathscr{D}_1^{\mathrm{gr}} + \mathscr{D}_2^{\mathrm{gr}} + \mathscr{D}_1^{\mathrm{m}} + \mathscr{D}_2^{\mathrm{m}})\,\mathrm{d}^4x\,. \tag{10.37}$$

To obtain this form, we have used the definition of β from (10.15) and the background equations of motion (10.14a, b). $\mathscr{D}_2^{\mathrm{m}}$ denotes a further total derivative term containing contributions from both $\delta_2 S_{\mathrm{m}}$ and $\delta_2 S_{\mathrm{gr}}$,

$$\mathscr{D}_2^{\mathrm{m}} = (4\mathscr{H}a^2\psi E_{,ii} + 2\mathscr{H}a^2E_{,ii}^2 - 6\mathscr{H}a^2\psi^2)' + [2a^2(2\mathscr{H}' + \mathscr{H}^2)(E_{,ij}E_{,j} - E_{,jj}E_{,i}) - 4\mathscr{H}a^2\psi B_{,i}]_{,i}\,. \tag{10.38}$$

In the above form, the action looks rather complicated. It is possible to achieve a substantial simplification: by using the constraint equations, we can rewrite the action in terms of a minimal set of gauge-invariant variables which describes simultaneously the gravitational and matter inhomogeneities. This form is then convenient for quantization.

The first constraint equation follows immediately by varying the action (10.37) with respect to $B_{,i}$,

$$(\psi' + \mathscr{H}\phi)_{,i} = -\beta(\xi^{i\prime} + B_{,i}) = -\tfrac{3}{2}l^2a^2(\varepsilon_0 + p_0)u_0(\delta u^i + a^{-1}B_{,i})\,, \tag{10.39}$$

where

$$\delta u^i = a^{-1}\xi^{i\prime} \tag{10.40}$$

is the three-velocity of the fluid in linear approximation. Equation (10.39) is the (0–i) Einstein equation. From it, we conclude that the velocity potential φ_v can be defined as

$$\varphi_{v,i} = (2\beta^{1/2}a/c_s)(\xi^{i\prime} + B_{,i}) = -(2\beta^{1/2}a^2/c_s)\,\delta u_i\,. \tag{10.41}$$

This potential is not gauge-invariant. However, it is easy to construct the corresponding gauge-invariant quantity

$$\varphi_v^{(\mathrm{gi})} \equiv \varphi_v - 2(a\beta^{1/2}/c_s)(B - E')\,. \tag{10.42}$$

Using $\varphi_v^{(\mathrm{gi})}$ and Ψ we can construct the following gauge-invariant variable:

$$v \equiv \frac{1}{\sqrt{6}l}\,(\varphi_v^{(\mathrm{gi})} - 2z\Psi) = \frac{1}{\sqrt{6}l}\,(\varphi_v - 2z\psi)\,, \quad z = \frac{a\beta^{1/2}}{\mathscr{H}c_s}\,. \tag{10.43a, b}$$

The goal now is to prove that the action (10.37) can be rewritten in terms of v alone by means of the constraint equations.

Before doing that we demonstrate how to obtain the other Einstein equations from the action principle. Varying the action (10.37) with respect to ϕ, ψ and $E_{,ii}$ one gets

$$\Delta\psi - 3\mathscr{H}\psi' - 3\mathscr{H}^2\phi - \mathscr{H}(B - E')_{,ii} = \beta(3\psi - E_{,ii} - \xi^i_{,i}) = \tfrac{3}{2}l^2a^2\,\delta_1\varepsilon\,, \tag{10.44}$$

$$\psi'' + 2\mathscr{H}\psi' + \mathscr{H}\phi' + (2\mathscr{H}' + \mathscr{H}^2)\phi + \tfrac{1}{3}\{\Delta(\phi - \psi) + a^{-2}\Delta[a^2(B - E')]'\}$$
$$= c_s^2\beta(3\psi - E_{,ii} - \xi^i_{,i}) = \tfrac{3}{2}l^2a^2\,\delta_1 p\,, \tag{10.45}$$

$$\psi'' + 2\mathscr{H}\psi' + \mathscr{H}\phi' + (2\mathscr{H}' + \mathscr{H}^2)\phi = \tfrac{3}{2}l^2a^2\,\delta_1 p\,. \tag{10.46}$$

To simplify the right-hand side of (10.44) and (10.45) we have taken into account that

$$\delta_1 u^0 = -a^{-1}\phi\,, \qquad \delta_1\varepsilon = (\varepsilon_0 + p_0)(3\psi - E_{,ii} - \xi^i_{,i})\,, \tag{10.47}$$

$$\delta_1 p = c_s^2\,\delta_1\varepsilon\,. \tag{10.48}$$

Comparing (10.45) and (10.46), we conclude that

$$\Delta(\phi - \psi + a^{-2}[a^2(B - E')]') = 0 \tag{10.49}$$

[see eq. (3.13) and discussion thereof], which from (3.13) implies that the two gauge-invariant metric potentials Φ and Ψ are equal,

$$\Phi = \Psi\,. \tag{10.50}$$

Thus, we have rederived the basic results starting from the action.

Finally, by varying the action with respect to ξ^i we obtain the matter equation of motion

$$[\beta a^2(\xi^{i\prime}+B_{,i})]'+\beta a^2 c_s^2[3\psi - E_{,jj}-\xi^j_{,j}+c_s^{-2}\phi]_{,i}=0\,. \tag{10.51}$$

This equation is redundant. It in fact follows from (10.43a) and (10.48).

We now use eqs. (10.41) and (10.51) to exclude ξ^i and its derivatives from the action (10.37) expressing these quantities in terms of ϕ, ψ, B, E and φ. First, eq. (10.41) can be written in the form

$$\xi^{i\prime}+B_{,i}=\tfrac{1}{2}(c_s/\beta^{1/2}a)\varphi_{v,i}\,. \tag{10.52}$$

From the constraint equation (10.39) we conclude that

$$\psi' + \mathcal{H}\phi = -\tfrac{1}{2}(c_s\beta^{1/2}/a)\varphi_v\,, \tag{10.53}$$

thus, taking into account (10.52), we can rewrite (10.51) as

$$3\psi - E_{,ii}+c_s^{-2}\phi - \xi^i_{,i} = -\tfrac{1}{2}(c_s\beta^{1/2}a\varphi_v)'/c_s^2\beta a^2\,. \tag{10.54}$$

Using (10.52) and (10.54), the last two terms in the action (10.37) can be converted in the following manner [here we do not use the constraint equation (10.53)]:

$$\begin{aligned}
2\beta a^2(\xi^{i\prime}+B_{,i})(\xi^{i\prime}+B_{,i}) &= c_s\beta^{1/2}a\varphi_{v,i}(\xi^{i\prime}+B_{,i})\\
&= [c_s\beta^{1/2}a\varphi_v(\xi^{i\prime}+B_{,i})]_{,i} - c_s\beta^{1/2}a\varphi_v(\xi^{i\prime}_{,i}+B_{,ii})\\
&= [c_s\beta^{1/2}a\varphi_v(\xi^{i\prime}+B_{,i})]_{,i} - c_s\beta^{1/2}a\varphi_v\\
&\quad\times\left[\frac{1}{2}\left(\frac{(c_s\beta^{1/2}a\varphi_v)'}{c_s^2\beta a^2}\right)' + 3\psi' - E'_{,ii} + \left(\frac{1}{c_s^2}\,\phi\right)' + B_{,ii}\right]\\
&= \frac{1}{2}\,\frac{(c_s\beta^{1/2}a\varphi_v)'^2}{c_s^2\beta a^2} - 3c_s\beta^{1/2}a\varphi_v\psi' - c_s\beta^{1/2}a\varphi_v\left(\frac{1}{c_s^2}\,\phi\right)'\\
&\quad - c_s\beta^{1/2}a\varphi_v(B-E')_{,ii} - \frac{1}{4}\left(\frac{(c_s^2\beta a^2\varphi_v^2)'}{c_s^2\beta a^2}\right)' + \tfrac{1}{4}(c_s^2\varphi_v^2)_{,ii}\,,
\end{aligned} \tag{10.55}$$

$$\begin{aligned}
-2\beta a^2c_s^2(3\psi - E_{,ii}-\xi^i_{,i}+c_s^{-2}\phi)^2 &= (c_s\beta^{1/2}a\varphi_v)'(3\psi - E_{,ii}-\xi^i_{,i}+c_s^{-2}\phi)\\
&= -c_s\beta^{1/2}a\varphi_v[3\psi' - E'_{,ii}-\xi^{i\prime}_{,i}+(c_s^{-2}\phi)']\\
&\quad + [c_s\beta^{1/2}a\varphi_v(3\psi - E_{,ii}-\xi^i_{,i}+c_s^{-2}\phi)]'\\
&= -c_s\beta^{1/2}a\varphi_v[3\psi' + (c_s^{-2}\phi)' + (B-E')_{,ii}]\\
&\quad - \tfrac{1}{2}c_s^2\varphi_{v,i}\varphi_{v,i} + \tfrac{1}{4}(c_s^2\varphi_v^2)_{,ii} - \tfrac{1}{4}[(c_s^2\beta a^2\varphi_v^2)'/c_s^2\beta a^2]'\,.
\end{aligned} \tag{10.56}$$

Using (10.55) and (10.56), the action (10.37) takes on the following form:

$$\delta_2 S = \frac{1}{6l^2}\int\Big[-6a^2(\psi' + \mathcal{H}\phi)^2 + \frac{2\beta a^2}{c_s^2}\,\phi^2 - 2a^2\psi_{,i}(2\phi_{,i} - \psi_{,i}) + \frac{1}{2}\,\frac{(c_s\beta^{1/2}a\varphi_v)'^2}{c_s^2\beta a^2}$$
$$- \tfrac{1}{2}c_s^2\varphi_{v,i}\varphi_{v,i} - 6c_s\beta^{1/2}a\varphi_v\psi' - 2c_s\beta^{1/2}a\varphi_v(c_s^{-2}\phi)'$$
$$- 4a^2\Big(\psi' + \mathcal{H}\phi + \frac{c_s\beta^{1/2}}{2a}\,\varphi_v\Big)(B - E')_{,ii}\Big]\,d^4x$$
$$+ \frac{1}{6l^2}\int(\mathcal{D}_1^{gr} + \mathcal{D}_2^{gr} + \mathcal{D}_1^{m} + \mathcal{D}_2^{m} + \mathcal{D}_3^{m})\,d^4x\,, \qquad (10.57)$$

where $\mathcal{D}_3^{m}$ is a further total derivative term,

$$\mathcal{D}_3^{m} = 3l^2\Big[(c_s^2\varphi_v^2)_{,ii} - \Big(\frac{(c_s^2\beta a^2\varphi_v^2)'}{c_s^2\beta a^2}\Big)\Big]'\,. \qquad (10.58)$$

Finally, we express φ_v and ψ' in eq. (10.57) in terms of the gauge-invariant variable v and ψ and ϕ using the definition of v in (10.43a) and the constraint equation (10.53). Taking into account the background equations, the constraint equation follows by varying the action with respect to $B - E'$. The result takes on a very simple form,

$$\delta_2 S = \frac{1}{2}\int\Big(v'^2 - c_s^2 v_{,i}v_{,i} + \frac{z''}{z}\,v^2\Big)\,d^4x + \frac{1}{6l^2}\int\Big(\mathcal{D}_1^{gr} + \mathcal{D}_2^{gr} + \sum_{i=1}^{4}\mathcal{D}_i^{m}\Big)\,d^4x\,, \qquad (10.59)$$

where $\mathcal{D}_4^{m}$ is a total derivative term

$$\mathcal{D}_4^{m} = \Big[\frac{1}{2}\,\frac{(\mathcal{H}c_s^2 z)'}{\mathcal{H}c_s^2 z}\,\varphi_v^2 + 2\,\frac{a^2}{\mathcal{H}}\,\psi_{,i}\psi_{,i} + (z^2)'\psi^2 - 2\mathcal{H}z\phi\varphi_v$$
$$- 2\,\frac{\beta}{\mathcal{H}}\,z^2\psi^2 + 2\sqrt{6}\,l\Big(z' - \frac{\mathcal{H}c_s^2 z^3}{a^2}\Big)v\psi - 2l^2\,\frac{\mathcal{H}c_s^2 z^2}{a^2}\,v^2\Big]'\,. \qquad (10.60)$$

The above action is like the action of the scalar field v with a time-dependent mass. Its quantization will be discussed in chapter 11.

The action (10.59) was derived for a flat universe ($\mathcal{K} = 0$). To generalize this result to closed and open universes ($\mathcal{K} = \pm 1$) we define the gauge-invariant variable v by

$$v = (1/\sqrt{6}\,l)[\varphi_v^{(gi)} - 2z\psi + (2a\mathcal{K}/\mathcal{H}\beta^{1/2}c_s)\psi]\,, \qquad (10.61)$$

and – omitting the total derivatives – obtain essentially the same form for the action,

$$\delta_2 S = \frac{1}{2}\int\Big(v'^2 - c_s^2\gamma^{ij}v_{,i}v_{,j} + \frac{z''}{z}\,v^2\Big)\sqrt{\gamma}\,d^3x\,d\eta\,, \qquad (10.62)$$

where z and β were defined in (10.43b) and (10.15).

10.3. The action of scalar-field matter

The derivation of the action for scalar-field matter is less involved than that for hydrodynamical matter since the action can be immediately expanded to second order in the perturbations [20]. Here and in the following section we shall only concentrate on the case of $\mathcal{K}=0$. We consider a theory with total action given by

$$S=-\frac{1}{16\pi G}\int R(-g)^{1/2}\,\mathrm{d}^4x+\int[\tfrac{1}{2}\varphi_{,\alpha}\varphi^{,\alpha}-V(\varphi)](-g)^{1/2}\,\mathrm{d}^4x\,. \tag{10.63}$$

It is useful to recall the two basic background equations (see chapter 6)

$$\mathcal{H}^2=l^2[\tfrac{1}{2}\varphi_0'^2+V(\varphi_0)a^2]\,,\qquad 2\mathcal{H}'+\mathcal{H}^2=3l^2[-\tfrac{1}{2}\varphi_0'^2+V(\varphi_0)a^2]\,. \tag{10.64a, b}$$

These two equations can be combined to give

$$\mathcal{H}^2-\mathcal{H}'=\tfrac{3}{2}l^2\varphi_0'^2\,. \tag{10.65}$$

From section 10.1, the form of the contribution to the first term in (10.63) of second order in the fluctuation variables is known. When evaluating $\delta_2 S_\mathrm{m}$, where S_m stands for the matter part of the action (10.63), several terms must be considered,

$$\delta_2 S_\mathrm{m}=\int \mathrm{d}^4x\,(-g_0)^{1/2}\left(\frac{\delta_2(-g)^{1/2}}{(-g_0)^{1/2}}\,\mathcal{L}_0+\frac{2\,\delta_1(-g)^{1/2}\,\delta_1\mathcal{L}}{(-g_0)^{1/2}}+\delta_2\mathcal{L}\right), \tag{10.66}$$

where subscripts 0 denote the homogeneous background values and

$$\mathcal{L}(\varphi)=\tfrac{1}{2}\varphi_{,\alpha}\varphi^{,\alpha}-V(\varphi)\,. \tag{10.67}$$

$\delta_1\mathcal{L}$ and $\delta_2\mathcal{L}$ can be immediately read off by expanding (10.67) in a Taylor series about φ_0. Applying the background equations (10.64a, b) and (10.65), integrating by parts and combining $\delta_2 S_\mathrm{m}$ with $\delta_2 S_\mathrm{gr}$ from (10.11), we obtain

$$\begin{aligned}\delta_2 S=\delta_2 S_\mathrm{gr}+\delta_2 S_\mathrm{m}=\frac{1}{6l^2}\int\{&a^2[-6\psi'^2-12\mathcal{H}\phi\psi'-2\psi_{,i}(2\phi_{,i}-\psi_{,i})-2(\mathcal{H}'+2\mathcal{H}^2)\phi^2\\&+3l^2(\delta\varphi'^2-\delta\varphi_{,i}\,\delta\varphi_{,i}-V_{,\varphi\varphi}a^2\,\delta\varphi^2)+6l^2[\varphi_0'(\phi+3\psi)'\,\delta\varphi-2V_{,\varphi}a^2\phi\,\delta\varphi]\\&+4(B-E')_{,ii}(\tfrac{3}{2}l^2\varphi_0'\,\delta\varphi-\psi'-\mathcal{H}\phi)]+\mathcal{D}_1^\mathrm{gr}+\mathcal{D}_2^\mathrm{gr}+\mathcal{D}_3\}\,\mathrm{d}^4x,\end{aligned} \tag{10.68}$$

where $\mathcal{D}_3$ is a total derivative term

$$\begin{aligned}\mathcal{D}_3=&[6l^2a^2\varphi_0'\,\delta\varphi'\,(E_{,ii}-\phi-3\psi)+2\mathcal{H}a^2(E_{,ii}^2+2\psi E_{,ii}-3\psi^2)]'\\&+\{a^2[2(2\mathcal{H}'+\mathcal{H}^2)(E_{,ji}E_{,j}-E_{,jj}E_{,i})-4\mathcal{H}\psi B_{,i}-6l^2\varphi_0'B_{,i}\,\delta\varphi]\}_{,i}\,.\end{aligned} \tag{10.69}$$

By varying (10.68) with respect to $B - E'$, we get the following constraint equation:

$$\psi' + \mathcal{H}\phi = \tfrac{3}{2} l^2 \varphi_0' \, \delta\varphi \, . \tag{10.70}$$

In analogy to section 10.2, we introduce a gauge-invariant potential

$$v = a[\delta\varphi + (\varphi_0'/\mathcal{H})\psi] = a[\delta\varphi^{(\mathrm{gi})} + (\varphi_0'/\mathcal{H})\Psi] \, , \tag{10.71}$$

$$\delta\varphi^{(\mathrm{gi})} = \delta\varphi + \varphi_0'(B - E') \, , \tag{10.72}$$

where $\delta\varphi^{(\mathrm{gi})}$ is the gauge-invariant scalar-field variation. Using (10.70) and (10.71) to express ψ' and $\delta\varphi$ in terms of ϕ, ψ and v, and after some straightforward but rather lengthy calculations, we obtain the following simple action:

$$\delta_2 S = \frac{1}{2} \int \left(v'^2 - v_{,i} v_{,i} + \frac{z''}{z} v^2 + \frac{1}{3l^2} \sum_{i=1}^{4} \mathcal{D}_i \right) \mathrm{d}^4 x \, , \tag{10.73}$$

$$z = a\varphi_0'/\mathcal{H} \, , \tag{10.74}$$

$$\mathcal{D}_4 = -3l^2 \Big[2 \frac{a^2}{\mathcal{H}} \left(\frac{\varphi_0'}{a} \right)' v\psi + \frac{3l^2 \psi_0'^2}{2\mathcal{H}} v^2 - 2a^2 \varphi_0' v \phi - \frac{2a^2}{3l^2 \mathcal{H}} \psi_{,i} \psi_{,i}$$
$$+ 2 \frac{\varphi_0'^2}{\mathcal{H}} a^2 \phi\psi + \frac{1}{2} \left(\frac{\varphi_0'}{a} \right)'^2 \frac{a^4}{\mathcal{H}} \psi^2 + \mathcal{H} v^2 \Big]' \, , \tag{10.75}$$

a further total divergence. Once again, the final action is that of a scalar field in flat space–time with time-dependent mass.

10.4. *The action for higher-derivative theories of gravity*

Finally, we consider a gravity theory with metric $g_{\mu\nu}$ and action

$$S = -\frac{1}{6l^2} \int f(R) \sqrt{-g} \, \mathrm{d}^4 x \, . \tag{10.76}$$

Our goal is to find $\delta_2 S$, the perturbation of S about a homogeneous background to second order in the fluctuation variables [120]. Using the conformal transformation technique discussed in chapter 7, it is easy to reduce the problem to the case discussed in the previous section.

We recall from (7.1)–(7.10) that the equations of motion for $g_{\mu\nu}$ which follow by varying S are equivalent to the Einstein equations, and the corresponding action to the Einstein action for a rescaled metric,

$$\tilde{g}_{\mu\nu} = F(R) g_{\mu\nu} \, , \tag{10.77}$$

and a scalar field $\varphi(R)$ with potential $V(\varphi)$, provided that $F(R)$ and $V(\varphi)$ are chosen appropriately, namely

$$F(R) = \partial f/\partial R\,, \qquad V(\varphi) = (1/6l^2)[f(R) - R\,\partial f/\partial R](\partial f/\partial R)^{-2}\,. \tag{10.78a, b}$$

The scalar field itself is given by

$$\varphi(R) = (1/\sqrt{2}\,l)\ln F(R)\,. \tag{10.79}$$

At this point we can immediately apply the results of section 10.3 and express $\delta_2 S$ as

$$\delta_2 S = \frac{1}{2}\int\left(\tilde{v}'^2 - \tilde{v}_{,i}\tilde{v}_{,i} + \frac{z''}{z}\,\tilde{v}^2 + \mathscr{D}\right)\mathrm{d}^4x\,, \tag{10.80}$$

where $\mathscr{D}$ is a total derivative term. The variable $\tilde{v}$ is a conformal gauge-invariant potential defined by [see (10.71)]

$$\tilde{v} = \tilde{a}[\delta\varphi + (\tilde{a}\varphi_0'/\tilde{a}')\tilde{\psi}]\,, \qquad \tilde{z} = \tilde{a}\varphi_0'/\tilde{\mathscr{H}}\,, \tag{10.81}$$

where the variables with tilde denote those which appear in the conformal metric $\tilde{g}_{\mu\nu}$ [see (7.34)],

$$\mathrm{d}\tilde{s}^2 = F\,\mathrm{d}s^2 = \tilde{a}^2(\eta)\{(1+2\tilde{\phi})\,\mathrm{d}\eta^2 - 2\tilde{B}_{,i}\,\mathrm{d}x^i - [(1-2\tilde{\psi})\delta_{ij} + 2\tilde{E}_{,ij}]\,\mathrm{d}x^i\,\mathrm{d}x^j\}\,. \tag{10.82}$$

The relation of the variables with tilde and those without which appear in $\mathrm{d}s^2$ was worked out in eqs. (7.34)–(7.36). In particular,

$$\tilde{a} = F_0^{1/2}a\,, \qquad \tilde{\psi} = \psi - (\partial\ln F_0^{1/2}/\partial R)\,\delta R\,. \tag{10.83a, b}$$

Using (10.79) and (10.83a, b), $\tilde{v}$ can be written in terms of variables related to the original metric $g_{\mu\nu}$,

$$\tilde{v} = \frac{\sqrt{2}}{l}\,aF_0^{1/2}\left(\frac{2F_0}{\partial F_0/\partial R}\,\mathscr{H} + R_0'\right)^{-1}(R_0'\psi + \mathscr{H}\,\delta R)\,, \tag{10.84}$$

or, in terms of gauge-invariant variables Ψ and $\delta R^{(\mathrm{gi})}$,

$$\tilde{v} = \frac{\sqrt{2}}{l}\,a\mathrm{F}_0^{1/2}\left(\frac{2F_0}{\partial F_0/\partial R}\,\mathscr{H} + R_0'\right)^{-1}(R_0'\Psi + \mathscr{H}\,\delta R^{(\mathrm{gi})})\,, \tag{10.85}$$

where we recall that

$$\delta R^{(\mathrm{gi})} = \delta R + R_0'(B - E')\,. \tag{10.86}$$

This completes our derivation of the action for perturbations. In all three cases of interest, the problem has been reduced to that of quantizing a scalar field with time-dependent mass.

11. Quantization

In all three cases under consideration we have been able to reduce the action for cosmological perturbations about a classical FRW background solution to the following form:

$$\delta_2 S = \int \mathscr{L}\sqrt{\gamma}\, d^4x = \frac{1}{2}\int \left(v'^2 - c_s^2\gamma^{ij}v_{,i}v_{,j} + \frac{z''}{z}v^2\right)\sqrt{\gamma}\, d^4x\,, \tag{11.1}$$

up to total derivatives which have been omitted. In the above, γ^{ik} is the metric on the background η = constant hypersurfaces, γ is its determinant, v is the gauge-invariant potential to be quantized, and z is a time-dependent function. Except for hydrodynamical perturbations, $c_s^2 = 1$.

The action (11.1) effectively describes a scalar field with time-dependent mass,

$$m^2 = -z''/z\,. \tag{11.2}$$

In the case of a time-dependent c_s^2 (for example during the transition between radiation and matter-dominated periods of the evolution of the universe), the spatial gradient term in (11.1) also depends on time. Here, for the sake of simplicity we will only consider the case of time-independent c_s^2. If there is a slow time dependence, it can be treated adiabatically to a first approximation. For a flat universe, $\gamma^{ik} = \delta^{ik}$ and $\sqrt{\gamma} = 1$.

The quantization of the classical action (11.1) is analogous to the quantization of Minkowski space–time scalar fields in external fields [127] and also has similarities to the analysis of scalar fields in an expanding universe [128]. The time dependence in our case is entirely due to the variable background gravitational field. Thus, we can formulate the quantization prescription in analogy with well-studied examples.

The first step in canonically quantizing (11.1) is to determine the momentum π canonically conjugate to v,

$$\pi(\eta, \boldsymbol{x}) = \partial\mathscr{L}/\partial v' = v'(\eta, \boldsymbol{x})\,. \tag{11.3}$$

Then, the Hamiltonian is

$$H = \int (v'\pi - \mathscr{L})\sqrt{\gamma}\, d^3x = \frac{1}{2}\int \left(\pi^2 + c_s^2\gamma^{ij}v_{,i}v_{,j} - \frac{z''}{z}v^2\right)\sqrt{\gamma}\, d^3x\,. \tag{11.4}$$

In quantum theory, the variables v and π become operators $\hat{v}$ and $\hat{\pi}$ satisfying the standard commutation relations on the η = constant hypersurface,

$$[\hat{v}(\eta, \boldsymbol{x}), \hat{v}(\eta, \boldsymbol{x}')] = [\hat{\pi}(\eta, \boldsymbol{x}), \hat{\pi}(\eta, \boldsymbol{x}')] = 0\,, \qquad [\hat{v}(\eta, \boldsymbol{x}), \hat{\pi}(\eta, \boldsymbol{x}')] = i\delta(\boldsymbol{x} - \boldsymbol{x}')\,, \tag{11.5}$$

where the delta function is normalized by

$$\int \delta(\boldsymbol{x} - \boldsymbol{x}')\sqrt{\gamma}\, d^3x = 1\,. \tag{11.6}$$

Varying (11.1) with respect to v gives the equation of motion for v, which is also the field equation for $\hat{v}$,

$$\hat{v}'' - c_s^2\Delta\hat{v} - (z''/z)\hat{v} = 0\,. \tag{11.7}$$

This equation is equivalent to the Heisenberg equations

$$\mathrm{i}\hat{v}' = [\hat{v}, \hat{H}] , \qquad \mathrm{i}\hat{\pi}' = [\hat{\pi}, \hat{H}] , \tag{11.8}$$

where the operator Hamiltonian $\hat{H}$ is simply the Hamiltonian H written in terms of operators $\hat{v}$ and $\hat{\pi}$.

We shall work in the Heisenberg representation in which the state vectors are time-independent and all the time dependence is carried by the operators. The operator $\hat{v}$ can be expanded over a complete orthonormal basis of the solution of the (classical) field equation (11.7). These solutions will be denoted $\psi_J(x)v_J^*(\eta)$ and can be obtained using the method of separation of variables. $\psi_J(x)$ are eigenfunctions of the Laplace–Beltrami operator Δ with eigenvalue k_J^2

$$(\Delta + k_J^2)\psi_J(x) = 0 . \tag{11.9}$$

In particular, for a spatially flat universe ($\mathcal{K} = 0$) we can take a basis of plane waves,

$$\psi_J(x) = (2\pi)^{-3/2} \exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{k}) . \tag{11.10}$$

Substituting the ansatz for the solutions back into eq. (11.7) yields the following equation for $v_J(\eta)$:

$$v_J''(\eta) + E_J^2 v_J(\eta) = 0 , \quad E_J^2 = c_s^2 k_J^2 - z''/z . \tag{11.11a, b}$$

In terms of the modes discussed above, the expansion of the operator $\hat{v}$ is

$$\hat{v} = (1/\sqrt{2}) \int \mathrm{d}J \, [\psi_J(x)v_J^*(\eta)a_J^- + \psi_J^*(x)v_J(\eta)a_J^+] , \tag{11.12}$$

where the symbol $\int \mathrm{d}J$ denotes summation or integration over the modes. If the eigenvalues k_J are discrete, then $\int \mathrm{d}J$ is summation, if k_J is continuous, then $\int \mathrm{d}J$ denotes actual integration. For a flat universe ($\mathcal{K} = 0$), $\mathrm{d}J = \mathrm{d}^3k$. The creation and annihilation operators $\hat{a}_J^+$ and $\hat{a}_J^-$ in (11.12) satisfy the standard commutation relations for bosons

$$[\hat{a}_J^-, \hat{a}_{J'}^-] = [\hat{a}_J^+, \hat{a}_{J'}^+] = 0 , \qquad [\hat{a}_J^-, \hat{a}_{J'}^+] = \delta_{JJ'} , \tag{11.13}$$

where $\delta_{JJ'}$ is the usual Dirac function for continuous J and the Kronecker δ symbol for discrete J. It is straightforward to verify that the commutation relations (11.5) and (11.13) are only consistent if the following normalization conditions for $v_J(\eta)$ are satisfied:

$$v_J'(\eta)v_J^*(\eta) - v_J^{*\prime}(\eta)v_J(\eta) = 2\mathrm{i} . \tag{11.14}$$

The second step in the canonical quantization is the construction of the Fock representation of the Hilbert space of states on which the operators $\hat{v}$ and $\hat{\pi}$ act. Recall that for a free scalar field in flat space–time with constant mass m there is a unique vacuum state $|0\rangle$ defined by

$$\hat{a}_{\boldsymbol{k}}^- |0\rangle = 0 \; \forall \boldsymbol{k} , \tag{11.15}$$

where in the mode expansion (11.12), the annihilation operators $\hat{a}_{\boldsymbol{k}}^-$ are the operator coefficients of the positive-frequency modes

$$v_k(\eta) \sim \exp(\mathrm{i} w_k \eta)\,, \qquad w_k^2 = k^2 + m^2\,. \tag{11.16}$$

All other states in the Fock basis can be obtained by acting on $|0\rangle$ with a product of creation operators.

What renders the above prescription unique is that there is a distinguished time direction, and that the notion of positive and negative frequency is time-invariant. When quantizing the homogeneous component of a scalar field in an expanding FRW universe, there is no definite notion of time [128]. In addition, given some choice of time, solutions which are positive-frequency at time t_1 with respect to this time are no longer pure positive-frequency at a later time t_2. When quantizing perturbations about a curved background space–time, the background provides a distinguished time direction. However, the notion of a positive-frequency mode is still not time-invariant. Hence, there will be no unique definition of the vacuum state.

If we pick a time η_0, it is possible to find a linear combination of the two fundamental solutions of the time-dependent mode equation (11.11a) which is positive-frequency (for oscillating solutions) at η_0. This can be done by demanding

$$v_J(\eta_0) = E_J^{-1/2}(\eta_0)\,, \qquad v_J'(\eta_0) = \mathrm{i} E_J^{1/2}(\eta_0) \tag{11.17}$$

if E_J is positive for all modes J. The initial conditions (11.17) in most cases give consistent initial conditions to solve (11.11a). A vacuum state $|0_{\eta_0}\rangle \equiv |\psi_0\rangle$ can be defined in analogy to (11.5) by

$$a_J^- |0_{\eta_0}\rangle = 0 \,\forall J\,, \tag{11.18}$$

where the modes in (11.12) have been determined using the initial conditions (11.17). A more mathematical way to describe the above procedure for determining the vacuum $|\psi_0\rangle$ is to demand that the Hamiltonian (11.4) be diagonal in the Fock basis at time $\eta = \eta_0$ [128].

The time dependence of the notion of positive frequency has immediate and important consequences. An observer at time η_0 will define as the state empty of particles the state $|\psi_0\rangle$ described above. However, at a later time $\eta_1 > \eta_0$, the modes which are positive-frequency at η_0 according to (11.17) will no longer be positive-frequency at η_1. However, an observer at time η_1 will still define as the state empty of particles a state $|\psi_1\rangle$ which satisfies

$$\hat{b}_J^- |\psi_1\rangle = 0 \,\forall J\,, \tag{11.19}$$

where the b_J^- are the operator expansion coefficients of the modes of (11.7) which are positive-frequency at η_1. If we denote the positive- (negative-) frequency modes at η_i by $v_J^{(i)+}$ ($v_J^{(i)-}$), where $i = 0, 1$, then

$$v_J^{(1)+} = \alpha_J v_J^{(0)+} + \beta_J v_J^{(0)-}\,, \qquad v_J^{(1)-} = \beta_J^* v_J^{(0)+} + \alpha_J^* v_J^{(0)-}\,, \quad |\alpha|^2 - |\beta|^2 = 1\,. \tag{11.20}$$

Since the mode equation (11.7) is linear, it is clear that the modes defined to be positive- and negative-frequency at different times are related by a linear transformation (11.20) which is called a Bogoliubov transformation [128]. The Bogoliubov coefficients α_J and β_J are obtained by taking the inner products

$$\alpha_J = \langle v_J^{(0)+}, v_J^{(1)+} \rangle\,, \qquad \beta_J = \langle v_J^{(0)-}, v_J^{(1)+} \rangle\,. \tag{11.21}$$

From (11.20) we immediately obtain the transformation of creation and annihilation operators,

$$\hat{a}_J^- = \alpha_J \hat{b}_J^- + \beta_J^* \hat{b}_J^+ , \qquad \hat{a}_J^+ = \beta_J \hat{b}_J^- + \alpha_J^* \hat{b}_J^+ . \tag{11.22}$$

The observer at η_1 will define the following number operator for the number of particles in the Jth mode: $\hat{N}_J^1 = \hat{b}_J^+ \hat{b}_J^-$. The transformation (11.22) can immediately be inverted to obtain the operators $\hat{b}_J$ as functions of $\hat{a}_J^-$ and $\hat{a}_J^+$. Then, evaluating the number operator $\hat{N}_J^1$ in the state $|\psi_0\rangle$ we obtain

$$\langle \psi_0 | \hat{N}_J^1 | \psi_0 \rangle = |\beta_J|^2 . \tag{11.23}$$

Thus, if the Bogoliubov transformation (11.20) is nontrivial, the observer at η_1 will see the state $|\psi_0\rangle$ as containing a nonvanishing number of particles.

To summarize this section, we have seen that nontrivial external fields will lead in general to particle production from an initial "vacuum" state. Applied to the problem at hand, due to the time dependence of the background space–time, quantum fluctuations will be produced from an initial "vacuum" state. This process is responsible for example for the generation of perturbations in inflationary universe models.

Having defined a vacuum state $|\psi_0\rangle$ and constructed the Fock basis of the Hilbert space of states by acting on $|\psi_0\rangle$ with products of creation operators a_J^+, we can calculate the expectation value for any combination of field operators at an arbitrary moment of time η in the state $|\psi_0\rangle$ and in any other state of the Hilbert space. However, the prescription (11.17) only makes sense if $E^2 > 0$ for all values of k_J. From (11.11b) it is obvious that this will be the case if $z''/z \leq 0$. For hydrodynamical matter with a time-independent equation of state $p = \frac{1}{3}\rho$ the above condition is satisfied since $z'' = 0$. This can be verified using the explicit expression for $z(\eta)$ given in (10.42) and the expression for β from (10.15). Hence, in a radiation-dominated universe

$$v_J(\eta_0) = 1/\sqrt{w_J} , \qquad v_J'(\eta_0) = \mathrm{i}\sqrt{w_J} , \quad w_J = k_J/\sqrt{3} . \tag{11.24}$$

In the general case and in particular for scalar-field matter and for higher-derivative gravity, the applicability of (11.17) is no longer assured. In particular, during the inflationary period

$$z''/z \approx a''/a > 0 , \tag{11.25}$$

and hence (11.17) is inapplicable. However, in this case it is possible to define the so-called de Sitter invariant vacuum [128] given by the conditions

$$v_k(\eta_0) = (1/k^{3/2})(\mathscr{H}_0 + \mathrm{i}k)\exp(\mathrm{i}k\eta_0) , \qquad v_k'(\eta_0) = (\mathrm{i}/k^{1/2})(\mathscr{H}_0 + \mathrm{i}k - \mathrm{i}\mathscr{H}_0'/k)\exp(\mathrm{i}k\eta_0) \tag{11.26}$$

(restricting attention to the case $\mathscr{K} = 0$). Here, $\mathscr{H}_0 = a'(\eta_0)/a(\eta_0)$. Note that for $k \gg \mathscr{H}_0$, the prescriptions (11.17) and (11.26) converge.

The above result that the vacuum definitions (11.17) and (11.26) agree points to a fairly general feature. The ambiguity in the definition of the vacuum is related to the ambiguity in the definition of the notion of particles for modes with wavelength larger than the curvature radius of the background. For physically reasonable vacuum definitions [in particular for those given by the same prescription as (11.17) evaluated at different times], the leading terms in the k expansions of $v_k(\eta_0)$ and $v_k'(\eta_0)$ agree for large k, whereas for small k, $v_k(\eta_0)$ and $v_k'(\eta_0)$ may depend very sensitively on the particular choice of vacuum.

Fortunately, for the applications we have in mind, i.e., the computation of perturbation spectra in inflationary universe models, the results depend only on the short-wavelength part of the initial vacuum spectrum which is independent of most choices of the vacuum. Quite generally, $v_k(\eta_0) \sim k^{-1/2}$ and $v_k'(\eta_0) \sim k^{1/2}$ as $k \to \infty$. Hence, in the following we shall use the following general initial conditions to define the vacuum at $\eta = \eta_0$:

$$v_k(\eta_0) = k^{-1/2} M(k\eta_0) , \qquad v_k'(\eta_0) = \mathrm{i} k^{1/2} N(k\eta_0) , \tag{11.27}$$

where [as follows from (11.14)] N and M obey the normalization condition

$$NM^* + N^*M = 2 ; \qquad |M(k\eta_0)| \to 1 , \quad |N(k\eta_0)| \to 1 \quad \text{for} \quad k\eta_0 \gg 1 . \tag{11.28}$$

12. Spectrum of density perturbations for hydrodynamical matter

In the following three chapters we apply the quantization prescription developed in chapter 11 to calculate the spectra of density perturbations generated from quantum fluctuations in several models. In this chapter, a model with hydrodynamical matter will be considered [42]. Chapter 13 will be an analysis of inhomogeneities in models with scalar-field matter, and in Chapter 14 higher-derivative theories of gravity will be analyzed.

In the first section we will derive the mathematical relation between the quantities used to describe fluctuations and the quantization variables introduced in chapter 10. Next, the generation of long-wavelength (wavelength much larger than the Hubble radius during the excitation period) perturbations will be studied. In the third section, the results will be applied to a pure hydrodynamical universe ($p + \varepsilon \simeq \varepsilon$), and to a model with a time-dependent cosmological constant. The second example will illustrate why quantum perturbations are able to generate sufficiently large fluctuations on galactic scales today if during the early universe there was a period in which the energy–momentum tensor was dominated by the cosmological constant, whereas in models with positive p the quantum perturbations always remain small [see also part I, chapter 6 and in particular, eq. (6.67) for an entirely classical point of view on this issue].

12.1. Measures of fluctuations

In this review article only linearized theories are considered. In this case, the spectrum of perturbations is in general Gaussian. Heuristically speaking (in the case $\mathcal{K} = 0$), there are no nonrandom correlations between fluctuations on different wavelengths. This implies that the spectrum can be characterized entirely by the two-point correlation function of the gauge-invariant potential Φ

$$\xi_\Phi(\eta, \boldsymbol{r}) \equiv \overline{\Phi(\eta, \boldsymbol{x}) \Phi(\eta, \boldsymbol{x} + \boldsymbol{r})} , \tag{12.1}$$

where by an overbar we denote spatial averaging.

For the purpose of comparing with observations, we also need the two-point correlation function of the gauge-invariant energy-density fluctuation

$$\xi_\varepsilon(\eta, r) \equiv \overline{\frac{\delta\varepsilon}{\varepsilon_0}(\eta, x)\,\frac{\delta\varepsilon}{\varepsilon_0}(\eta, x+r)}\,. \tag{12.2}$$

For inhomogeneities with length scale smaller than the Hubble radius, $\delta\varepsilon/\varepsilon_0$ to first approximation coincides with the energy-density perturbation in any gauge (gauge differences are only dominant on scales larger than H^{-1}). The relation between ξ_Φ and ξ_ε can be derived from (4.15). Considering wavelengths smaller than the Hubble radius, only the term on the left-hand side proportional to $\Delta\Psi$ contributes. Using $\Psi = \Phi$ we obtain

$$\xi_\varepsilon(\eta, r) \simeq \tfrac{4}{9}(1/l^4\varepsilon^2a^4)\Delta^2\xi_\Phi(\eta, r)\,. \tag{12.3}$$

Expressions (12.1) and (12.2) are valid in the classical theory. In quantum theory we must substitute the Heisenberg operators $\hat{\Phi}$ and $\delta\hat{\varepsilon}$ instead of the classical variables Φ and $\delta\varepsilon$. In addition, the spatial average becomes the quantum expectation value in the state $|\psi\rangle$ of the system (various choices for which were discussed in the previous chapter). Thus, for example,

$$\xi_\Phi(\eta, r) = \langle\psi|\hat{\Phi}(\eta, x+r)\hat{\Phi}(\eta, x)|\psi\rangle\,. \tag{12.4}$$

If the quantum state $|\psi\rangle$ is homogeneous, then the right-hand side of (12.4) is indeed independent of x.

However, the quantization prescription developed in chapter 11 is not in terms of Φ, but rather in terms of the velocity potential v. To relate these variables, we return to the equations for hydrodynamical matter derived in chapter 10. Rewritten in terms of gauge-invariant variables, eqs. (10.44), (10.53) and (10.54) read

$$\Delta\Phi - 3\mathcal{H}\Phi' - 3\mathcal{H}^2\Phi + 3\mathcal{K}\Phi = \beta(3\Phi - \xi^{(\mathrm{gi})j}_{\ ,j})\,, \tag{12.5}$$

$$\Phi' + \mathcal{H}\Phi = -\tfrac{1}{2}(c_s\beta^{1/2}/a)\varphi_v^{(\mathrm{gi})}\,, \tag{12.6}$$

$$3\Phi - \xi^{(\mathrm{gi})j}_{\ ,j} = -(1/c_s^2)\Phi - \tfrac{1}{2}(c_s\beta^{1/2}a\varphi_v^{(\mathrm{gi})})'/c_s^2\beta a^2\,, \tag{12.7}$$

where $\xi^{(\mathrm{gi})}$ is the gauge-invariant shift vector associated with ξ, defined as in (5.15). Substituting (12.7) in the right-hand side of (12.5) and expressing $\varphi^{(\mathrm{gi})}$ and Φ' exclusively in terms of Φ and v by means of (10.43a) and (12.6), one finds the following relation between Φ and v:

$$\Delta\Phi = -\sqrt{\tfrac{3}{2}}\,l(\beta/\mathcal{H}c_s^2)(v/z)'\,, \tag{12.8}$$

where β and z were defined in (10.15) and (10.43b). This relation is also valid on the quantum level if we replace Φ and v by the corresponding operators.

For adiabatic perturbations, the operators analog of the equation of motion (5.22) for Φ is

$$\hat{\Phi}'' + 3\mathcal{H}(1+c_s^2)\hat{\Phi}' - c_s^2\Delta\hat{\Phi} + [2\mathcal{H}' + (1+3c_s^2)(\mathcal{H}^2 - \mathcal{K})]\hat{\Phi} = 0\,. \tag{12.9}$$

We can also expand $\hat{\Phi}$ in terms of the basis $\psi_J(x)$ of eigenfunctions of Δ (the same basis used to expand $\hat{v}$ in the previous chapter) and the associated creation and annihilation operators a_J^+ and a_J^-,

$$\hat{\Phi} = \sqrt{\tfrac{3}{4}}\, l\, \frac{\beta^{1/2}}{a} \int \mathrm{d}J\, [\psi_J(\mathbf{x}) u_J^*(\eta) a_J^- + \psi_J^*(\mathbf{x}) u_J(\eta) a_J^+] . \tag{12.10}$$

By inserting (12.10) into (12.9) we find that the temporal mode functions $u_J(\eta)$ satisfy the following equation:

$$u_J''(\eta) + [c_s^2 k_J^2 - (1/c_s z)''(1/c_s z)^{-1}] u_J(\eta) = 0 . \tag{12.11}$$

Comparison with (12.8) leads to the following relation between $u_J(\eta)$ and the mode functions $v_J(\eta)$ arising in the expansion (11.12) of the operator $\hat{v}$:

$$u_J(\eta) = (z/k_J^2 c_s)(v_J(\eta)/z)' . \tag{12.12}$$

Let us assume that the state $|\psi\rangle$ of the system is the "initial vacuum state" $|\psi_0\rangle$ determined by the initial conditions (11.17) for the mode functions $v_J(\eta)$. This also determines the initial values at time η_0 for the mode functions $u_J(\eta)$ by (12.12). At this point, the correlation function $\xi_\Phi(\eta, \mathbf{r})$ can be computed by inserting (12.10) into the defining relation (12.4) and by using the standard canonical commutation relations. Note that in this calculation, no renormalization is required for correlation functions. The result is

$$\xi_\Phi(\eta, r) = \begin{cases} \displaystyle\int |\delta_\Phi(\eta, k)|^2 \frac{\sin kr}{k f(r)} \frac{\mathrm{d}k}{k} , & \mathcal{K} = 0, -1 , \\ \displaystyle\sum_{k=1}^{\infty} |\delta_\Phi(\eta, k)|^2 \frac{\sin kr}{k^2 f(r)} , & \mathcal{K} = 1 , \end{cases} \tag{12.13}$$

where

$$|\delta_\Phi(\eta, k)|^2 = (3l^2/8\pi^2)(\beta/a^2)|u_k|^2 k^3 \tag{12.14}$$

is a measure of the square of the amplitude of fluctuations in Φ at comoving wavelength $1/k$ and k is the wavenumber which is related to the eigenvalue k_J^2 of Δ by

$$k_J^2 = k^2 - \mathcal{K} . \tag{12.15}$$

As is already implicit in the notation of (12.13), for flat and open universes k is a continuous variable ranging from 0 to ∞ whereas in a closed universe, k takes on integer values $1, 2, 3, \ldots$ The function $f(r)$ is

$$f(r) = \sinh r , \quad \mathcal{K} = -1 ; \qquad f(r) = r , \quad \mathcal{K} = 0 ; \qquad f(r) = \sin r , \quad \mathcal{K} = 1 , \tag{12.16}$$

and r is the geodesic comoving distance between $\mathbf{x}$ and $\mathbf{x} + \mathbf{r}$. In deriving (12.13) and (12.14), we used the following forms of the measure $\mathrm{d}J$:

$$\mathrm{d}J = \int_0^\infty \mathrm{d}k \sum_{l=0}^{\infty} \sum_{m=-l}^{l} , \quad \mathcal{K} = 0, -1 ; \qquad \mathrm{d}J = \sum_{k=1}^{\infty} \sum_{l=0}^{k-1} \sum_{m=-l}^{l} , \quad \mathcal{K} = 1 , \tag{12.17}$$

where l and m are the standard angular-momentum eigenvalues. We also made use of a handy summation formula for the eigenfunctions of Δ (see for example ref. [127]),

$$\sum_{l,m} \psi_J^*(x)\psi_J(x+r) = \frac{k}{2\pi^2}\,\frac{\sin kr}{f(r)}\,. \tag{12.18}$$

Note that the integral (sum) in (12.13) may diverge. Sometimes, physically motivated subtraction schemes are possible (see, e.g., ref. [118]). However, in this article we will focus on single Fourier modes since the main interest will be the amplitude and spectrum (i.e., k dependence) of the power spectrum.

To simplify the algebra, we shall consider the case of a flat universe ($\mathcal{K}=0$). Combining (12.14) and (12.12), the expression for the metric perturbations on scale k becomes

$$|\delta_\Phi(\eta,k)| = \left(\frac{3l^2}{8\pi^2}\right)^{1/2} \frac{\beta^{1/2}z}{ac_s}\,\frac{1}{k^{1/2}}\left|\left(\frac{v_k(\eta)}{z}\right)'\right|, \tag{12.19}$$

where the mode functions $v_k(\eta)$ satisfy the equation

$$v_k''(\eta) + E_k^2 v_k(\eta) = 0\,, \quad E_k^2 = c_s^2k^2 - z''/z\,, \tag{12.20}$$

with initial conditions [see (11.11a, b) and (11.17)]

$$v_k(\eta_0) = E_k^{-1/2}(\eta_0)\,, \quad v_k'(\eta_0) = \mathrm{i}E_k^{1/2}(\eta_0)\,, \tag{12.21}$$

at some time η_0 where we suppose that $z''/z=0$, e.g., when $p=\varepsilon/3$ (at which time the state $|\psi_0\rangle$ is the vacuum state).

12.2. Spectrum of long-wavelength perturbations

In a radiation dominated universe with equation of state $p=\frac{1}{3}\varepsilon$, it follows from (10.15) and (10.43b) that $z'(\eta)=0$. Hence, the potentially time-dependent effective potential U vanishes,

$$U = -z''/z = 0\,. \tag{12.22}$$

Thus, the equation for the modes is a simple harmonic-oscillator equation and there is no particle production.

Let us now assume that the equation of state differs from $p=\frac{1}{3}\varepsilon$ and the condition $U=0$ is violated in some finite time interval $\eta_0<\eta<\eta_e$. We will calculate the amplitude of perturbations whose wavelength is larger than the Hubble radius during the above time interval, i.e., $k\eta_e \ll 1$. For applications to cosmology, this is a useful calculation, since in the very early universe when deviations from $U=0$ are expected, galactic scales are far outside the Hubble radius.

The solution of (12.20) with initial conditions (12.21) takes the following form for $\eta<\eta_0$:

$$v_\omega(\eta) = (1/\sqrt{\omega})\exp[\mathrm{i}\omega(\eta-\eta_0)]\,, \quad \omega = c_sk = (1/\sqrt{3})k\,. \tag{12.23}$$

For $\eta>\eta_e$, the potential U again vanishes and $v_\omega(\eta)$ can be written as

$$v_\omega(\eta) = (1/\sqrt{\omega})\{\alpha^+_\omega \exp[i\omega(\eta - \eta_e)] + \alpha^-_\omega \exp[-i\omega(\eta - \eta_e)]\}\,, \tag{12.24}$$

where the coefficients α^+_ω and α^-_ω obey the following normalization condition [see the third of eqs. (11.20)],

$$|\alpha^+_\omega|^2 - |\alpha^-_\omega|^2 = 1\,. \tag{12.25}$$

These coefficients can be written in terms of $v_\omega(\eta_e)$ and $v'_\omega(\eta_e)$,

$$\alpha^\pm_\omega = (1/2i\omega^{1/2})[\pm v'_\omega(\eta_e) + i\omega v_\omega(\eta_e)]\,. \tag{12.26}$$

To calculate $v_\omega(\eta_e)$ and $v'_\omega(\eta_e)$ we use the integral form of the differential equation (12.20) for the modes $v_k(\eta)$,

$$v_k(\eta) = z(\eta)\left(\frac{v^0_k}{z_0} + (v'^0_k z_0 - v^0_k z'_0)\int_{\eta_0}^{\eta} \frac{d\eta}{z^2} - k^2 \int_{\eta_0}^{\eta} \frac{d\tilde{\eta}}{z^2} \int_{\eta_0}^{\tilde{\eta}} c_s^2 z v_k\, d\eta\right), \tag{12.27}$$

where the superscript 0 indicates that the expression is to be evaluated at time η_0. v^0_k and v'^0_k are given by (12.21).

So far, all the equations are exact. Now we make an approximation which is good only for long wavelengths. The last (and most complicated) term in (12.27) will be neglected since it is suppressed compared to the other ones by a factor of $k\eta \ll 1$. In this case, the expressions for $v_k(\eta_e)$ and $v'_k(\eta_e)$ follow directly from (11.27), inserting them into (11.26) we obtain

$$\alpha^\pm_\omega = (1/2i\omega)[\pm C + i\omega D^\pm + \omega O(\omega\eta_e)]\,, \tag{12.28}$$

$$C = z'_0 z'_e \int_{\eta_0}^{\eta_e} \frac{1}{z^2}\, d\eta + \frac{z'_0}{z_e} - \frac{z'_e}{z_0} = z'_0 z'_e \int_{\eta_0}^{\eta_e} \frac{z''}{z'^2 z}\, d\eta\,, \tag{12.29}$$

$$D^\pm = z'_0/z'_e \pm z'_e/z'_0 + (z_e/z'_e \mp z_0/z'_0)C\,, \tag{12.30}$$

where both (12.29) and (12.30) are independent of ω. The symbol $O(\omega\eta_e)$ indicates higher-order terms in $\omega\eta_e$. The integral indentity used in (12.29) is a result of integration by parts.

A second approximation is to assume $|C| \gg |\omega D|$, which will be true if the rate of change of z for $\eta_0 \leq \eta \leq \eta_e$ is larger than ω (strong violation of the condition $U = 0$). In this case, the first term in (12.28) dominates and $\alpha^\pm_\omega \simeq \pm C/2i\omega$. Substituting this into (12.24) and the resulting expression for $v_\omega(\eta)$ into (12.19), we obtain the following result for the measure of metric perturbations with wavenumber $k = \sqrt{3}\,\omega$ for times $\eta > \eta_e$:

$$|\delta_\Phi(\eta, \omega)| \simeq (3\sqrt{3}/4\pi^2)^{1/2}(lC/a\omega^2\eta^2)|\omega\eta \cos\omega(\eta - \eta_e) - \sin\omega(\eta - \eta_e)|\,. \tag{12.31}$$

Note that this result is valid provided $\omega\eta_e \ll 1$, i.e., that the wavelength is larger than the Hubble radius at the end of the period when the potential $U \neq 0$. However, there are no conditions on $\omega\eta$. We can also use this result for η after the scale of the perturbation has entered the Hubble radius. However, as long as this is not the case, i.e., for $\omega \ll 1/\eta \ll 1/\eta_e$, it follows from (12.31) that

$$|\delta_\Phi(\eta, \omega)| \simeq (3\sqrt{3}/4\pi^2)^{1/2}(lC/a_e)\eta_e\omega\ ; \quad \omega\eta \ll 1\ , \tag{12.32}$$

where, as we stress again, C is independent of ω.

Galactic scales are well inside the Hubble radius at the time of recombination. When evaluating (12.31) for these times and scales, $\omega\eta \gg 1$ and only the first term in (12.31) contributes, and

$$|\delta_\Phi(\eta, \omega)| \simeq (3\sqrt{3}/4\pi^2)^{1/2}[lC/a(\eta)\omega\eta]|\cos\omega(\eta - \eta_e)|\ . \tag{12.33}$$

for $1/\eta \ll \omega \ll 1/\eta_e$. In this case, we can apply (12.3) to find $|\delta_\varepsilon(\eta, \omega)|$ and from it to calculate the two-point correlation function of the energy-density perturbation,

$$\xi_\varepsilon(\eta, r) = \int |\delta_\varepsilon(\eta, k)|^2 \frac{\sin kr}{kr}\frac{\mathrm{d}k}{k}\ , \tag{12.34}$$

where $\delta_\varepsilon(\eta, \omega)$ in the range $1/\eta \ll \omega \ll 1/\eta_e$ is given by

$$|\delta_\varepsilon(\eta, \omega)| \simeq (3\sqrt{3}/\pi^2)^{1/2}(lC\eta_e/a_e)\omega|\cos\omega(\eta - \eta_e)|\ . \tag{12.35}$$

The important conclusion to be drawn from (12.33) and (12.35) is that $|\delta_\Phi(k)| \sim k^{-1}$ and $|\delta_\varepsilon(k)| \sim k$ on scales which were outside the Hubble radius when the perturbations were generated, but which are inside when the fluctuations are evaluated, i.e. $1/\eta \ll \omega \ll 1/\eta_e$. On scales which are still outside the Hubble radius at time η, i.e., $\omega \ll 1/\eta$, the spectrum is different, namely $|\delta_\Phi(k)| \sim |\delta_\varepsilon^{(\mathrm{gi})}(k)| \sim k$ as follows from (12.32).

It is possible to generalize the analysis to the more realistic case when at late times (around recombination) the equation of state is again time-dependent, as it will be in a universe containing both cold matter and radiation. For scales obeying $1/c_s\eta \ll k \ll 1/\eta_e$, the mode equation (12.20) can be solved in a WKB approximation with the result

$$v_k(\eta) \simeq \frac{\sqrt{3}C}{2k^{3/2}c_s^{1/2}}\left[\exp\left(\mathrm{i}\int kc_s\,\mathrm{d}\eta\right) - \exp\left(-\mathrm{i}\int kc_s\,\mathrm{d}\eta\right)\right]. \tag{12.36}$$

Substituting this solution in (12.19) and using the relationship (12.3) between δ_k and δ_ε, we obtain

$$|\delta_\varepsilon(\eta, k)| \simeq \frac{\sqrt{3}}{2\pi}\left[\left(1 + \frac{p}{\varepsilon}\right)\frac{1}{c_s\varepsilon a^4}\right]^{1/2} kC\left|\sin\left(\int kc_s\,\mathrm{d}\eta\right)\right|,\quad \frac{1}{c_s\eta} \ll k \ll \frac{1}{\eta_e}\ . \tag{12.37}$$

Thus, the spectrum has the same shape as in (12.35).

To summarize this section: for hydrodynamical perturbations generated during a period $\eta_0 < \eta < \eta_e$ when their scale was outside the horizon, i.e. $k\eta_e \ll 1$, the shape of the spectrum is independent of the particular mechanism for the generation, provided that the initial quantum state is the vacuum. The spectral dependence at a late time $\eta \gg \eta_e$ is

$$\delta_\Phi(k) \sim \delta_\varepsilon(k) \sim k\ , \quad k \ll 1/\eta \ll 1/\eta_e \tag{12.38}$$

(scale outside the Hubble radius at time η), and

$$\delta_\Phi(k) \sim k^{-1}\,, \qquad \delta_\varepsilon(k) \sim k\,, \quad 1/\eta \ll k \ll 1/\eta_e \tag{12.39}$$

(scale inside the Hubble radius at time η). The amplitude of the spectrum depends sensitively on the particular model and will be discussed in the following section.

12.3. The amplitude of the spectrum of adiabatic perturbations

In this section it will be shown that in order for the quantum fluctuations in a universe with hydrodynamical matter to obtain a large amplitude on galactic scales today, the background model has to have an equation of state with a time-dependent cosmological constant which, although negligible today, was dominant in the very early universe.

First, we consider a universe without cosmological term in which $p = \gamma\varepsilon$ with $0 < \gamma < 1$ and $\gamma \neq 1/3$ during the period $\eta_0 < \eta < \eta_e$ in which the equation of state differs from that of a radiation-dominated universe. We will assume that the velocity of sound c_s is constant during this period. Then, using the background equations of motion (10.14a, b) and (10.15), it can be shown that the effective potential U is

$$U = -z''/z = \tfrac{1}{2} l^2 \varepsilon a^2 (1 - 3c_s^2)\,. \tag{12.40}$$

Substituting (12.40) in (12.29), one estimates the coefficient C, and then from (12.37) the following estimate of density perturbations on scales

$$t_m a(t_r)/a(t_m) \ll \lambda(t_r) \ll t_r \tag{12.41}$$

evaluated at the time t_r of recombination:

$$|\delta_\varepsilon(\lambda(t_r))| \sim \frac{a(t_r)}{a(t_m)} \frac{t_m}{t_r} \left(\frac{\varepsilon_m}{\varepsilon_{Pl}}\right)^{1/2} \frac{t_r}{\lambda(t_r)} (1 - 3c_s^2)\,, \tag{12.42}$$

where $\varepsilon_{Pl} \sim l^{-4}$ is the Planck density, $t = \int a \, d\eta$ is physical time and $\lambda(t_r) \sim a(t_r)/k$ is the physical scale of the inhomogeneities. The index m indicates that the corresponding variables are evaluated at the time t_m when U/z'^2 takes on its maximum.

From (12.42) it follows that the perturbation amplitude δ_ε increases as t_m decreases. However, for $t_m < l$ the perturbative approach breaks down. Thus, within the limits of the present theory, the maximal value of δ_ε is attained for $t_m \sim l$. Since $a(t) \sim t^{2/3(\gamma+1)}$ if $\gamma = \text{const}$, the maximum value of (4.12) is taken on for $\gamma = 0$. With $a(t_r)/a(t_{Pl}) \sim 10^{37}$ we obtain the following upper bound on δ_ε: $\delta_\varepsilon \leq 10^{-18} t_r/\lambda(t_r)$. For scales corresponding to galaxies and clusters, $t_r/\lambda(t_r) \sim 1$. In order to form structures by linear growth of perturbations, an amplitude $\delta_\varepsilon \sim 10^{-4}$ is required [105]. Hence, we conclude that initial quantum fluctuations are insufficient to generate the primordial perturbations needed for galaxy formation if the equation of state always has positive pressure. This conclusion does not change if we allow for rapid oscillations of the equation of state $p(\varepsilon)$ during the interval $\eta_0 < \eta < \eta_e$.

Let us now turn to a model with a "quasi-vacuum" epoch, an epoch with a time-dependent cosmological constant which yields an effective equation of state satisfying $p + \varepsilon \ll \varepsilon$. Recall that a cosmological constant Λ gives

$$p = -\varepsilon = -\Lambda/3l^2 . \tag{12.43}$$

In this subsection, we will demonstrate that in this case it is in principle possible for quantum fluctuations to generate perturbations of suffícently high amplitude for galaxy formation. Since the particular model is not very realistic, only rough estimates of the magnitude of the effects will be given. Realistic models which yield time intervals with an equation of state similar to (12.43) will be analyzed in the following chapters.

To justify the hydrodynamical approximation, we shall assume that during some period in the early universe, there is some amount of radiation in addition to the cosmological constant, i.e.,

$$\varepsilon = \varepsilon_v + \varepsilon_r , \qquad p = -\varepsilon_v + \tfrac{1}{3}\varepsilon_r , \tag{12.44a, b}$$

where ε_r is the energy density in radiation which decays as a^{-4}. The vacuum energy ε_v is time-independent and homogeneous in space. Hence, matter perturbations are inhomogeneities in radiation. Therefore, they can be described using the hydrodynamical formalism discussed in this chapter. The only role of the vacuum energy is in determining the evolution of the background in the "quasi-vacuum" epoch. We assume that the vacuum is metastable and decays over a time interval t_d. In order to exclude unphysical effects due to instantaneous phase transitions, the transition is taken to be regular. The phase transition produces relativistic particles [107].

Initially at some early time when $\varepsilon_r \gg \varepsilon_v$, $c_s^2 = \frac{1}{3}$ and the effective potential $U \approx 0$, ($z \propto a$, $a \propto \eta$, $U \propto z''$). At a time t_m, $\varepsilon_r(t_m) = \varepsilon_v$, and for $t > t_m$ the period of quasi-vacuum expansion starts. (We are assuming that the vacuum decays well after t_m, i.e., $t_d \gg t_m$.) Since during the "quasi-vacuum" epoch it can be seen from the definition of z that

$$z \sim a[(\varepsilon + p)/\varepsilon]^{1/2} \sim a^{-1} \sim \eta^{-1} , \tag{12.45}$$

the effective potential $U \propto z''$ also vanishes for $t \gg t_m$. Hence the main contribution to the constant C in (12.29) comes from times $t \sim t_m$, and we can estimate δ_ε as in the previous section [see also (10.43a)] and obtain

$$\delta_\varepsilon(\lambda(t_r))| \sim (\varepsilon_v/\varepsilon_{Pl})^{1/4}\lambda_\gamma[a(t_d)/a(t_m)]/\lambda(t_r) , \tag{12.46}$$

where t_d denotes the end of the quasi-vacuum period and where λ_γ is the characteristic wavelength of the background radiation at the time t_r of recombination. Equation (12.46) is valid on scales satisfying

$$t_m a(t_r)/a(t_m) \sim \lambda_\gamma(\varepsilon_v/\varepsilon_{Pl})^{-1/4}[a(t_d)/a(t_m)] < \lambda(t_r) < t_r . \tag{12.47}$$

This is the condition for applicability of the long-wavelength approximation ($k\eta_d \ll 1$).

Combining (12.46) and (12.47), we may conclude that the maximum amplitude of δ_ε is given by

$$|\delta_\varepsilon| < (\varepsilon_v/\varepsilon_{Pl})^{1/2} . \tag{12.48}$$

The above analysis is only applicable if $\varepsilon_v < \varepsilon_{Pl}$. If $\varepsilon_v \sim 10^{-8}\varepsilon_{Pl}$, quantum fluctuations in this model are of the right order of magnitude to generate the primordial perturbations required for galaxy formation. This amplitude is obtained on scales

$$\lambda(t_r) \sim \lambda_\gamma(\varepsilon_{Pl}/\varepsilon_v)^{1/4} a(t_d)/a(t_m) . \tag{12.49}$$

Provided that the quasi-vacuum period lasts long enough, $t_d \geq 50(\varepsilon_v l^2)^{-1/2}$, this scale is larger than or equal to galactic scales.

The hydrodynamical formalism can only be applied on scales which contain more than one particle. This condition is satisfied on scales

$$\lambda(t_m) > l(\varepsilon_v/\varepsilon_{Pl})^{-1/4} , \tag{12.50}$$

or, evaluated at t_r,

$$\lambda(t_r) > \lambda_\gamma a(t_d)/a(t_m) . \tag{12.51}$$

Thus, for $\varepsilon_v < \varepsilon_{Pl}$, (12.51) is satisfied on scales where the long-wavelength approximation can be applied [see (12.47)].

If $t_d \gg 50(\varepsilon_v l^2)^{-1/2}$, then the hydrodynamical and long-wavelength approximations hold only on scales which are much larger than galactic scales. In this case the field-theoretic analysis of chapter 13 is needed to calculate the density perturbations on scales which are smaller than the ones considered above.

In this section, we have demonstrated that it is in principle possible to generate sufficiently large primordial density perturbations for galaxy formation from quantum fluctuations. The physical reason is the following: the amplitude of the perturbation is defined by the smoothness of the decay hypersurface. If this surface is very inhomogeneous, large fluctuations will be produced. In our case, the information about the smoothness of the hypersurface is entirely contained in v and v' which must be continuous on this surface.

During the period $t_m \leq t \leq t_d$ of quasi-exponential expansion, the nondecaying mode of v is practically constant since $U \rightarrow 0$. Thus, it saves the information about inhomogeneities on corresponding comoving scales. If this period is sufficiently long, i.e., $t_d \geq 50(\varepsilon_v l^2)^{-1/2}$, then the comoving scale $t_m a(t_m)^{-1}$ on which the fluctuations can be large if t_m is small becomes comparable with galactic scales. Now, provided ε_v is not much smaller than ε_{Pl}, the metric fluctuations on this scale will be large at t_m since the physical length is comparable to the Planck length. In fact, the relative fluctuations on Planck scales are of order unity. As a result of the exponential expansion, the information about large inhomogeneities at time t_m on a physical scale t_m becomes information characterizing the inhomogeneity of vacuum decay hypersurface on galactic comoving scales. Thus, the role of the de Sitter (quasi-vacuum) period in generating large fluctuations is to expand Planck scales on which fluctuations are large to galactic scales.

One can verify that the gauge-invariant potential Φ characterizing the metric perturbations decays exponentially during the "quasi-vacuum" epoch. Thus the metric perturbations at the end of this stage are negligibly small. However, during the phase transition, the variables β and z in eq. (12.8) change by large factors while v has essentially a constant magnitude. As a result, significant fluctuations of the metric are created. However, they are due to the inhomogeneity of the vacuum decay hypersurface and occur after the phase transition.

To calculate the amplitude of density perturbations more quantitatively, an understanding of the physical processes which determine the decay hypersurface is required. This will be discussed in the next chapters.

13. Spectrum of density perturbations in inflationary universe models with scalar-field matter

Many of the currently popular models of structure formation assume that the primordial energy-density perturbations are due to quantum vacuum fluctuations which were present just before a period of exponential expansion of the scale factor $a(t)$ in the very early universe. In this chapter, we will calculate the spectrum of density perturbations produced in an inflationary universe with scalar-field matter [20].

Starting point is the equation of motion (6.49) for the gauge-invariant potential Φ,

$$\Phi'' + 2(a/\varphi_0')'(\varphi_0'/a)\Phi' - \Delta\Phi + 2\varphi_0'(\mathcal{H}/\varphi_0')'\Phi = 0\,. \tag{13.1}$$

This equation can also be directly derived starting from the action (10.68). Varying (10.68) with respect to ϕ and ψ and setting $B - E' = 0$ at the end of the calculation (i.e., going to longitudinal gauge) gives $(\mathcal{K} = 0)$

$$\Delta\Psi - 3\mathcal{H}\Psi' - (\mathcal{H}' + 2\mathcal{H}^2)\Phi = \tfrac{3}{2}l^2(\varphi_0'\,\delta\varphi^{(\mathrm{gi})\prime} + V_{,\varphi}a^2\,\delta\varphi^{(\mathrm{gi})})\,, \tag{13.2}$$

$$\tfrac{1}{3}\Delta(\Phi - \Psi) + \Psi'' + \mathcal{H}\Phi' + 2\mathcal{H}\Psi' + (\mathcal{H}' + 2\mathcal{H}^2)\Phi = \tfrac{3}{2}l^2(\varphi_0'\,\delta\varphi^{(\mathrm{gi})\prime} - V_{,\varphi}a^2\,\delta\varphi^{(\mathrm{gi})})\,. \tag{13.3}$$

Variation with respect to $B - E'$ yields the constraint equation (10.70)

$$\Psi' + \mathcal{H}\Phi = \tfrac{3}{2}l^2\,\varphi_0'\,\delta\varphi^{(\mathrm{gi})}\,. \tag{13.4}$$

Substituting $\delta\varphi^{(\mathrm{gi})}$ from (13.4) in (13.3) leads to $\Phi = \Psi$, and, combining (13.4) and (13.2) we recover the above equation of motion (13.1) for Φ.

For the sake of simplicity, we consider only a spatially flat universe $(\mathcal{K} = 0)$. In this case, the field operator $\hat{\Phi}$ can be expanded in a Fourier integral in terms of creation and annihilation operators a_k^+ and a_k^-,

$$\hat{\Phi}(\boldsymbol{x}, \eta) = \frac{1}{\sqrt{2}}\,\frac{\varphi_0'}{a}\int \frac{\mathrm{d}^3k}{(2\pi)^{3/2}}\,[u_k^*(\eta)\,\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}}a_k^- + u_k(\eta)\,\mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}}a_k^+]\,. \tag{13.5}$$

From (13.1), it follows that the mode functions $u_k(\eta)$ satisfy the equation

$$u_k''(\eta) + [k^2 - (1/z)''(1/z)^{-1}]u_k(\eta) = 0\,, \quad z = a\varphi_0'/\mathcal{H}. \tag{13.6}$$

The connection between $\hat{\Phi}$ and the operator $\hat{v}$ used in the general quantization scheme of chapter 10 follows from (13.2) making use of (13.4) and the definition (10.71) of v to express the variables $\delta\hat{\varphi}^{(\mathrm{gi})}$ and $\hat{\Psi}'$ in (13.2) exclusively in terms of $\hat{v}$ and $\hat{\Psi}$,

$$\Delta\hat{\Phi} = \tfrac{3}{2}l^2(\varphi_0'^2/\mathcal{H})(\hat{v}/z)' . \tag{13.7}$$

For the Fourier mode coefficient functions $u_k(\eta)$ and $v_k(\eta)$, the relation is

$$u_k(\eta) = -\tfrac{3}{2}l^2(z/k^2)(v_k/z)' . \tag{13.8}$$

In chapter 11, the initial conditions for v_k were expressed in terms of functions $M(k\eta_i)$ and $N(k\eta_i)$ (η_i being the initial time) which obey a simple normalization condition (11.28) and tend to 1 in the limit $k\eta_i \gg 1$ [see (11.27)]. Using (13.8), the initial conditions for $u_k(\eta)$ become

$$\begin{aligned} u_k(\eta_i) &= -\tfrac{3}{2}l^2\left(\frac{i}{k^{3/2}}\, N(k\eta_i) - \frac{z'(\eta_i)}{z(\eta_i)}\,\frac{1}{k^{5/2}}\, M(k\eta_i)\right), \\ u_k'(\eta_i) &= -\tfrac{3}{2}l^2\left[\frac{1}{k^{3/2}}\, M(k\eta_i) + 3\,\frac{z'(\eta_i)}{z(\eta_i)}\left(\frac{i}{k^{3/2}}\, N(k\eta_i) - \frac{z'(\eta_i)}{z(\eta_i)}\,\frac{1}{k^{5/2}}\, M(k\eta_i)\right)\right]. \end{aligned} \tag{13.9}$$

Before considering the solution of (13.6), we will define the power spectrum of metric perturbations. The power spectrum $|\delta_k|^2$ is a measure of the two-point correlation function of $\hat{\Phi}$,

$$\langle 0|\hat{\Phi}(\boldsymbol{x},\eta)\hat{\Phi}(\boldsymbol{x}+\boldsymbol{r},\eta)|0\rangle = \int_0^\infty \frac{\mathrm{d}k}{k}\,\frac{\sin(kr)}{kr}\,|\delta_k|^2 . \tag{13.10}$$

Inserting the expansion (13.5) into the left-hand side of this equation and making use of the canonical commutation relations, we obtain

$$|\delta_k(\eta)|^2 = (1/4\pi^2)(\varphi_0'^2/a^2)|u_k(\eta)|^2k^3 , \tag{13.11}$$

where $|\delta_k|^2$ characterizes the squared perturbation amplitude on a comoving scale k. Note that the above definition of the power spectrum is analogous to the definition (12.14) in the case of hydrodynamical matter.

We now turn to the calculation of the power spectrum in inflationary universe models. The solution of the mode equation (13.6) for short-wavelength perturbation for which $k^2 \gg (1/z)''(1/z)^{-1}$ is

$$u_k(\eta) = u_k(\eta_i)\cos[k(\eta-\eta_i)] + [u_k'(\eta_i)/k]\sin[k(\eta-\eta_i)] . \tag{13.12}$$

For long-wavelength perturbations $[k^2 \ll (1/z)''(1/z)^{-1}]$ the solution of (13.6) is

$$u_k(\eta) = A_k\,\frac{1}{\varphi_0'}\left(\frac{1}{a}\int a^2\,\mathrm{d}\eta\right)' . \tag{13.13}$$

To directly derive (13.13), u_k is first written as

$$u_k = B_k\,\frac{1}{z}\int z^2\,\mathrm{d}\eta\,, \quad B_k = \text{constant}\,. \tag{13.14}$$

It can be verified by direct substitution that (13.14) satisfies the long-wavelength limit of (13.7). Next, the definition of z is inserted in (13.14). Inside the integral, $\varphi_0'^2$ is replaced by a function of a using the background equation $\dot{H} = \frac{3}{2}l^2\dot{\varphi}_0^2$. After integrating the resulting equation by parts, the expression for u_k reduces to (13.13).

If we consider inflationary models which solve the horizon problem, all scales smaller than the Hubble radius at the present time η_0 are inside the Hubble radius at the beginning of the de Sitter phase ($\eta = \eta_i$). Only such scales will be considered in the following. As mentioned at the end of chapter 11, the ambiguities in the choice of the state of the system are unimportant on these scales provided the asymptotic conditions on the functions $M(k\eta_i)$ and $N(k\eta_i)$ – mentioned above (13.9) – are satisfied.

Thus, the perturbations we are interested in are initially inside the Hubble radius and evolve according to (13.12). Later they leave the Hubble radius and subsequently their evolution can be described in (13.13). One way to determine the coefficient A_k in (13.13) is to use the junction conditions between solutions (13.12) and (13.13) at the time of Hubble radius crossing, i.e., we could demand that at that time $u_k(\eta)$ and $u_k'(\eta)$ are continuous. However, there is a more elegant way to join the solutions. From (11.11a), it follows that $v_k(\eta)$ has oscillatory solutions for $k^2 \gg z''/z$. By (13.8), this implies that the solution (13.12) for $u_k(\eta)$ is also valid for $(1/z)''(1/z)^{-1} \gg k^2 \gg z''/z$. Hence, in the time interval for which

$$(1/z)''(1/z)^{-1} \gg k^2 \gg z''/z\ , \tag{13.15}$$

both of the solutions (13.12) and (13.13) for $u_k(\eta)$ are valid if this interval exists.

For an inflationary universe, the interval (13.15) is not empty since in this case (13.15) takes the form

$$\mathcal{H}a(\eta) \gg k \gg V_{,\varphi\varphi}^{1/2}a(\eta)\ . \tag{13.16}$$

During inflation $\mathcal{H} \gg V_{,\varphi\varphi}^{1/2}$, and therefore the above interval does not vanish. Thus, for fixed k, from the time η_1 when $k \sim \mathcal{H}a(\eta)$ until η_2 when $k \sim V_{\varphi\varphi}^{1/2}a(\eta_2)$, both solutions (13.12) and (13.13) are valid. Comparing these solutions during this time interval, it follows that for perturbations with k in the range

$$\mathcal{H}(\eta)a(\eta) > k > H(\eta_i)a(\eta_i)\ , \tag{13.17}$$

the solution for $u_k(\eta)$ can be written as

$$u_k(\eta) = \frac{u_k(\eta_i)\cos(k\eta_i) - [u_k'(\eta_i)/k]\sin(k\eta_i)}{\{(1/\varphi_0')[(1/a)\int a^2\,\mathrm{d}\eta]'\}_{\eta_{\mathcal{H}}(k)}}\ \frac{1}{\varphi_0'}\left(\frac{1}{a}\int a^2\,\mathrm{d}\eta\right)'\ . \tag{13.18}$$

[To see this, evaluate (13.18) at $\eta_{\mathcal{H}}(k)$. The result must agree with (13.12) when taking into account that $|\eta_i| \gg |\eta_{\mathcal{H}}(k)|$.] In the above, the index $\eta_{\mathcal{H}}(k)$ means that the expression inside the bracket is to be evaluated at the time $t_{\mathcal{H}}(k)$ when the scale k crosses the Hubble radius. Note that (13.18) is also true after inflation for scales which have left the Hubble radius during inflation. It is worth noting that the expression inside the square brackets [with index $\eta_{\mathcal{H}}(k)$] in (13.18) changes only insignificantly during the time interval $\eta_1 < \eta < \eta_2$ at the inflation stage; we could take $\eta_{\mathcal{H}}(k)$ for any η for this time interval.

Now, we can combine the above results to determine the power spectrum δ_k, which is given in terms of $u_k(\eta)$ by (13.11). The solutions for $u_k(\eta)$ in the various wavelength intervals are expressed in terms of the initial values in (13.12) and (13.18). The initial values for $u_k(\eta)$ and $u_k'(\eta)$ can be taken from (13.9). We first consider the power spectrum during inflation. On scales which at time t are still inside the Hubble radius $[k_{\rm ph}(t) > H(t)]$, $u_k(\eta)$ is given by (13.12) and the power spectrum is

$$|\delta_k| \simeq (3l^2/4\pi^2)|\dot{\varphi}_0(t)|\,, \quad k_{\rm ph}(t) > H(t)\,, \tag{13.19}$$

where t stands for the physical time $t = \int a\,{\rm d}\eta$ and $k_{\rm ph} = k/a(\eta)$ is the wavenumber in physical coordinates. On scales which were inside the Hubble radius at the beginning of inflation but which are outside at time t, we use (13.18) to obtain

$$|\delta_k| \simeq \frac{3l^2}{4\pi^2}\left(\frac{\dot{\varphi}_0 H^2}{\dot{H}}\right)_{t_{\mathscr{H}}(k)}\left(\frac{1}{a}\int a\,{\rm d}t\right)', \quad H(t) > k_{\rm ph}(t) > H(t_{\rm i})\,\frac{a(t_{\rm i})}{a(t)}\,. \tag{13.20}$$

In deriving (13.19) and (13.20) we took into account that $|M| \to 1$ and $|N| \to 1$ for $k\eta_{\rm i} \gg 1$ in (13.9), used the normalization condition (11.28), and kept only the leading terms in k in the final expressions. Equations (13.19) and (13.20) are valid on the corresponding wavelengths both during and after inflation.

Let us first apply (13.19) and (13.20) to study the time evolution of the spectrum of perturbations during inflation. Using the Friedmann background equation (6.11) to express $\dot{H}$ in terms of $\dot{\varphi}_0^2$ and the equation of motion (6.27) in the slow roll approximation to express $\dot{\varphi}$ in terms of $V_{,\varphi}$ and V, we obtain – for a scalar field with general potential $V(\varphi)$ – the following power spectrum:

$$|\delta_k| \simeq \begin{cases} (l/4\pi)\,V_{,\varphi}/V^{1/2}\,, & k_{\rm ph} > H(t)\,, \\ \dfrac{l}{4\pi}\left(\dfrac{V^{3/2}}{V_{,\varphi}}\right)_{\eta_{\mathscr{H}}(k)}\dfrac{V_{,\varphi}^2}{V^2}\,, & H(t) > k_{\rm ph} > H_{\rm i}\,a(t_{\rm i})/a(t)\,. \end{cases} \tag{13.21}$$

In the case of a potential $V(\varphi) = (\lambda/n)\varphi^n$, eq. (13.21) becomes (when using $a(t) = a(t_{\rm i})\exp\{-(3l^2/2n)[\varphi^2(t) - \varphi^2(t_{\rm i})]\}$)

$$|\delta_k| \simeq \frac{l\lambda^{1/2}n^{1/2}}{4\pi}\left(\frac{2n}{3l^2}\ln\frac{a_{\rm r}}{a(t)}\right)^{n/4-1/2} \times \begin{cases} 1\,, & k_{\rm ph} > H(t)\,, \\ \left(1 + \dfrac{\ln(\lambda_{\rm ph}H)}{\ln[a_{\rm r}/a(t)]}\right)^{n/4+1/2}, & H(t) > k_{\rm ph} > H_{\rm i}\,\dfrac{a(t_{\rm i})}{a(t)}\,. \end{cases} \tag{13.22}$$

These equations are true only during inflation, i.e., for $a(t) < a_{\rm r}$, where $a_{\rm r}$ is the scale factor at the end of inflation. $\lambda_{\rm ph}$ is the wavelength in physical coordinates.

As an example, consider the simplest potential which leads to chaotic inflation, namely $V(\varphi) = \frac{1}{2}m^2\varphi^2$. In this case (13.22) becomes

$$|\delta_k| \simeq \frac{2^{1/2}}{4\pi}\frac{m}{m_{\rm Pl}} \begin{cases} 1\,, & k_{\rm ph} > H(t)\,, \\ \left(1 + \dfrac{\ln[\lambda_{\rm ph}H(t)]}{\ln[a_{\rm r}/a(t)]}\right), & H(t) > k_{\rm ph} > H_{\rm i}\,\dfrac{a(t_{\rm i})}{a(t)}\,. \end{cases} \tag{13.23}$$

The time dependence of the spectrum (13.23) is sketched in fig. 13.1. Note that the k region for which there are logarithmic corrections to a flat spectrum increases in time.

The period of inflation ends at the time t_r when $H^2(t_r) \simeq V_{,\varphi\varphi}$. Let us assume that after inflation the time evolution of the scale factor $a(t)$ is given by $a(t) \sim t^p$, where p is some real number smaller than 1. In the case of a quadratic potential, $p = 2/3$. On scales which are inside the Hubble radius at the beginning of inflation ($t = t_i$) but outside at the end ($t = t_r$), i.e., for which

$$H(t_r)\, a(t_r)/a(t) > k_{ph}(t) > H_i\, a(t_i)/a(t)\,, \tag{13.24}$$

the power spectrum is obtained from (13.20). Thus, for any $t > t_r$

$$|\delta_k| \simeq \frac{3H^3}{4\pi(p+1)} \left(\frac{\dot{\varphi}_0 H^2}{\dot{H}}\right)_{t_{\mathcal{H}}(k)} \simeq \frac{3H^2}{2\pi(p+1)} \left(\frac{V^{3/2}}{V_{,\varphi}}\right)_{t_{\mathcal{H}}(k)}. \tag{13.25}$$

In particular, for the potential $V(\varphi) = (\lambda/n)\varphi^n$

$$|\delta_k| \simeq \frac{l\lambda^{1/2}}{\pi(p+1)n^{1/2}} \left(\frac{2n}{3l^2}\right)^{n/4-1/2} [\ln(\lambda_{ph}/\lambda_\gamma)]^{n/4+1/2}\,, \tag{13.26}$$

where λ_γ is the characteristic wavelength of the cosmic background radiation (see fig. 13.2). From (13.26) it follows that in inflationary universe models the spectrum of adiabatic metric fluctuations is nearly scale-invariant. There is a logarithmic correction factor. The strength of the logarithmic correction depends on the particular model.

It is interesting to consider two special examples and confront the theoretical predictions with observational constraints. As discussed in part III, chapter 17 of this review, the isotropy of the microwave background radiation constrains $|\delta_k|$ to be smaller than about 10^{-5} on large scales. Note that in the following equations $p = 2/3$ is assumed.

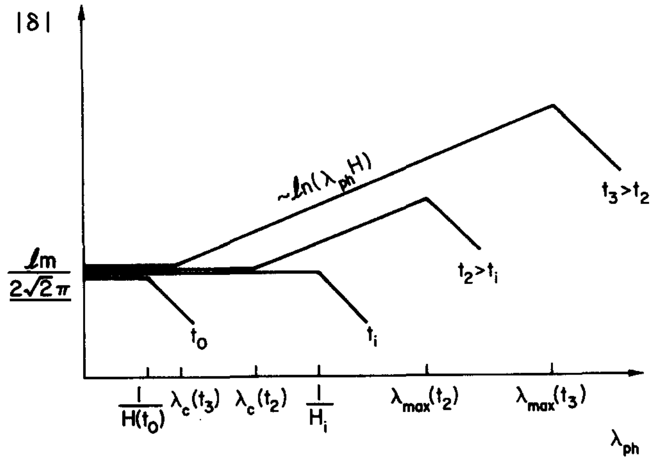

Fig. 13.1. The power spectrum during inflation in the model of chaotic inflation with potential $V(\varphi) = \frac{1}{2}m^2\varphi^2$. t_0 is some time before the onset of inflation at time t_i, and t_2 and t_3 are two fixed times during inflation. The maximal wavelength $\lambda_{max}(t)$ is $H^{-1}a(t)a(t_i)^{-1}$, and $\lambda_c(t)$ is given in terms of the scale factor $a(t_r)$ at the end of inflation by $H^{-1}a(t_r)a(t)^{-1}$.

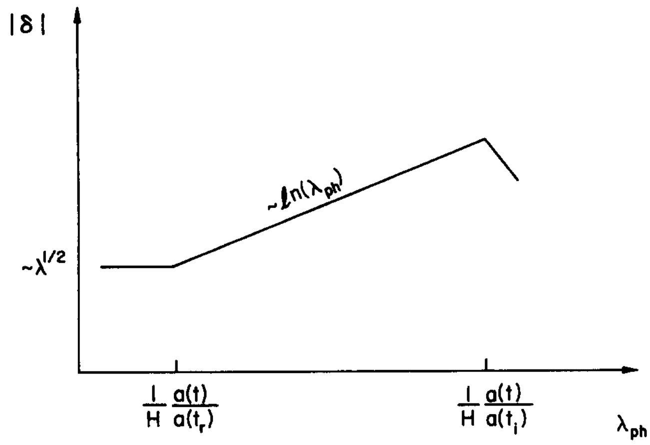

Fig. 13.2. The power spectrum after inflation in the model of chaotic inflation discussed in fig. 13.1. Note the logarithmic deviation from a scale-invariant spectrum.

In a model with quadratic potential $V(\varphi) = \frac{1}{2}m^2\varphi^2$, it follows from (13.26) that

$$|\delta_k| \simeq (3/5\pi\sqrt{2})(m/m_{\mathrm{Pl}}) \ln(\lambda_{\mathrm{ph}}/\lambda_\gamma)\,. \tag{13.27}$$

In very large scales the logarithm is about 10^2. Hence, the observational limits on $|\delta_k|$ yield the constraint

$$m/m_{\mathrm{Pl}} < 10^{-6}\,. \tag{13.28}$$

For a quartic potential $V(\varphi) = \frac{1}{4}\lambda\varphi^4$, the power spectrum is

$$|\delta_k| \simeq (3/5\pi)\tfrac{2}{3}^{1/2}\lambda^{1/2}[\ln(\lambda_{\mathrm{ph}}/\lambda_\gamma)]^{3/2}\,. \tag{13.29}$$

In this example, the observational constraints give

$$\lambda < 10^{-14}\,. \tag{13.30}$$

Thus, in order that the quantum fluctuations from inflation do not produce too large density perturbations, either extremely small values of coupling constants (13.30) or a mass hierarchy (13.28) is required.

In models with a very small coupling constant as in (13.30), it is unnatural to assume that φ was in thermal equilibrium during the very early universe. The above conditions on coupling constants and masses also apply to new inflationary universe models. In this case, however, thermal equilibrium was crucial to justify the initial conditions. Hence, the constraint (13.30) is a major problem for models of new inflation. In contrast, in chaotic inflation φ is not assumed to be in thermal equilibrium in the very early universe, and (13.30) or (13.28) do not lead to an internal inconsistency.

In conclusion, in this chapter a consistent derivation of the power spectrum of metric perturbations caused by vacuum quantum fluctuations of scalar-field matter has been presented. Unlike earlier derivations [53, 52, 118], no artificial splitting between a quantum generation period and classical growth phase has to be assumed. Our method is based on writing the action for metric perturbations in terms of one gauge-invariant variable, and applying standard quantization to that action (see chapter 10). The main results are as follows. Metric perturbations always exist. Their amplitude increases during inflation and due to the exponential expansion, the final power spectrum has an amplitude which is nearly the same over all scales of cosmological relevance. As was shown, the results do not depend on the choice of vacuum state, as long as some basic asymptotic limits are taken on correctly [see above (13.9)].

14. Spectrum of density perturbations in higher-derivative gravity models

As discussed in chapter 7, higher-derivative gravity theories quite naturally lead to inflation. Hence it is of interest to compute the spectrum of metric perturbations in these models [120]. Since we shall consider the quantum theory of cosmological perturbations, the starting point will be the action principle for these perturbations. For a gravity theory given by

$$S = -\frac{1}{6l^2}\int f(R)\sqrt{-g}\,\mathrm{d}^4x\,, \tag{14.1}$$

the action for cosmological perturbations has been obtained in section 10.4 in the case of a spatially flat universe ($\mathcal{K}=0$),

$$\delta_2 S = \frac{1}{2}\int\left(\tilde{v}'^2 - \tilde{v}_{,i}\tilde{v}_{,i} + \frac{z''}{z}\tilde{v}^2\right)\mathrm{d}^4x\,, \tag{14.2}$$

$$z = \frac{1}{\sqrt{2}l}\,\frac{a^2F_0'}{(aF_0^{1/2})'}\,,\qquad \tilde{v} = \frac{z}{R_0'}\,(\mathcal{H}\,\delta R^{(\mathrm{gi})} + R_0'\Psi)\,,\quad F = \frac{\partial f}{\partial R}\,. \tag{14.3a, b, c}$$

The variables z and $\tilde{v}$ have been expressed in terms of original (nonconformal) variables exclusively. Note that the index 0 stands for background model quantities.

The quantization of the system with action (14.2) has been considered in detail in chapter 11. The operator $\hat{\tilde{v}}$ corresponding to $\tilde{v}$ can be expanded in the usual way in terms of creation and annihilation operators $\hat{a}_k^+$ and $\hat{a}_k^-$,

$$\hat{\tilde{v}}(\eta, \boldsymbol{x}) = \frac{1}{\sqrt{2}}\int\frac{\mathrm{d}^3k}{(2\pi)^{3/2}}\,[\tilde{v}_k^*(\eta)\,\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\hat{a}_k^- + \tilde{v}_k(\eta)\,\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\hat{a}_k^+]\,, \tag{14.4}$$

where the temporal modes satisfy the equation

$$\tilde{v}_k''(\eta) + (k^2 - z''/z)\tilde{v}_k(\eta) = 0\,, \tag{14.5}$$

with initial conditions

$$\tilde{v}_k(\eta_\mathrm{i}) = k^{-1/2}M(k\eta_\mathrm{i})\,,\qquad \tilde{v}_k'(\eta_\mathrm{i}) = \mathrm{i}k^{1/2}N(k\eta_\mathrm{i})\,, \tag{14.6}$$

at the initial time η_i. The functions M and N have the asymptotic limits $|M(k\eta_\mathrm{i})|\to 1$ and $|N(k\eta_\mathrm{i})|\to 1$ when $k\eta_\mathrm{i} \gg 1$. They also satisfy the normalization condition (11.28).

Since the observables are correlation functions of the operators $\hat{\Phi}$ and $\hat{\Psi}$, it is necessary to first find the relations between $\hat{\Phi}$, $\hat{\Psi}$ and $\hat{v}$. As was shown in chapter 7, the gravity theory given by (14.1) is conformally equivalent to an Einstein theory for the conformal metric, $\tilde{g}_{\alpha\beta} = F(R) g_{\alpha\beta}$, with scalar-field matter, the scalar field φ being related to the Ricci scalar curvature R of the original metric via, $\varphi = (1/\sqrt{2}\, l) \ln F(R)$. Hence, we can use (13.7) to express the conformal potentials $\tilde{\Phi}$ and $\tilde{\Psi}$ in terms of $\tilde{v}$,

$$\Delta \tilde{\Phi} = \frac{3}{2\sqrt{2}}\, l\, \frac{(\ln F)'}{F^{1/2} a}\, z\left(\frac{\tilde{v}}{z}\right)' , \tag{14.7}$$

where z was defined in (14.3a). On the other hand, the relation between the conformal potentials $\tilde{\Phi} = \tilde{\Psi}$ and the original potentials Φ and Ψ was derived in chapter 7 [see (7.42)],

$$\Phi = -\tfrac{2}{3}(F^2/F'a)[(a/F)\tilde{\Phi}]' , \qquad \Psi = \tfrac{2}{3}(1/FF'a)(aF^2\tilde{\Phi})' . \tag{14.8}$$

Thus, from (14.7) and (14.8) we can immediately express Φ and Ψ in terms of $\tilde{v}$. The same relations are valid in the quantum theory for the corresponding operators.

To calculate the perturbation spectrum, it is necessary to evaluate the correlation function of the potentials Φ or Ψ. The operator $\hat{\Phi}$ can be expanded in terms of the creation and annihilation operators defined in (14.4),

$$\hat{\Phi}(\eta, \boldsymbol{x}) = \frac{1}{\sqrt{2}} \int \frac{d^3k}{(2\pi)^{3/2}} \left[\Phi_k^*(\eta)\, e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \hat{a}_k^- + \Phi_k(\eta)\, e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \hat{a}_k^+\right] , \tag{14.9}$$

where the mode functions $\Phi_k(\eta)$ are related to the functions $v_k(\eta)$ appearing in (14.4) by the identities (14.7) and (14.8). The time dependence of $\Phi_k(\eta)$ was studied in chapter 7 [see eqs. (7.51) and (7.52)],

$$\Phi_k(\eta) = -\frac{1}{F^{1/2}} \left[\left(\frac{\ddot{F}}{\dot{F}} - \frac{5}{2}\frac{\dot{F}}{F} + H + \frac{ik}{a}\right) c_k \exp\left(ik \int \frac{dt}{a}\right) + \text{c.c.}\right] \tag{14.10}$$

(where c.c. stands for complex conjugate) for short-wavelength perturbations with $k^2 \gg (1/z)''(1/z)^{-1}$ and

$$\Phi_k(\eta) \simeq A_k \left(\frac{1}{aF} \int aF\, dt\right)^{\cdot} , \tag{14.11}$$

for long-wavelength perturbations with $k^2 \ll (1/z)''(1/z)^{-1}$.

To fix the coefficients c_k and A_k in these equations, we use (14.7) and (14.8) to relate the initial values of Φ_k and Φ_k' to the values of $\tilde{v}$ and $\tilde{v}'$ at the initial time η_i, which in turn are given by (14.6). In inflationary universe models, scales of interest to cosmology are far inside the Hubble radius at the beginning of inflation. Hence, the relevant wavenumbers k satisfy $k\eta_i \gg 1$. Thus, from (14.6)–(14.8) and keeping only the leading terms in an expansion in powers of k^{-1} we obtain

$$\Phi_k(\eta_i) \simeq \frac{1}{\sqrt{2}} \frac{l}{a(\eta_i) F_0^{1/2}(\eta_i)} \frac{1}{k^{1/2}} + O(k\eta_i) , \qquad \Phi_k'(\eta_i) \simeq -\frac{1}{\sqrt{2}} \frac{il}{a(n_i) F_0^{1/2}(\eta_i)} k^{1/2} + O(k\eta_i) , \tag{14.12}$$

where $O(k\eta_i)$ denotes terms which can be neglected when $k\eta_i \gg 1$.

In order to make the following considerations specific, the above formalism will now be applied to the inflationary model in R^2 gravity. The action for this theory is

$$S = -\frac{1}{6l^2}\int\left(R - \frac{R^2}{6M^2}\right)\sqrt{-g}\,d^4x\,. \tag{14.13}$$

As discussed in section 7.2, a period of inflation rises under quite natural initial conditions in this model. We are interested in perturbations which were inside the Hubble radius at the beginning of inflation. From (14.10) and (14.12) it follows that on scales which satisfy the condition

$$k \gg (1/z)''(1/z)^{-1} \tag{14.14}$$

for the applicability of the short-wavelength approximation, the solutions for $\Phi_k(\eta)$ are

$$\Phi_k(\eta) = \frac{l}{\sqrt{2}F^{1/2}}\left(\frac{\ddot{F}}{F} - \frac{5}{2}\frac{\dot{F}}{F} + H + \frac{ik}{a}\right)\left(\frac{i}{k^{3/2}} + O(k\eta)\right)\exp[ik(\eta - \eta_i)]\,. \tag{14.15}$$

In the inflationary phase of the R^2-gravity theory, eq. (14.14) takes the form $k > Ma(\eta)$.

Once the scale k comes to satisfy the condition $k \ll Ma(\eta)$ during the inflationary period, then $\Phi_k(\eta)$ evolves according to (14.11). To fix the coefficient A_k in (14.11) for values of k which correspond to scales which were initially inside the Hubble radius the solutions (14.11) and (14.15) must be joined. The method is analogous to the consideration in chapter 13 for scalar-field perturbations. For fixed k, both solutions (14.11) and (14.15) must be valid in the time interval when $H(\eta)a(\eta) \gg k \gg Ma(\eta)$ (during inflation). Comparing these solutions in the time interval when both are valid and taking into account that during inflation $F \simeq 4H^2/M^2$ one obtains

$$\Phi_k(\eta) \simeq \frac{3l}{\sqrt{2}}\frac{i}{k^{3/2}}\,e^{-ik\eta}\,\frac{H^2_{t_M(k)}}{M}\left(\frac{1}{aF}\int aF\,dt\right)' \tag{14.16}$$

for perturbations on scales with $Ma(\eta) \gg k \gg H(\eta_i)a(\eta_i)$. Note that in (14.16), the time when $k = Ma(\eta)$ is denoted by $t_M(k)$. Equation (14.16) is valid also after inflation, but only for scales which have crossed the Hubble radius before the end of inflation.

The power spectrum $|\delta_{\Phi,k}|$ of metric perturbations is defined as in chapter 13 [see (13.10)],

$$\langle 0|\hat{\Phi}(\boldsymbol{x},\eta)\hat{\Phi}(\boldsymbol{x}+\boldsymbol{r}_1\eta)|0\rangle = \int_0^\infty \frac{dk}{k}\frac{\sin(kr)}{kr}|\delta_{\Phi_k}|^2\,, \tag{14.17}$$

and is related to the mode functions $\Phi_k(\eta)$ via

$$|\delta_{\Phi_k}(\eta)|^2 = (1/4\pi^2)|\Phi_k(\eta)|^2k^3\,. \tag{14.18}$$

Since we have already determined $\Phi_k(\eta)$, the power spectrum can be calculated in a straightforward manner by evaluating (14.16) using the equations of chapter 7 for the background model. During inflation, the result is

$$|\delta_{\Phi}(\lambda_{\rm ph}, t)| = (1/4\pi\sqrt{2})Ml \begin{cases} 1/\lambda_{\rm ph}M\,, & \lambda_{\rm ph} \ll 1/H(t)\,, \\ 1 + \dfrac{\ln(\lambda_{\rm ph}M)}{\ln[a(t_{\rm r})/a(t)]}\,, & \dfrac{1}{H(t)} \ll \lambda_{\rm ph} \ll \dfrac{1}{H(t_{\rm i})}\dfrac{a(t)}{a(t_{\rm i})}\,. \end{cases} \tag{14.19}$$

The answer has been expressed in terms of physical time t and physical wavelength $\lambda_{\rm ph} = a(t)/k$. As before, $t_{\rm r}$ is the time corresponding to the end of inflation.

The spectrum (14.19) coincides up to a numerical coefficient with the spectrum in a scalar-field model with potential $v = \frac{1}{2}M^2\varphi^2$, but only for the long-wavelength perturbations $\lambda_{\rm ph} > 1/H(t)$. For short wavelengths the amplitude of the power spectrum in R^2-gravity increases as $\lambda_{\rm ph}^{-1}$, while it is constant in the case of scalar-field perturbations.

To find the power spectrum after inflation, we use eq. (7.56) (which states that up to oscillating terms, Φ is constant for long-wavelength perturbations), and eq. (14.16) to define the amplitude in eq. (7.56). Hence, the following spectrum results after inflation:

$$|\delta_{\Phi}(\lambda_{\rm ph}, t)| \simeq (3/10\pi\sqrt{2})Ml\,\ln(\lambda_{\rm ph}/\lambda_\gamma)\,, \quad \lambda_\gamma = (1/M)a(t)/a(t_{\rm r})\,, \tag{14.20}$$

where λ_γ is the characteristic wavelength of the background radiation. Equation (14.20) is valid on scales satisfying

$$\lambda_\gamma < \lambda_{\rm ph} < [1/H(t_{\rm i})]a(t)/a(t_{\rm i})\,, \tag{14.21}$$

i.e., scales which are inside the Hubble radius at the beginning of inflation but have crossed by the end.

The observational constraints from the absence of detected microwave background radiation anisotropies give an upper bound $|\delta_{\Phi}(\lambda_{\rm ph})| < 10^{-4}$ on cosmological distance scales. This implies a bound $M < 10^{14}$ GeV. For M in the range 10^{13} GeV–10^{14} GeV, the power spectrum has the appropriate magnitude for galaxy formation.

15. Conclusions

In part II of this review article we studied the quantum theory of cosmological perturbations. This enabled us to consider the main problem which cannot be addressed in a classical theory of perturbations, namely the origin of fluctuations. The spectrum of density inhomogeneities originating in quantum fluctuations was calculated in three classes of cosmological models: Friedman cosmologies with hydrodynamical matter, inflationary universe models in which the exponential expansion is driven by a scalar field, and higher-derivative gravity models.

The calculation of the spectrum of density perturbations can be reduced to the evaluation of the two-point function of either the gauge-invariant metric potential Φ or of the gauge-invariant density perturbation $\delta\varepsilon^{(\rm gi)}/\varepsilon$ for the initial quantum state $|\psi\rangle$ of the system. In models with inflation, it was shown that the spectra depend only weakly on the state $|\psi\rangle$ if the length scales under consideration are smaller than the Hubble radius at the time when initial conditions are imposed. We concluded that quantum fluctuations in universes with hydrodynamical matter are insufficient for galaxy formation. However, in inflationary universe models the amplitude of density perturbations at the present time can be sufficiently large for galaxy formation. In fact, the danger is that they might be too large without fine tuning the particle physics model. It was shown that in inflationary universe models the spectrum is in general nearly scale-invariant.

The main advantages of the formalism presented in this article are gauge invariance and physical consistency. In addition, the treatment gives a unified picture of the generation and evolution of perturbations. The first step in the analysis is the reduction of the action to that of a single gauge-invariant variable. This reduced action is then quantized according to the canonical quantization prescription. This allows for the evaluation of the two-point correlation function of physical, gauge-invariant operators. These operators can be expanded in creation and annihilation operations. The coefficient functions are shown to satisfy the same classical equations as the classical perturbation variables of part I. This establishes a deep connection between the first two parts of this review article.

PART III. EXTENSIONS

16. Introduction

In this part we demonstrate the power of the gauge-invariant formalism by considering several important physical applications of cosmological perturbation theory. Specifically, we consider microwave background anisotropies, gravitational waves, entropy perturbations and statistical fluctuations. All of these topics address issues which directly relate primordial perturbations to observable quantities, hence providing constraints on models of the early universe. A gauge-invariant approach simplifies and clarifies the analysis in all four cases (the analysis of gravitational waves is automatically gauge-invariant).

Chapter 17 is devoted to a study of anisotropies in the cosmic microwave background radiation (CBR). At the present time, the most stringent constraints on the large wavelength power spectra in models of structure formation come from the absence of detected CBR anisotropies (except for the dipole component which is due to the motion of the solar system) [130]. In our analysis, the CBR temperature anisotropies are directly related to the gauge-invariant potential for metric perturbations.

In chapter 18, the theory of generation and evolution of metric perturbations is extended to the tensor perturbations which describe gravitational waves. In principle, gravitational waves provide a way to discriminate between various theories of formation of structure in the universe, since different models lead to different spectra of gravitational radiation. Because of the weak coupling between gravitational waves and matter, information about the earliest stages of the evolution of the universe is stored in the gravitational wave spectrum. We consider the generation of gravitational waves in an inflationary universe. The quantum fluctuations during the period of inflation lead to a late time spectrum of gravitational waves which has a flat logarithmic energy-density spectrum over a wide range of wavelengths. In addition, we extend the analysis to demonstrate that interesting specific features in the spectrum of gravitational waves in double-inflation models can be produced.

In parts I and II of this review article – especially in the chapters on scalar-field perturbations – most of the emphasis was on adiabatic perturbations. However, in models with several components of matter, e.g., in a model with several scalar fields, entropy perturbations are generic. In the simplest examples, entropy perturbations in inflationary universe models also lead to a scale-invariant Harrison–Zel'dovich spectrum. Recent observations of large-scale structure [129] indicate that such a spectrum might not provide sufficient power on large scales. It is fairly easy to construct models with non-scale-invariant spectra for entropy perturbations. In chapter 19 we use a simple model of two interacting scalar fields to illustrate the different features in the spectrum which can be obtained. However, it is worth noting that all of these models are quite unrealistic since they demand fine tuning of the parameters in order that the specific features in the spectrum appear on cosmologically interesting scales.

In chapter 20, we apply the quantum theory developed in part II to give a physically consistent and gauge-invariant definition of statistical fluctuations in models with hydrodynamical matter. These perturbations are shown to be insufficient for galaxy formation in usual scenarios without inflation.

17. Microwave background anisotropies

The observational upper bounds on temperature anisotropies in the cosmic microwave background radiation (CBR) provide strong constraints on the spectrum of metric perturbations. In fact, the approximate isotropy of the CBR yields strong evidence that the early universe can be considered to be approximately homogeneous on large scales.

Photons travel on geodesies. Therefore, as first realized by Sachs and Wolfe [2], metric perturbations will induce anisotropies in the temperature of the CBR. The basic idea is illustrated in fig. 17.1. There are three contributions to anisotropies of the CBR. Density fluctuations at the time of last scattering lead to differences in the length of the geodesics on which photons reach the observer at the present time t_0 from different directions. Secondly, density perturbations between the time $t_{\rm rec}$ of last scattering and t_0 lead to deviations of the geodesics and hence to temperature fluctuations. Finally, any peculiar velocity of the observer or of the atoms emitting photons at last scattering gives rise to Doppler temperature perturbations. Apart from the dipole contribution [130], which is believed to be due to our peculiar velocity relative to the rest frame of the CBR, no anisotropies in the microwave background have been observed at a sensitivity of 10^{-4} or better. For detailed reviews of the observational status see, e.g., refs. [131, 1]. The upper limits on temperature fluctuations lead to interesting bounds on the amplitude of the spectrum of density perturbations. In fig. 17.1, the last scattering surface is idealized as being infinitesimally thin. This is only a good approximation on angular scales corresponding to a distance at last scattering larger than the Hubble radius at that time. In the following, we shall only consider such large angular scales. We also neglect Doppler type anisotropies.

There has been a lot of theoretical work on microwave background anisotropies following the pioneering paper by Sachs and Wolfe [2]. Some early papers are listed in refs. [132–134]. Peebles [135]

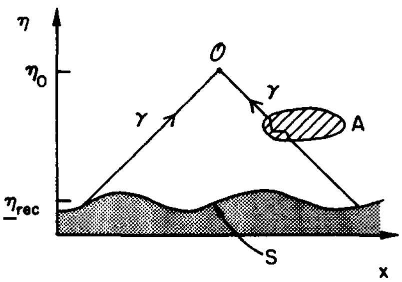

Fig. 17.1. Conformal space–time diagram showing two sources of fluctuations in the CBR: perturbations of the last scattering surface S and time delays due to inhomogeneities (like region A) along the geodesics γ reaching the observer O from different directions.

has performed a detailed analysis of CBR fluctuations, including small scale anisotropies, which were also considered in ref. [136]. More recent analyses are in refs. [137, 138], where the causality constraints on large-scale CBR anisotropies were studied. The first studies of CBR anisotropies produced by density perturbations in inflationary universe models were given in ref. [139]. Bond and Efstathiou [140] have done detailed analytical and numerical analyses of the CBR fluctuations in various cosmological models based on inflation (for a recent view see, e.g., ref. [141]).

Most of the analyses in the literature use synchronous gauge and are very complicated. Here, we will give an analysis of the large-scale CBR anisotropies using the gauge-invariant framework (based on ref. [142]). A similar formalism was developed in ref. [143]. The observed anisotropy will be directly connected to gauge-invariant characteristics of the perturbation. A different gauge-invariant approach to CBR fluctuations has also been put forward in ref. [144].

In this chapter the gauge-invariant framework will be applied to derive the relation between the relative temperature fluctuation $\delta T/T$ and the metric potential Φ. In particular, this general relation will be applied to the quadrupole temperature anisotropy. Since last scattering occurs after the time of equal matter and radiation, the equations for metric perturbations as specialized to a matter-dominated universe may be applied. We shall again only consider a spatially flat metric ($\mathcal{K}=0$). From (5.19), it follows that the equation of motion for the potential Φ is

$$\Phi'' + (6/\eta)\Phi' = 0\,. \tag{17.1}$$

The general solution of (17.1) is

$$\Phi(\boldsymbol{x},\eta) = A(\boldsymbol{x}) + B(\boldsymbol{x})\eta^{-5}\,. \tag{17.2}$$

The B mode decays rapidly and can hence be neglected.

From the (0, 0) and (0, i) Einstein equations, the gauge-invariant energy-density perturbation $\delta\varepsilon/\varepsilon$ and the gauge-invariant peculiar velocity v^i can be expressed in terms of the potential Φ [see (5.28) and (5.29)],

$$\delta\varepsilon/\varepsilon = (\eta^2/6)\Delta\Phi - 2\Phi\,, \qquad v^i = -\tfrac{1}{3}\eta\delta^{ij}\,\partial\Phi/\partial x^j\,, \quad v^i = a\,\delta u^i\,, \tag{17.3}$$

where we omitted the superscript (gi). Until last scattering, radiation is coupled to matter. Since the energy density in radiation scales as $\varepsilon_{\rm r} \sim a^{-4}$ while the energy density in matter obeys $\varepsilon_{\rm m} \sim a^{-3}$, we have

$$\delta\varepsilon_{\rm r}/\varepsilon_{\rm r} = \tfrac{4}{3}\,\delta\varepsilon/\varepsilon \tag{17.4}$$

(using $\varepsilon_{\rm r} \ll \varepsilon_{\rm m}$). Also, since $p_{\rm r} = \frac{1}{3}\varepsilon_{\rm r}$, it follows that

$$(T^i_{\ 0})_{\rm r} = \tfrac{4}{3}\varepsilon_{\rm r}v^i \tag{17.5}$$

(where a subscript r indicates "radiation"). Equations (17.4) and (17.5) specify the initial conditions for radiation at the time of last scattering.

To derive the relation between $\delta T/T$ and Φ, we will work in longitudinal gauge in which the line element is

$$ds^2 = a^2(\eta)[(1+2\phi)\,d\eta^2 - (1-2\phi)\delta_{ij}\,dx^i\,dx^j]\,, \tag{17.6}$$

with ϕ equal to the gauge invariant potential Φ (see chapter 3). The equation of motion for photons is the geodesic equation. If p_α stands for the four-momentum of the photon, then the geodesic equation can be written as

$$\mathrm{d}p_\alpha/\mathrm{d}\eta = 2p\, \partial\Phi/\partial x^\alpha \;, \qquad \mathrm{d}x^i/\mathrm{d}\eta = l^i(1+2\Phi) \;, \tag{17.7}$$

where by definition $l^i = -(1/p)p_i$ and $p^2 = p_i p_i$. These equations are obtained by evaluating the Christoffel symbols in the metric (17.6).

The temperature enters the system of equations through the phase-space distribution function $f(x^\alpha, p_i)$ for the ensemble of photons;

$$\mathrm{d}N = f(x^\alpha, p_i)\, \mathrm{d}^3x^k\, \mathrm{d}^3p_k \tag{17.8}$$

gives the number of photons at time x^0 in the infinitesimal phase-space volume $\mathrm{d}^3x^k\, \mathrm{d}^3p_k$ about the point (x^j, p_i). After last scattering, f obeys the collisionless Boltzmann equation

$$(\mathrm{d}x^\alpha/\mathrm{d}\eta)\, \partial f/\partial x^\alpha + (\mathrm{d}p_i/\mathrm{d}\eta)\, \partial f/\partial p_i = 0 \;. \tag{17.9}$$

Small fluctuations in the microwave background may conveniently be described by perturbations of the temperature parameter T in the unperturbed Planck distribution $\bar{f}$

$$f(x^\alpha, p_i) = \bar{f}(p/[\bar{T} + \delta T(x^\alpha, l^i)]) \;. \tag{17.10}$$

We stress that $\bar{f}$ is a function of a single variable v which in an unperturbed universe would simply be $v = p/T$. In a perturbed metric, f depends on x^α and l^i via the dependence of T on these variables. Note that δT is independent of p since the equations of motion (17.7) are homogeneous in p.

Inserting (17.9) and (17.10) and making use of (17.7), we obtain

$$(\partial_\eta + l^i\, \partial_i)\, \delta T/T = -2l^i\, \partial_i \Phi \;. \tag{17.11}$$

Since in the matter-dominated period $\partial_\eta \Phi = 0$ for the nondecaying mode of the metric perturbation [see (17.2)], eq. (17.11) can be rewritten as

$$(\partial_\eta + l^i\, \partial_i)(\delta T/T + 2\Phi) = 0 \;. \tag{17.12}$$

However, the operator $(\partial_\eta + l^i\, \partial_i)$ is just the total time derivative along the world lines of the photons. Hence, we find that along a null geodesic $\boldsymbol{x}(\eta)$

$$(\delta T/T + 2\Phi)(\boldsymbol{x}, \eta) = \text{const} \;. \tag{17.13}$$

As the final step in the derivation, the initial conditions (17.3)–(17.5) at the time of last scattering are used to determine the constant in (17.13). Radiation with distribution function f will have [134] the stress-energy tensor

$$(T^\alpha_\beta)_\mathrm{r} = \frac{1}{\sqrt{-g}} \int \mathrm{d}^3p\, f\, \frac{p^\alpha p_\beta}{p} \;. \tag{17.14}$$

Hence,

$$\frac{1}{4}\frac{\delta\varepsilon_{\rm r}}{\varepsilon_{\rm r}} = \Phi + \int \frac{{\rm d}^2 l}{4\pi}\frac{\delta T}{T}\,, \qquad \frac{(\delta T^i_0)_{\rm r}}{4\varepsilon_{\rm r}} = \int \frac{{\rm d}^2 l}{4\pi}\, l^i\, \frac{\delta T}{T}\,. \tag{17.15a, b}$$

To obtain (17.15a) from (17.14), the null vector condition $p^\alpha p_\alpha = 0$ is used to express $p^0 p_0$ in terms of p^2. The square root of g, the determinant of the metric tensor, is also expanded to first order in the perturbation variable Φ. To perform the integration, it is useful to introduce $z = p/T$ as a new variable. We separate the integration $d^3 p$ into a radial integral dp and an angular integral. Keeping only terms of first order in the perturbation variables Φ and δT, subtracting $\varepsilon_{\rm r}$ and dividing by $\varepsilon_{\rm r}$ gives (17.15a). The analysis for (17.15b) is similar.

Comparing (17.15a, b) with (17.3)–(17.5) we find that before last scattering

$$\delta T/T = -\tfrac{5}{3}\Phi + \tfrac{1}{18}\eta^2 \Delta\Phi - \tfrac{1}{3}\eta l^i\,\partial_i \Phi\,. \tag{17.16}$$

Thus, using (17.16) as the initial conditions in (17.13), we can rewrite (17.13) in the following final form:

$$(\delta T/T + 2\Phi)(\boldsymbol{x}_0, \eta_0) = (\tfrac{1}{3}\Phi + \tfrac{1}{18}\eta^2\Delta\Phi - \tfrac{1}{3}\eta l^i\,\partial_i\Phi)(\boldsymbol{x}_0 - \boldsymbol{l}(\eta_0 - \eta_{\rm rec}), \eta_{\rm rec})\,, \tag{17.17}$$

where η_0 is the present time, and $\eta_{\rm rec}$ the time of recombination (last scattering).

Let us briefly discuss the physical meaning of the result (17.13) and (17.17). From (17.13) it follows that

$$(\delta T/T)(\eta_0) = (\delta T/T)(\eta_{\rm rec}) - 2[\Phi(\eta_0) - \Phi(\eta_{\rm rec})]\,, \tag{17.18}$$

where each term is evaluated along the light ray $\boldsymbol{x}(\eta)$ at the corresponding time. Therefore, if the last scattering surface is unperturbed, only the second term on the right-hand side of (17.18) would remain. This term is the line-of-sight contribution to $\delta T/T$. In the case under consideration, the last scattering surface is perturbed and (17.17) can be rewritten [using (17.3)] as

$$(\delta T/T)(\eta_0) + 2\Phi(\eta_0) = \tfrac{1}{3}\Phi(\eta_{\rm rec}) + l^i v_i(\eta_{\rm rec}) - \tfrac{1}{18}\eta^2_{\rm rec}\,\Delta\Phi(\eta_{\rm rec})\,, \tag{17.19}$$

where all terms are evaluated on the same null geodesic of the unperturbed metric. On scales larger than the Hubble radius at the time of recombination the last term on the right-hand side of (17.19) is negligible. The first term is the Sachs–Wolfe result which includes both the energy-density perturbation on the last scattering surface and the line-of-sight distortions. The second term is the Doppler contribution which is nonvanishing if the emitters have a nonzero peculiar velocity.

We now consider two light rays approaching the observer with an angle of separation θ. The relative temperature difference $(\Delta T/T)(\theta)$ between the two photons is obtained by taking the difference of $\delta T/T(\eta_0)$ corresponding to the two photon paths. (Note that ΔT stands for the temperature difference between two specific light rays, whereas δT denotes the deviation of the temperature from its background value.) On large angular scales,

$$(\Delta T/T)(\theta) \simeq \tfrac{1}{3}\Delta\Phi(\theta, \eta_{\rm rec}) + \Delta(l^i v_i)(\eta_{\rm rec})\,. \tag{17.20}$$

It is important to keep in mind the conditions under which (17.17) and (17.20) are valid. It has been assumed that metric perturbations are linear, and that the decaying mode can be neglected. Furthermore, the finite thickness of the last scattering surface has not been taken into account. Therefore, (17.17) and (17.20) are valid only on angular scales larger than about 1°, corresponding to the Hubble radius at last scattering. In addition, possible reionization of the universe after η_{rec} has been neglected. If the universe is reionized between times η_1 and $\eta_2 > \eta_1$, then (17.17) will hold only on scales larger than the Hubble radius at η_2.

As an example of the application of (17.19), the quadrupole anisotropy $\delta_2 T/T$ of the CBR becomes (in the absence of a quadrupole component to the velocity perturbation on the last scattering surface)

$$\delta_2 T/T = \tfrac{1}{3}\delta_2 \Phi \,, \qquad (17.21)$$

where $\delta_2 \Phi$ is the quadrupole moment of the potential Φ. Recent observations [145] constrain the magnitude of $\delta_2 T/T$ to be smaller than 3×10^{-5}.

The potential Φ on a scale k is proportional to the magnitude of the energy-density perturbation on that scale at the time when the corresponding wavelength crosses the Hubble radius. Thus, on an angular scale θ, the temperature fluctuation ΔT is proportional to the amplitude of $\delta\varepsilon/\varepsilon$ on the scale k which dominates the contribution to $\Delta T(\theta)$ evaluated at the time $t_{\mathscr{H}}(k)$ when k enters the Hubble radius, and the observational limits on temperature fluctuations translate into constraints on the amplitude of the primordial density perturbations. Many low Ω models are already ruled out [146, 140, 176] by recent observational results. One way to circumvent the tight constraints is to have a recent phase of loitering [147], a phase during which the scale factor of the universe is essentially constant, which leads to an exponential growth of perturbations.

18. Gravitational waves

Linear tensor perturbations of the metric, gravitational waves, do not couple to energy and pressure and hence do not contribute to gravitational instability. Nevertheless, gravitational waves are of interest in their own right as a specific signature of metric theories of gravity. The classical theory of the evolution of gravitational waves in expanding space–time backgrounds was first analyzed in some of the basic articles on cosmological perturbations (see, e.g., refs. [3, 5]). It was shown that in synchronous gauge the growing modes which exist for scalar-type metric perturbations are absent for gravitational radiation.

The quantum theory of gravitational waves was first considered by Grishchuk [148] who introduced the term "parametric amplification" to describe the evolution of tensor-type metric perturbations. It was found that the expectation values of operators which measure the amplitude of gravitational radiation have a nontrivial time dependence. A state which initially has vacuum-state occupation numbers, at later times corresponds to a squeezed state with occupation numbers which differ from those of the later-time vacuum state [149–151]. Thus, in an expanding universe gravitons can be produced.

As we shall see, the evolution of gravitational wave perturbations is closely related to the evolution of minimally coupled scalar fields on the background space–time. As seen in chapter 13, scalar-field perturbations are generated in inflationary universe models from initial quantum fluctuations. In analogy, we expect that gravitational waves will be generated during a period of inflation from quantum fluctuations. The detailed process will be analyzed in this chapter. In the context of inflationary universe models, the first computations of the spectrum of gravitational waves were presented in refs. [152–155]. For later work see, e.g., refs. [144, 156–160].

The outline of this chapter is as follows: first, the connection between the quantization of gravitational waves and of a scalar field in de Sitter universe is shown. Next, the methods used in chapter 13 to quantize a scalar field are applied to derive a general expression for the tensor operator corresponding to gravitational waves. The third major topic is a discussion of the observables which measure the strength of gravitational radiation. The fourth issue is the evaluation of the spectrum of gravitational waves in de Sitter space. We then proceed to study the spectrum at late times in an inflationary universe. The chapter concludes with an analysis [161] of special signatures in the gravitational wave spectrum in models of double inflation [162].

18.1. Quantization

We shall consider the generation of gravitational waves in inflationary universe scenarios. Since inflation can easily be realized in higher-derivative gravity theories, we will for the sake of generality consider a theory given by the action

$$S = -\frac{1}{16\pi G}\int f(R)\sqrt{-g}\,\mathrm{d}^4x\,. \tag{18.1}$$

As discussed in chapter 7, the theory defined by the above action is conformally equivalent to a pure Einstein theory with scalar-field matter. In linear theory, the gravitational waves decouple from the matter fields. Thus, the only role of this matter in the problem at hand is to define the background model and to determine the relation between the conformal metrics. To simplify the discussion we will assume an exact de Sitter model for the inflationary period. As shown, e.g., in ref. [156], the analysis and results in models of generalized inflation are quite similar.

The background metric is given by

$$\mathrm{d}s^2 = a^2(\eta)(\mathrm{d}\eta^2 - \gamma_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j)\,. \tag{18.2}$$

As mentioned in chapter 2, gravitational wave perturbations are given by a modified line element $\mathrm{d}s^2 + \delta\,\mathrm{d}s^2$ with

$$\delta\,\mathrm{d}s^2 = -a^2(\eta)h_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j\,. \tag{18.3}$$

Here, h_{ij} is a symmetric, traceless and divergenceless three-tensor, i.e., it obeys the constraints

$$h^i_i = h^{|j}_{ij} = 0\,, \tag{18.4}$$

where the notation of chapter 2 has been used. In particular, the indices are raised and lowered using the metric γ_{ij} and its inverse.

Tensor metric perturbations (18.3) are gauge-invariant. They are also invariant under a conformal transformation of type (7.4),

$$\tilde{h}_{ij} = h_{ij}\,. \tag{18.5}$$

Conformal invariance can be used to give an easy derivation of the action for tensor perturbations in models with action (18.1). In the Einstein frame, the action is expanded to second order in perturbation variables. At the end, for the background functions in the Einstein frame their expressions in terms of functions in the original frame can be substituted. We obtain the following second-order action for tensor perturbations:

$$\delta_2 S = \frac{1}{24l^2} \int \tilde{a}^2 (h^{i\prime}_k h^{k\prime}_i - h^i_{k|l} h^{k|l}_i - 2\mathcal{K} h^i_k h^k_i) \sqrt{\gamma}\, \mathrm{d}^4 x \,, \tag{18.6}$$

where $l^2 \equiv 8\pi G/3$, $\tilde{a} = F^{1/2} a(\eta)$ and $F = \partial f/\partial R$. In deriving (18.6), the conformal invariance (18.5) and the identities (18.4) have been used. By varying the action (18.6) with respect to h^i_j we obtain the following equation of motion:

$$h^{i\prime\prime}_j + 2(\tilde{a}'/\tilde{a}) h^{i\prime}_j - \Delta h^i_j + 2\mathcal{K} h^i_j = 0 \,. \tag{18.7}$$

Equation (18.7) is applicable to describe the evolution of gravitational waves at any stage of the evolution of the universe, even in higher-derivative theories of gravity. We will be interested in the generation of gravitational waves during a period of inflation. Thus, we shall focus on the quantum theory of gravitons in a de Sitter phase of a universe with space curvature $\mathcal{K} = 0$. This background model evolves according to

$$a(\eta) = -1/H\eta \,, \qquad H = a'/a^2 = \text{const.} \,, \tag{18.8}$$

where η ranges from $-\infty$ to 0.

To study the evolution of gravitational waves in a de Sitter background, it is convenient to make the following change of variables in (18.7):

$$h^i_j = (1/\tilde{a}^2)[\tilde{a}\mathcal{D}^i_j(\boldsymbol{x}, \eta)]' \,, \tag{18.9}$$

where the tensor function $\mathcal{D}^i_j$ is traceless and divergenceless,

$$\mathcal{D}^i_i = \mathcal{D}^i_{j,i} = 0 \,. \tag{18.10}$$

Substituting (18.9) into (18.7) we find

$$(\mathcal{D}^{i\prime\prime}_j - \Delta\mathcal{D}^i_j)' + (\tilde{a}'/\tilde{a})(\mathcal{D}^{i\prime\prime}_j - \Delta\mathcal{D}^i_j) = 0 \,, \tag{18.11}$$

and hence by integrating (18.11) one obtains

$$\mathcal{D}^{i\prime\prime}_j - \Delta\mathcal{D}^i_j = C^i_j(\boldsymbol{x})/\tilde{a} \,, \tag{18.12}$$

where $C^i_j(\boldsymbol{x})$ are time-independent functions. It is possible to set $C^i_j(\boldsymbol{x}) = 0$, since a nonvanishing $C^i_j(\boldsymbol{x})$ would lead to the following particular solution of (18.12):

$$\mathcal{D}^i_j(\boldsymbol{x}, \eta) = F^i_j(\boldsymbol{x})/\tilde{a} \,, \quad \Delta F^i_j(\boldsymbol{x}) = C^i_j(\boldsymbol{x}) \,. \tag{18.13}$$

As is obvious from (18.9), this solution corresponds to vanishing h^i_j. Hence, from (18.12) it follows that $\mathscr{D}^i_j(\boldsymbol{x}, \eta)$ satisfies the usual flat space–time wave equation

$$\mathscr{D}^{i\prime\prime}_j - \Delta\mathscr{D}^i_j = 0\,. \tag{18.14}$$

Substituting the ansatz (18.9) into the action (18.6), and taking into account that in de Sitter space the Hubble parameter H is a constant yields the following result:

$$\delta_2 S = \frac{1}{24 l^2} \int (\mathscr{D}^{i\prime\prime}_j \mathscr{D}^{j\prime\prime}_i - \mathscr{D}^{i\prime}_{j,k} \mathscr{D}^{j,k\prime}_i)\, \mathrm{d}^4 x\,, \tag{18.15}$$

where total derivatives have been omitted. Under variation with respect to $\mathscr{D}^{i\prime}_j$, we obtain the first time derivative of (18.14).

The quantization of gravitational waves can be reduced to the quantization of a scalar field $\varphi(\boldsymbol{x}, \eta)$ in the background given by (18.2). As a first step, we expand $\mathscr{D}^{i\prime}_j(\boldsymbol{x}, \eta)$ in a Fourier series

$$\mathscr{D}^{i\prime}_j(\boldsymbol{x}, \eta) = \int \frac{\mathrm{d}^3 k}{(2\pi)^3}\, \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} G^i_j(\boldsymbol{k}) v_k(\eta)\,, \tag{18.16}$$

where $G^i_j(\boldsymbol{k})$ is called the polarization tensor. Note that in units of k, both G^i_j and v_k have dimensions $-3/2$. Using the above expansion, we can define a scalar field $\varphi(\boldsymbol{x}, \eta)$ in the following manner:

$$\varphi(\boldsymbol{x}, \eta) = \left(\frac{1}{12 l^2}\right)^{1/2} \int \frac{\mathrm{d}^3 k}{(2\pi)^3} [G^i_j(k) G^j_i(k)]^{1/2} v_k(\eta)\,. \tag{18.17}$$

In terms of this scalar field, the action (18.15) can be rewritten in the form

$$\delta_2 S = \frac{1}{2} \int (\varphi'^2 - \varphi_{,i}\varphi^{,i})\, \mathrm{d}^4 x\,. \tag{18.18}$$

It can be shown [by substituting (18.17) in (18.18) and (18.16) in (18.15)] that the two actions (18.18) and (18.15) are identical up to total derivative terms.

The field equation for φ is the usual flat-space wave equation. It can be obtained by varying the action (18.18) with respect to φ,

$$\varphi'' - \Delta\varphi = 0\,. \tag{18.19}$$

To express h^i_j in terms of φ, we use the equation of motion (18.14) to rewrite (18.9) in the form

$$\Delta h^i_j = (1/\tilde{a}^2)(\tilde{a}\mathscr{D}^{i\prime\prime}_j)'\,. \tag{18.20}$$

The quantization of the scalar field $\varphi(\boldsymbol{x}, \eta)$ is standard. The quantum operator $\hat{\varphi}(\boldsymbol{x}, \eta)$ corresponding to the classical field $\varphi(\boldsymbol{x}, \eta)$ can be expanded in terms of creation and annihilation operators

$$\hat{\varphi} = \frac{1}{2} \int \frac{\mathrm{d}^3 k}{(2\pi)^{3/2}} \frac{1}{\sqrt{k}} [F(k\eta_{\mathrm{i}})\, \mathrm{e}^{-\mathrm{i}(k\eta - \boldsymbol{k}\cdot\boldsymbol{x})} \hat{b}^-_{\boldsymbol{k}} + F^*(k\eta_{\mathrm{i}})\, \mathrm{e}^{+\mathrm{i}(k\eta - \boldsymbol{k}\cdot\boldsymbol{x})} \hat{b}^+_{\boldsymbol{k}}]\,, \tag{18.21}$$

where $\hat{b}_k^+$ and $\hat{b}_k^-$ satisfy the canonical commutation and $\exp[\pm i(k\eta - \boldsymbol{k}\cdot\boldsymbol{x})]$ are a basis of solutions of (18.19). The exact form of the functions $F(k\eta_i)$ is defined by the state of the graviton field at the time η_i. Since $h^i_j(\boldsymbol{x},\eta)$ can be expressed in terms of $\varphi(\boldsymbol{x},\eta)$ using (18.20), (18.16) and (18.17), we can immediately write down the expansion of the operators $\hat{h}^i_j$ in terms of creation and annihilation operators,

$$\hat{h}^i_j = (6l^2)^{1/2}\frac{1}{\tilde{a}^2}\int \frac{d^3k}{(2\pi)^{3/2}}\,k^{-3/2}\,\frac{G^i_j(\boldsymbol{k})}{[G^m_n(\boldsymbol{k})G^n_m(\boldsymbol{k})]^{1/2}}$$
$$\times [F(k\eta_i)(\tilde{a}' - ik\tilde{a})\,e^{-i(k\eta-\boldsymbol{k}\cdot\boldsymbol{x})}\hat{b}_k^- + F^*(k\eta_i)(\tilde{a}' + ik\tilde{a})\,e^{i(k\eta-\boldsymbol{k}\cdot\boldsymbol{x})}\hat{b}_k^+]\,. \tag{18.22}$$

Recall that $\tilde{a} = F^{1/2}a$, where $F = \partial f(R)/\partial R$ is constant in the de Sitter phase since R is constant.

The spectrum of gravitational waves can be calculated by taking expectation values of products of operators $\hat{h}^i_j$ for the quantum state of the system. It is important to consider which are the important physical observables.

18.2. Observables

We will focus on two physical quantities: the power spectrum $\delta_h(k)^2$ and the logarithmic spectrum, i.e., the gravitational wave energy density per logarithmic k interval which will be denoted $k\,d\rho_g(k)/dk$. The "energy density" in gravitational waves is the 0–0 component of the energy-momentum pseudo-tensor (see, e.g., ref. [5]). For a single plane gravitational wave traveling in the z direction, the energy density is

$$\rho_g = (k^2/8\pi G)(|e_{11}|^2 + |e_{22}|^2)\,, \tag{18.23}$$

where e_{ij} is the classical polarization tensor. For a superposition of plane waves,

$$h^i_j(\boldsymbol{x},\eta) = \int d^3k\, e^i_j(\boldsymbol{k},\eta)h(\boldsymbol{k},\eta)\,e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\,, \tag{18.24}$$

with expectation value of e_{ij} normalized to a delta function,

$$\langle e^i_j(\boldsymbol{k})e^j_i(\boldsymbol{k}')\rangle = \delta^3(\boldsymbol{k}-\boldsymbol{k}')\,, \tag{18.25}$$

the expression for the energy density becomes

$$\rho_g = \int d^3k\,\frac{k^2}{8\pi G}\,h^2(k) = \frac{1}{2G}\int_0^\infty dk\,k^4h^2(k)\,. \tag{18.26}$$

Hence, the logarithmic spectrum is

$$k\,d\rho_g(k)/dk = (1/2G)k^2\delta_h(k)^2\,, \tag{18.27}$$

where $\delta_h(k)^2$ denotes the power spectrum

$$\delta_h(k)^2 = k^3 h(k)^2 . \tag{18.28}$$

Note that a flat logarithmic energy-density spectrum of gravitational radiation requires $\delta_h(k)^2 \sim k^{-2}$. As we shall see, this spectrum emerges at late times in inflationary universe models over a large range of wavenumbers. The stochastic gravitational background radiation from cosmic strings also has a flat logarithmic spectrum over a wide band of wavenumbers [163]. In both examples, the result is connected to the scale invariance of the fluctuation generation process [164] which also results in a Harrison–Zel'dovich spectrum of density perturbations [165].

18.3. Spectrum of gravitational waves in de Sitter space

We will first evaluate the spectrum of gravitational radiation in an eternal de Sitter universe. In order to calculate the power spectrum in gravitational waves, we must calculate the expectation value of the square of $\hat{h}$ in the quantum state of the system denoted by $|0\rangle$. The significant difference between gravitational waves (gravitons) and massive scalar matter fields is that the state $|0\rangle$ can be defined to be the de Sitter invariant vacuum state of the system at all times. This means that from the point of view of some observer, there is no graviton production in de Sitter space.

We define the state $|0\rangle$ by setting $F(k\eta_i) = 1$ in the basic expansion (18.21). In this case, the operators $\hat{b}_k^+$ and $\hat{b}_k^-$ are the creation and annihilation operators for gravitons. In particular, $\hat{b}_k^- |0\rangle = 0$ for all k. At this point, the two-point function can be calculated explicitly,

$$\langle 0|\hat{h}^i_j(x)h^j_i(x+r)|0\rangle = \frac{3l^2H^2}{\pi^2 F} \int \left[1 + \left(\frac{k_{\rm ph}}{H}\right)^2\right] \frac{\sin(k_{\rm ph}r_{\rm ph})}{k_{\rm ph}r_{\rm ph}} \frac{{\rm d}k_{\rm ph}}{k_{\rm ph}} , \tag{18.29}$$

where the subscript ph stands for physical quantities. In particular, $k_{\rm ph} = k/a$. To obtain this result, the expansion (18.22) is inserted into the left-hand side of (18.29), the normalization (18.25) for the polarization tensor is used to reduce the expression to one k integral, and the angular k integral is done explicitly, transforming the exponential $\exp(ik \cdot r)$ into the window function $\sin(k_{\rm ph}r_{\rm ph})/k_{\rm ph}r_{\rm ph}$. Since

$$\langle 0|h^i_j(x,\eta)h^j_i(x+r,\eta)|0\rangle = \int {\rm d}^3k\, {\rm e}^{ik\cdot r}h(k)^2 \sim \int_0^\infty \frac{{\rm d}k}{k} \frac{\sin kr}{kr} k^3h(k)^2 , \tag{18.30}$$

we conclude that the power spectrum in eternal de Sitter space is

$$\delta_h^2 \sim (Hl/F^{1/2})^2[1 + (k_{\rm ph}/H)^2] . \tag{18.31}$$

Figure 18.1 is a sketch of this spectrum.

As is apparent from (18.29), the spectrum of gravitational waves leads to an infrared singularity of logarithmic type in the correlation function. However, in an eternal de Sitter universe the "flat" coordinates used here do not cover the whole manifold. In a complete coordinate system which corresponds to the closed de Sitter universe, there is a maximum length scale and hence a small k cutoff. Since we are mainly interested in finite-duration de Sitter periods of a Friedmann cosmology, we shall not further discuss exact de Sitter space results. For more details on the graviton propagator in de Sitter space, see, e.g., ref. [166].

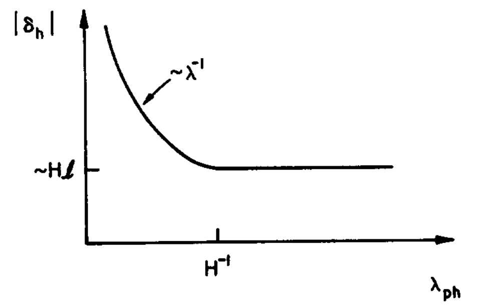

Fig. 18.1. Power spectrum of gravitational radiation in eternal de Sitter space.

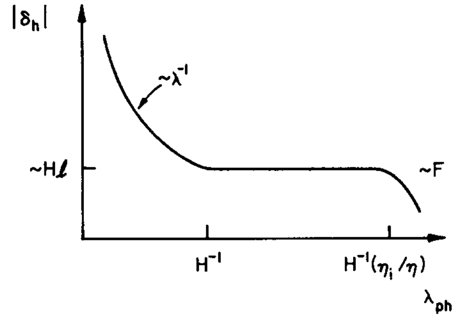

Fig. 18.2. Power spectrum of gravitational radiation at conformal time η during a period of inflation beginning at $\eta = \eta_i$.

18.4. Spectrum of gravitational waves in the inflationary universe

Next, we will compute the spectrum of gravitational waves during the inflationary phase of an inflationary universe. The symbol η_i is used to denote the conformal time at the beginning of the period of exponential expansion. Also, the quantum state $|0\rangle$ of the system is taken to be the vacuum state well before the beginning of the de Sitter phase. The expansion of the field operator $\hat{\varphi}(x, \eta)$ is again given by (18.21), and the operators $\hat{b}_k^+$ and $\hat{b}_k^-$ appearing in this expansion can be interpreted as creation and annihilation operators relative to the state $|0\rangle$. The coefficient functions $F(k)$ are determined by the evolution before η_i and by the junction conditions at η_i. In particular, F depends on the wavenumber k, in contrast to the case of eternal de Sitter space and can be expressed as a function of $k\eta_i$. Modes with $k\eta_i > 1$ are inside the Hubble radius at the beginning of the de Sitter phase, those with $k\eta_i < 1$ are outside.

For $k\eta_i \ll 1$, the form of $F(k\eta_i)$ is determined entirely by the evolution before the de Sitter phase, while modes with $k\eta_i \gg 1$ do not depend significantly on the evolution before the de Sitter phase. Hence,

$$F(k\eta_i) = 1 + O(1/k\eta_i) \quad \text{for} \quad k\eta_i \gg 1 . \tag{18.32}$$

We choose a state $|0\rangle$ without infrared divergences. Hence,

$$F(k\eta_i) \to 0 \quad \text{as} \quad k\eta_i \to 0 . \tag{18.33}$$

Figure 18.2 is a sketch of the resulting spectrum for δ_h^2 which, as before, can be read off from the two-point function

$$\langle 0|\hat{h}^i_j(x, \eta)\hat{h}^j_i(x + r, \eta)|0\rangle = \frac{3l^2H^2}{\pi^2 F} \int_0^\infty \frac{dk_{ph}}{k_{ph}} \frac{\sin(k_{ph} r_{ph})}{k_{ph} r_{ph}} \left| F\left(\frac{k_{ph}}{H} \frac{\eta_i}{\eta}\right)\right|^2 \left[1 + \left(\frac{k_{ph}}{H}\right)^2\right] . \tag{18.34}$$

Finally, we shall compute the spectrum of gravitational waves at late times, long after the end of a finite length de Sitter phase. The key point is that the graviton vacuum state is different in the de Sitter phase and in the post-inflationary period. Therefore, the state which smoothly matches to the de Sitter vacuum state is seen in the late time period as a state containing much larger gravitational wave perturbations than the minimal vacuum quantum fluctuations. This result can be interpreted as production of gravitons during the phase transition.

To simplify the calculations, we shall consider only the case $\mathcal{K}=0$. The time $\Delta t_{\rm dS}$ is the duration of the de Sitter phase. For general $a(\eta)$, the asymptotic forms of the solution of the momentum-mode evolution equations which follow from (18.7) are

$$h^i_j \simeq \begin{cases} G^i_j(\boldsymbol{k})\Big(C_{1k} + C_{2k}\displaystyle\int \mathrm{d}\eta/\tilde{a}^2(\eta)\Big)\,, & k\eta \ll 1\,, \\ G^i_j(\boldsymbol{k})\tilde{a}^{-1}(C_{1k}\sin k\eta + C_{2k}\cos k\eta)\,, & k\eta \gg 1\,. \end{cases} \tag{18.35}$$

Note that at late times $\tilde{a} = F^{1/2}a \simeq a$ since $F \simeq 1$. In turn, this follows since at late times R is small, and thus the corrections to the Einstein action from higher-derivative terms are negligible. On scales smaller than the Hubble radius, the solutions are damped periodic oscillations in conformal time η; whereas on scales larger than H^{-1}, the dominant mode is constant in time. The coefficients C_{1k} and C_{2k} are determined by the de Sitter phase results and by the junction conditions at the time of the phase transition. Consider first scales which cross the Hubble radius during the inflationary phase, i.e.,

$$H\exp(-H\,\Delta t_{\rm dS}) \ll k_{\rm ph} \ll H\,, \tag{18.36}$$

where $k_{\rm ph}$ is the physical wavenumber at the end of inflation. In this case, the power spectrum δ^2_h has constant amplitude $\sim(lH)^2/F$ at the end of inflation. In order to calculate the late time power spectrum, we shall take the standard Friedmann cosmology in which the universe is radiation-dominated and $a(t) \sim t^{1/2}$ from the end of inflation at time $t_1 \sim 10^{-35}$ s until the time $t_{\rm eq}$ of equal matter and radiation ($t_{\rm eq} \sim 10^{11}$ s) after which the universe becomes dominated by cold matter and $a(t) \sim t^{2/3}$.

Combining the results of fig. 18.2 and (18.35) it follows that on length scales $k > k_1$ which were smaller than the Hubble radius at the beginning of inflation [k_1 corresponds to the lower bound in (18.35)] the power spectrum remains constant if the scale is larger than the present Hubble radius. If the scale has entered the Hubble radius after $t_{\rm eq}$, the spectrum δ_h scales as k^{-2}, whereas on scales that enter before $t_{\rm eq}$, δ_h goes as k^{-1}. (Here, k is physical wavenumber at the present time $t_0 \sim 3h^{-1}10^{17}$ s). To summarize (see fig. 18.3),

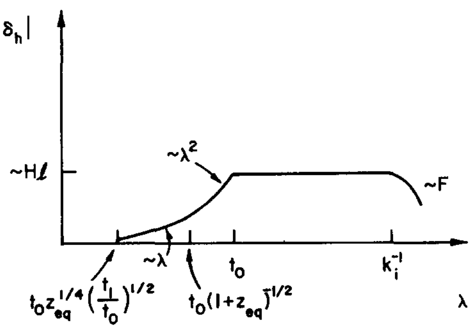

Fig. 18.3. Power spectrum of gravitational radiation in an inflationary universe after reheating.

$$|\delta_h(k)| \simeq (lH/\sqrt{2}\pi F^{1/2}) \begin{cases} |F(k)|\,, & 0<k<k_1\,, \\ 1\,, & k_1<k<t_0^{-1}\,, \\ (t_0k)^{-2}\,, & t_0^{-1}<k<t_0^{-1}(1+z_{eq})^{1/2}\,, \\ (t_0k)^{-1}z_{eq}^{-1/2}\,, & (1+z_{eq})^{1/2}t_0^{-1}<k<t_0^{-1}z_{eq}^{-1/4}(t_0/t_1)^{1/2}\,, \end{cases} \tag{18.37}$$

where $z_{eq} \simeq 10^4$ is the redshift at t_{eq}. The function $F(k)$ reflects the ambiguity on very large wavelengths which were outside the Hubble radius at t_1 and for which the power spectrum is determined by the pre-inflationary evolution.

For most physical questions, only the power spectrum for $k > t_0^{-1}$ is relevant. From the discussion following (18.28) we see that the logarithmic energy spectrum in gravitational waves is flat for $k > z_{eq}^{1/2}t_0^{-1}$. This is an interesting prediction of inflationary universe models.

Note that in inflationary universe models in R^2 gravity, $f(R) = R - (1/6M^2)R^2$. As discussed in chapter 7, during the de Sitter phase $R \gg M$. Hence,

$$F(R) \simeq -\tfrac{1}{3}R/M^2 = 4(H/M)^2\,. \tag{18.38}$$

In this case, the intrinsic amplitude of δ_h in (18.36) is

$$\delta_h \simeq lM/2\pi\sqrt{2}\,. \tag{18.39}$$

For a large mass parameter ($lM \sim 10^{-6}$), the gravity wave background is in the range of what could be observed in planned gravity wave detectors [167]. The comparison between the predicted spectrum of gravitational waves and observational limits is given in fig. 18.4. In this figure, the fraction $\Omega_{GW}(\lambda)$ of the critical density in gravitational radiation is compared with some planned and future detector limits.

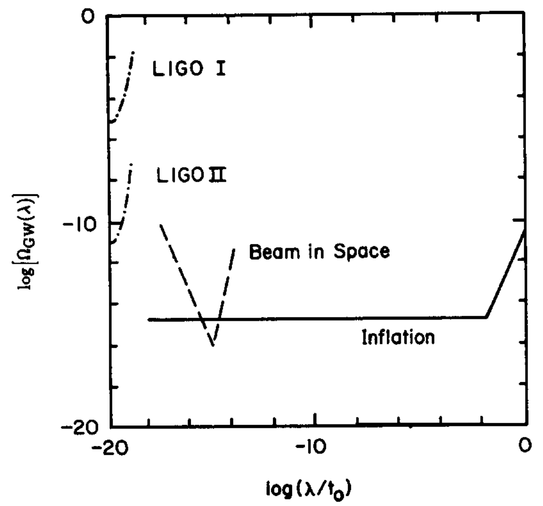

Fig. 18.4. The fraction $\Omega_{GW}(\lambda)$ of the critical energy density in gravitational radiation per octave as a function of the wavelength λ divided by the Hubble radius. The prediction from inflation with $lM \sim 10^{-6}$ is compared with the sensitivities of the LIGO gravitational wave detectors (CALTECH–MIT experiment) and of the proposed laser interferometer in space.

The quantity $\Omega_{GW}(\lambda)$ is obtained by inserting the spectrum (18.37) into (18.27) and dividing by the critical density.

Note that in models of extended inflation [111], there is an extra contribution to the gravity wave background which comes from bubble wall collisions [168]. The spectral shape of that contribution is very different from the component discussed here.

18.5. Spectrum of gravitational waves in double inflation models

In many scenarios of the very early universe there are several inflationary phases [162]. This property is rather generic for models with two or more scalar fields. Here we focus on a model of double inflation and ask if there are observational predictions which are specific to this model. Since gravitational waves interact only weakly with matter, it is convenient to look for features specific to double inflation in the spectrum of gravitational waves. The specific predictions for scalar metric perturbations will be considered in the following chapter. This section is based mainly on the recent work of ref. [161].

Let us consider the double inflation model in which the Hubble parameter evolves as sketched in fig. 18.5, i.e.,

$$H(a) = \begin{cases} H_1 & a < a_1 , \\ H_1(a_1/a)^n , & a_1 < a < a_2 , \\ H_2 & a_2 < a , \end{cases} \tag{18.40}$$

where n is a constant which specifies the equation of state in the interval $a_1 < a < a_2$ between the two periods of inflation. If the universe is radiation-dominated during this phase then $n = 2$, whereas $n = 3/2$ if the equation of state corresponds to dust-like matter.

Since our main interest is the qualitative physical results, we will only consider the asymptotic solutions of the equation of motion for gravitational waves. For a more detailed analysis see ref. [161]. The mode equation for gravitational perturbations in a $\mathcal{K} = 0$ universe is [see (18.7) and (18.24)]

$$h_k'' + 2(a'/a)h_k' + k^2 h_k = 0 , \quad h_k \equiv h(\boldsymbol{k}, \eta) . \tag{18.41}$$

As seen in (18.34), the general form of the solution of (18.41) is

$$h_k \sim C_{1k} + C_{2k} \int \frac{d\eta}{a^2} \tag{18.42}$$

for long-wavelength gravitational waves ($|k\eta| \ll 1$), and

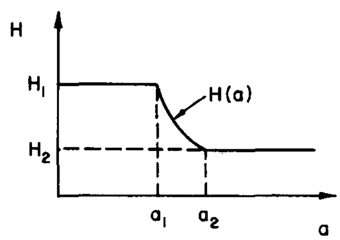

Fig. 18.5. The time dependence of the Hubble parameter in the double inflation model of (18.40).

$$h_k \simeq (l\sqrt{k})\,e^{ik\eta} \tag{18.43}$$

for short-wavelength gravitational waves ($|k\eta| \gg 1$) with the quantum initial conditions discussed in section 18.3.

Using the junction conditions at the time $\eta k = -1$ when the gravitational wave with given k crosses the Hubble radius, and neglecting the decaying mode in (18.42), it follows that the amplitude of h_k for long-wavelength gravitational waves which were initially inside the Hubble radius is

$$h_k \simeq A_k Hl/k^{3/2}\,, \tag{18.44}$$

where A_k is a numerical constant of the order unity.

The special feature of double inflation is that some of the scales which leave the Hubble radius late in the first period of inflation reenter again in the intermediate period $a_1 < a < a_2$ before the onset of the second inflationary phase. This is illustrated in fig. 18.6. Thus, after reentering the Hubble radius, the corresponding waves will begin to oscillate again. From fig. 18.6 it is obvious that only gravitational waves in the wavelength interval

$$H_2 a_2 \equiv k_2 < k < k_1 \equiv H_1 a_1 \tag{18.45}$$

will experience an intermediate oscillating phase.

Let us denote by $a_k^{(1)}$ the value of the scale factor when the scale k [asssumed to lie in the interval (18.45)] reenters the Hubble radius, and by $a_k^{(2)}$ the value of a when it exits again during the second period of inflation. Taking into account that for $a_1 < a < a_2$ when $H \sim a^{-n}$ the Hubble radius increases as

$$H^{-1} \sim t \simeq \int \frac{da}{Ha} \sim a^n\,, \tag{18.46}$$

one immediately obtains

$$a_k^{(1)} \simeq (H_1/k)^{1/(n-1)} a_1^{n/(n-1)}\,. \tag{18.47}$$

Since the value of h_k at the time of reentry is given by (18.43) with $H = H_1$, the evolution of h_k during

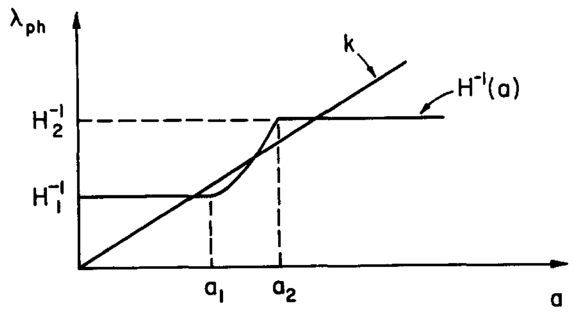

Fig. 18.6. A comoving scale k can leave the Hubble radius during the first period of inflation, reenter in the time interval between a_1 and a_2, and leave for a second time during the second period of inflation.

the period of oscillation is described by

$$h_k = A_k(H_1 l/k^{3/2})(a_k^{(1)}/a)\cos[k(\eta-\eta^{(1)})+\alpha]\,, \tag{18.48}$$

where α is independent of k (a very important fact.). Equation (18.48) is valid for k values in the interval (18.45), and for fixed k in the range $a_k^{(1)} < a < a_k^{(2)}$, where

$$a_k^{(2)} = k/H_2\,. \tag{18.49}$$

After leaving the Hubble radius at $a = a_k^{(2)}$, the value of h_k is once again constant. Hence the nondecaying mode of h_k is

$$\begin{aligned} h_k &\simeq A_k \frac{H_1 l}{k^{3/2}} \frac{a_k^{(1)}}{a_k^{(2)}} \cos\left(k \int_{a_k^{(1)}}^{a_k^{(2)}} \frac{da}{Ha^2} + \alpha\right) \\ &\simeq A_k \frac{H_2 l}{k^{3/2}} \left(\frac{H_1 a_1}{k}\right)^{n/(n-1)} \cos\left(\frac{n}{n-1}\frac{k}{H_2 a_2} + \beta\right). \end{aligned} \tag{18.50}$$

Thus, at the end of the second period of inflation, the power spectrum in gravitational waves is as follows:

$$|\delta_k^h| = (h_k^2 k^3)^{1/2} \sim \begin{cases} H_1 l\,, & H_1 a_i \ll k \ll H_2 a_2\,, \\ H_2 l \left(\dfrac{H_1 a_1}{k}\right)^{n/(n-1)} \left|\cos\left(\dfrac{n}{n-1}\dfrac{k}{H_2 a_2} + \beta\right)\right|, & H_2 a_2 \ll k \ll H_1 a_1\,, \\ H_2 l\,, & H_1 a_1 \ll k \ll H_2 a_f\,, \end{cases} \tag{18.51}$$

where a_i and a_f denote the value of the scale factor at the beginning of the first and end of the second period of inflation, respectively. Note that for $H_2 a_2 \ll k \ll H_1 a_1$, the amplitude of δ_k^h is modulated by $|\cos[n(n-1)^{-1}k/H_2 a_2 + \beta]|$ (see fig. 18.7). Thus, a characteristic signal for double inflation would be a

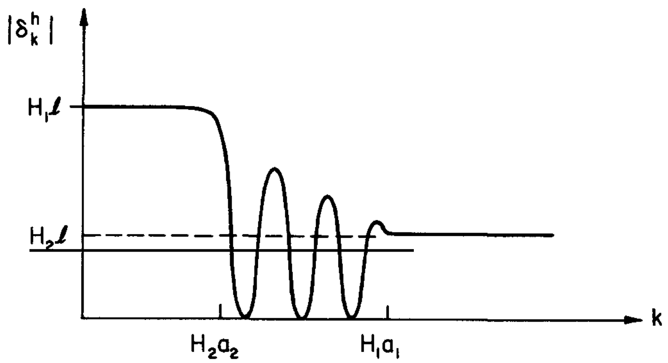

Fig. 18.7. Power spectrum $\delta_h \equiv h$ of gravitational radiation in the double inflation model of (18.40) with the ratio of Hubble constants $H_1/H_2 = 10$.

power spectrum of gravitational waves which is flat for large and small values of k and which manifests the characteristic modulation and decrease in amplitude sketched in fig. 18.7. The rate at which the amplitude decreases would give information about the equation of state in the phase between the two inflationary periods.

Note that (18.51) is the power spectrum at the end of inflation. In order to obtain the power spectrum at late times in the subsequent Friedmann cosmology, the same calculations as at the end of section 3.4 must be done. Note also that in order for the range of k values with nontrivial power spectrum to be observable, the second period of inflation must last less than $60\,H_2^{-1}$. This involves significant fine tuning of the theory.

19. Entropy perturbations

19.1. General remarks

The perturbations considered in most of parts I and II have been adiabatic perturbations, for which the density fluctuation is proportional to the pressure perturbation [see (5.3)]. Typically, only adiabatic perturbations arise in models with a single component for matter. In theories with more than one matter component, other types of perturbations may develop. In fact, the fluctuations will only be adiabatic if the perturbations in the different components of matter are proportional. If they are not, then, in general, entropy perturbations will result, in which case the density fluctuation depends not only on the pressure fluctuation. An isocurvature perturbation is a special case of entropy perturbations in which the energy density and hence the curvature are space independent at some initial moment in time. On scales larger than the Hubble radius, entropy and adiabatic perturbations are decoupled by causality (i.e., an entropy perturbation cannot generate an adiabatic fluctuation). However, on sub-Hubble radius scales, entropy perturbations induce an adiabatic component.

In the context of cosmological models with hydrodynamical matter, entropy perturbations arise if the fluctuations in cold matter and radiation are not proportional. Specific models for their formation have been discussed, e.g., in refs. [169, 74, 85]. They received special consideration since their growth rate is different from that of adiabatic perturbations on scales larger than the Hubble radius (see section 5.4) [6]. Hence, different power spectra can be generated.

In the context of modern cosmological models in which matter is described by fields, entropy perturbations arise quite generally if more than one scalar field contributes substantially to the energy–momentum tensor. Inflationary universe models with axion fields provide a good example [74]. A scalar field χ generates the inflationary period whereas fluctuations in the axion field φ_a develop into entropy perturbations. It was first believed [74] that in these models the adiabatic perturbations would be the dominant ones. However, it was later realized [76, 75] that in many cases the entropy perturbations could in fact dominate. An important reason for the increasing interest in entropy perturbations is the possibility of generating nonflat spectra in these models [76–78]. In addition, a few years ago Peebles [170] initiated a detailed study of structure formation in a particular model denoted the "minimal isocurvature model". Recently, a mechanism which yields the required spectrum of primordial perturbations was suggested [171, 172].

In the past few years, the kinds of spectra which can be generated in models with several scalar fields have been studied extensively [79]. The results depend on the evolution of the fields in the model and in particular on the coupling between the fields. In the following, we shall illustrate the basic mechanism for generating nonflat spectra. The analysis is based on ref. [80]. We conclude this section by noting that the evolution of entropy perturbations has also been studied by Kodama and Sasaki [72, 86] using a gauge invariant formalism. The quantum generation of axion fluctuations was investigated in ref. [173] using similar methods. It is also worth noting that interesting limits on the energy scale at which inflation takes place in axion models have been derived by Lyth [174].

19.2. A model for entropy perturbations

We consider a model consisting of two or more scalar fields. The field χ will be the inflaton. It is assumed to dominate the energy–momentum tensor $T_{\mu\nu}$ at very early times. It simplifies the following analysis to assume that the potential $V(\chi)$ and the initial state of χ are chosen such that a period of exact de Sitter expansion results, i.e., the scale factor $a(t)$ increases as $a(t) \sim \exp(Ht)$, where H is the Hubble constant. The other scalar fields are denoted $\varphi^{(i)}$, $i = 0, 1, \ldots$ We assume that only one of these fields, $\varphi^{(0)} \equiv \varphi$, becomes important at later times in influencing $T_{\mu\nu}$. Note, however, that the other fields may contribute significantly to the interaction terms in the equation of motion for φ.

The action for our model is given by

$$S = \int [\tfrac{1}{2}\chi_{;\mu}\chi^{;\mu} - V(\chi) + \tfrac{1}{2}\varphi_{;\mu}\varphi^{;\mu} - \tfrac{1}{2}m_0^2\varphi^2 - V_{\mathrm{I}}(\chi, \varphi, \ldots)]\sqrt{-g}\, \mathrm{d}^4x \,, \tag{19.1}$$

where V_{I} is the interaction Lagrangian. The first two terms in (19.1) are simply the action for the inflaton χ. The third and fourth terms specify the action of φ in the absence of interaction terms. In the case where φ is the axion, $m_0^2 = 0$ at temperatures higher than the scale of confinement. In the model given by (19.1), fluctuations in χ induce adiabatic perturbations. Their amplitude can be made arbitrarily small by reducing the coupling constants in $V(\chi)$. Hence, we shall neglect adiabatic perturbations in the following discussion. The field φ also induces fluctuations. However, if the energy density contributed by φ is small compared to $V(\chi)$ during the period of inflation, its effect on metric and temperature perturbations is negligible during and immediately after inflation. Hence, only entropy perturbations will be induced.

19.3. Evolution of the homogeneous field

Before discussing the spectrum of entropy perturbations, we will determine the evolution of the homogeneous component φ_0 of the field φ. On a given background, the evolution is defined not only by the scalar field φ_0 but also by other types of matter fields (e.g., χ). This is the main difference of the case under consideration with the model of a homogeneous universe which was considered in chapter 5. The reader only interested in fluctuations may skip to the final result, eq. (19.14), and proceed from there.

The equation of motion for φ_0 becomes

$$\ddot{\varphi}_0 + 3H\dot{\varphi}_0 - \tilde{m}^2_{\mathrm{eff}}(t)\varphi_0 = 0 \,, \tag{19.2}$$

where the effective mass $\tilde{m}_{\mathrm{eff}}(t)$ is determined by

$$\tilde{m}^2_{\mathrm{eff}}(t) = m_0^2 + \varphi_0^{-1}(\partial/\partial\varphi)V_{\mathrm{I}}(\chi, \varphi, \ldots) \,. \tag{19.3}$$

We shall solve (19.2) for the case when $\tilde{m}_{\rm eff}$ is constant. A particular example is when the coupling of φ to other fields is only important during some finite time interval. In this case, outside the interval $\tilde{m}_{\rm eff} = m_0$.

To solve (19.2) [with $\tilde{m}_{\rm eff}(t) = m_0$] it is convenient to use $a(t)$ as temporal variable and to introduce the rescaled field

$$u = H^{1/2} a^2 \varphi_0 \,. \tag{19.4}$$

In terms of u, eq. (19.2) takes the form

$$\frac{d^2u}{da^2} + \left(\frac{m_0^2}{H^2a^2} - \frac{1}{H^{1/2}a^2}\frac{d^2}{da^2}(H^{1/2}a^2)\right)u = 0 \,. \tag{19.5}$$

During the period of inflation H is constant and (19.5) becomes

$$d^2u/da^2 + (m_0^2/H^2 - 2)(1/a^2)u = 0 \,, \tag{19.6}$$

the solutions of which are

$$u \sim a^{1/2 \pm \nu} \,, \quad \nu = (\tfrac{9}{4} - m_0^2/H^2)^{1/2} \,. \tag{19.7a, b}$$

Thus, from (19.4) it follows that

$$\varphi_0 \sim a^{-3/2 \pm \nu} \,. \tag{19.8}$$

Note that if $m_0^2/H^2 > 9/4$, then ν is imaginary and $\varphi_0(t)$ contains an oscillating factor,

$$\varphi_0 \sim a^{-3/2}(\pm i|\nu| \ln a) \,. \tag{19.9}$$

On the other hand, if $m_0^2 \ll H^2$, then to a good approximation $\varphi_0 \sim$ const., and in this case, during inflation the energy density in φ remains constant. In particular, this applies to axion models.

After the end of inflation, H is no longer constant and hence (19.5) must be solved in a different way. It is, however, still possible to solve the equation of motion (19.5) in the interesting asymptotic limits. If $m_0^2 \ll H^2$, the first term in the brackets in (19.5) can be neglected. Thus, one obtains the solution

$$u = C_1 H^{1/2} a^2 + C_2 H^{1/2} a^2 \int da\, H^{-1} a^{-4} \,, \tag{19.10}$$

where C_1 and C_2 are the two integration constants. The dominant mode yields $\varphi_0 \sim$ const.

Once $H(a)$ drops well below m_0, the second term in the brackets in (19.5) can be neglected. If $H \propto a^{-n}$ (a behavior which arises in most interesting applications), then (19.5) becomes

$$d^2u/da^2 + (m_0^2/H_0^2 a_0^n) a^{2(n-1)} u = 0 \,, \tag{19.11}$$

where H_0 and a_0 are constants. The solutions of (19.11) can be expressed in terms of Bessel functions [175],

$$u \sim a^{1/2} J_{1/2n}((m_0/nH_0 a_0^n) a^n) = a^{1/2} J_{1/2n}(m_0/nH(a)) , \tag{19.12}$$

where J_p stands for any of the Bessel (Hankel) functions of order p. Using the asymptotic large-argument form for the Bessel functions [175] we obtain

$$u \sim H^{1/2} a^{1/2} \sin[m_0/nH(a) + \alpha] , \quad m_0/nH(a) \gg 1 , \tag{19.13}$$

where α is some phase constant, and hence

$$\varphi_0 \sim a^{-3/2} \sin[m_0/nH(a) + \alpha] . \tag{19.14}$$

In conclusion, our discussion has shown (see section 19.2) that while $m_0 \ll H(a)$, the nondecaying mode of φ_0 is constant both during and after inflation. Hence, the energy density in φ is constant. Even if during inflation the contribution of φ to $T_{\mu\nu}$ is negligible, it may come to dominate at some later time after the end of inflation. Note that this conclusion does not depend on what determines the background evolution. If the field φ itself determines the evolution of the background, then the condition $m_0^2 \ll H^2(a)$ for which $\varphi_0 \sim \text{const.}$ (and hence its equation of state is $p \sim -\varepsilon$) takes the form $l\varphi \gg 1$, since in this case $H^2(a) \simeq l^2 m_0^2 \varphi^2$. The above is the well known condition for chaotic inflation [100].

The second conclusion which can be drawn from the preceding discussion of the background evolution is that if $m_0^2 \gg H^2(a)$, then [as follows from (19.9) and (19.14)] φ_0 oscillates with an amplitude decreasing as $a^{-3/2}$. Hence, the energy density in φ_0 decreases as a^{-3}, which is the case for cold matter with an equation of state $p = 0$. Note that since the background energy density of the ultrarelativistic matter decreases as $a^{-4}(t)$, the φ field may still come to dominate the energy density at late times. Thus, in a hydrodynamical approach we can consider the homogeneous scalar field φ_0 on a given background as vacuum-like matter ($p \simeq -\varepsilon$) if $H^{-1}(a) < m_0^{-1}$, whereas if $H^{-1}(a) > m_0^{-1}$ then it behaves as dust-like matter ($p \simeq 0$).

Taking into account the above considerations, it follows that in the model discussed in this chapter, in which two scalar fields χ and φ dominate the evolution, the material content of the universe after inflation can be well approximated as consisting of one ultra-relativistic medium (radiation) which appears as a result of the decay of the scalar field χ and one cold component corresponding to the oscillating quasihomogeneous field φ. Then, to analyze the evolution of entropy perturbations generated by fluctuations in φ, the results obtained in section 5.4 can be applied.

19.4. Perturbations

In order to calculate the spectrum of entropy perturbations, we need to know the initial value of $\delta\varepsilon_m/\varepsilon_m$, the relative fluctuation of the energy density in the cold-matter component. Since in our model the field φ acts as cold matter, the relative perturbation is

$$\frac{\delta\varepsilon_m^{(gi)}}{\varepsilon_m} = \frac{\delta\varepsilon_\varphi^{(gi)}}{\varepsilon_\varphi} = \frac{[\delta T_0^{0(gi)}]_\varphi}{(T_0^0)_\varphi} \simeq \frac{-\dot\varphi_0^2 \Phi + \dot\varphi_0\, \delta\dot\varphi^{(gi)} + m_0^2 \varphi_0\, \delta\varphi^{(gi)}}{\frac{1}{2}(\dot\varphi_0^2 + m_0^2 \varphi_0^2)} , \tag{19.15}$$

where the subscript φ denotes the contribution of the field φ to the respective quantity. In (19.15), contributions from the interaction Lagrangian V_{I} in (19.1) were neglected since they are by assumption negligible. It is possible to further approximate (19.15). Because, as follows from (19.8), during inflation

$$\dot{\varphi}_0 \simeq (\nu - \tfrac{3}{2})H\varphi_0 \simeq \tfrac{2}{9}(m_0/H)m_0\varphi_0 \ll m_0\varphi_0 , \quad m_0 \ll H , \tag{19.16}$$

and since at the end of inflation

$$\Phi \sim -\tfrac{1}{2}\, \delta\varepsilon^{(\mathrm{gi})}/\varepsilon \simeq -\tfrac{1}{2}(\varepsilon_\varphi/\varepsilon_\chi)\, \delta\varepsilon_\varphi^{(\mathrm{gi})}/\varepsilon_\varphi \ll \delta\varepsilon_\varphi^{(\mathrm{gi})}/\varepsilon_\varphi \tag{19.17}$$

(using $\varepsilon_\varphi \ll \varepsilon_\chi$ at the end of inflation), we can neglect all but the last term in both the numerator and denominator of (19.15) and obtain

$$\delta\varepsilon_{\mathrm{m}}^{(\mathrm{gi})}/\varepsilon_{\mathrm{m}} \simeq 2\delta\varphi^{(\mathrm{gi})}/\varphi_0 , \qquad \Phi \simeq 0 , \tag{19.18a, b}$$

for scales which are outside the Hubble radius at the end of inflation. Equations (19.18a, b) hold at the end of inflation. However, by continuity at the time of the transition, the equations are also true immediately after inflation and can be used as initial conditions for the evolution in the post-inflationary phase. Note that in the first step of (19.17) we have used (5.55) which relates the metric potential Φ to the gauge-invariant density perturbation on scales larger than the Hubble radius.

Equations (19.18a, b) demonstrate that fluctuations in the scalar field φ generate entropy perturbations. Their evolution in time is shown in fig. 5.2. By (19.18a), the determination of the spectrum of entropy perturbations has been reduced to the evaluation of the spectrum of fluctuations in φ at the end of inflation. The equation of motion for $\delta\varphi^{(\mathrm{gi})}$ was derived in chapter 6 [see (6.46)]. Since [as follows from (19.12)] the metric potential Φ can be neglected, (6.46) takes the form

$$\delta\ddot{\varphi}^{(\mathrm{gi})} + 3H\, \delta\dot{\varphi}^{(\mathrm{gi})} - a^{-2}\nabla^2\, \delta\varphi^{(\mathrm{gi})} + m_{\mathrm{eff}}^2(t)\, \delta\phi^{(\mathrm{gi})} = 0 , \tag{19.19}$$

where the effective mass $m_{\mathrm{eff}}(t)$ [which in general can be different from the mass $\tilde{m}_{\mathrm{eff}}(t)$ appearing in the background equation (19.2)] is

$$m_{\mathrm{eff}}^2(t) = m_0^2 + (\partial^2/\partial^2\varphi)V_{\mathrm{I}}(\chi(a), \varphi(a), \ldots, a(t)) . \tag{19.20}$$

The interaction term V_{I} is crucial in determining the spectrum of $\delta\varphi^{(\mathrm{gi})}$. Different time dependence of the effective mass $m_{\mathrm{eff}}(t)$ can give rise to nontrivial spectra (see refs. [79, 80]). In the following, we shall omit the superscript gi for the gauge invariant scalar field perturbation.

Our goal is to study the different types of spectra which can be generated by appropriate choices of $m_{\mathrm{eff}}^2(t)$. As initial fluctuations for $\delta\varphi$ we will consider the quantum fluctuations discussed in part II of this review. The operator $\delta\hat{\varphi}$ corresponding to $\delta\varphi$ can be expanded in terms of creation and annihilation operators as follows:

$$\delta\hat{\varphi}(\boldsymbol{x}, t) = \frac{1}{\sqrt{2}} \int \frac{\mathrm{d}^3k}{(2\pi)^{3/2}}\, [\delta\varphi_k^*(t)\, \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\hat{a}_{\boldsymbol{k}}^- + \delta\varphi_k(t)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\hat{a}_{\boldsymbol{k}}^+] . \tag{19.21}$$

As discussed in chapter 9 [see also (13.10) and (13.11)], the root-square perturbations are given by the correlation function

$$\langle 0|\delta\hat{\varphi}(\boldsymbol{x}, t)\, \delta\hat{\varphi}(\boldsymbol{x}+\boldsymbol{x}, t)|0\rangle = \int_0^\infty \frac{\mathrm{d}k}{k}\, \frac{\sin(kr)}{kr}\, |\delta_k(t)|^2 \,, \tag{19.22}$$

where the power spectrum $|\delta_k(t)|$ is

$$|\delta_k(t)| = (1/2\pi)|\delta\varphi_k(t)|k^{3/2} \,, \tag{19.23}$$

and completely characterizes the fluctuations on a comoving length scale $\lambda \sim 1/k$ at time t. The mode function $\delta\varphi_k(t)$ obeys (19.19) with ∇^2 replaced by $-k^2$.

The initial conditions for $\delta\varphi_k(t)$ which correspond to choosing the quantum state as the vacuum state at time t_i are (see chapter 11)

$$\delta\varphi_k(t_i) = [1/a(t_i)k^{1/2}]M(k/Ha(t_i)) \,, \qquad \delta\dot{\varphi}_k(t_i) = [\mathrm{i}k^{1/2}/a(t_i)]N(k/Ha(t_i)) \,, \tag{19.24}$$

where the functions M and N obey the normalization condition (11.27) and have the asymptotic behavior $|M| \to 1$ and $|N| \to 1$ when $k \gg Ha(t_i)$.

If m_{eff} and H are constant, then the solution of (19.19) is given in terms of Bessel functions (as can be seen directly by introducing the variable $u = a^{3/2}\varphi_k$ and working in conformal time),

$$\delta\varphi_k(t) = a^{-3/2}[A_k J_{-\nu}(k/Ha) + B_k J_\nu(k/Ha)] = \delta\varphi_k(a) \,, \tag{19.25}$$

$$\nu = (\tfrac{9}{4} - m_{\mathrm{eff}}^2/H^2)^{1/2} \,. \tag{19.26}$$

The coefficients A_k and B_k are fixed by the initial conditions (19.24). The asymptotic limits for the solutions (19.25) can be obtained by combining the asymptotic limits of the Bessel functions [175] with the initial conditions (19.24). We obtain

$$\delta\varphi_k(a) \simeq (1/ak^{1/2}) \exp\{\mathrm{i}(k/H)[1/a - 1/a(t_i)]\} \,, \quad k \gg Ha \,, \tag{19.27}$$

$$\delta\varphi_k(a) \simeq H(Ha)^{-3/2}(k/Ha)^{-\nu} \,, \quad Ha \gg k \gg Ha(t_i) \,. \tag{19.28}$$

To obtain (19.28), we match the small- and large-argument asymptotic solutions at the time of Hubble-radius crossing, i.e., when $k = Ha$. Equations (19.27) and (19.28) will be used to calculate the spectrum of perturbations in various examples.

If the effective mass $m_{\mathrm{eff}}(t)$ is constant during the entire period of inflation, then [as follows from (19.26), (19.27) and (19.23)] at the time t_r corresponding to the end of inflation,

$$|\delta_k(t_r)| \simeq \frac{1}{2\pi} \begin{cases} k_{\mathrm{ph}} \,, & k_{\mathrm{ph}} > H \,, \\ H(k_{\mathrm{ph}}/H)^{3/2-\nu} \,, & H \gg k_{\mathrm{ph}} \gg Ha(t_i)/a(t_r) \,, \end{cases} \tag{19.29}$$

where $k_{ph} = k/a$ is the physical wavenumber. If $m_{eff}^2 \ll H^2$, then the spectrum is nearly flat on all scales which have left the Hubble radius during inflation.

We conclude that fluctuations in a scalar field with an effective mass m_{eff} which is constant in time and satisfies $m_{eff} \ll H$ generate entropy perturbations with a nearly scale-invariant spectrum. In order to obtain derivations from such a flat spectrum it is necessary to consider models with nontrivial time dependence of m_{eff}. In the following we shall demonstrate some examples of nonflat spectra which can be obtained. Note that many models with scale-invariant isocurvature spectra produce a large-angle microwave anisotropy in excess of observational bounds [176].

19.5. Mountain and valley spectra

To illustrate the basic mechanism which leads to nontrivial spectra, consider a model in which at time t_1 the effective mass jumps from some initial value m_1 to some final value m_2 with both masses smaller than $\frac{3}{2}H$.

In each of the time intervals $t_i < t < t_1$ and $t_1 < t < t_r$, the solutions for $\delta\varphi_k$ are given by (19.25) with the respective values of ν. Short-wave perturbations are always described by (19.27). Perturbations on scales which are larger than the Hubble radius at $t = t_1$ are given by

$$\delta\varphi_k(t) = A_k^{(j)} a^{\nu_j - 3/2} + B_k^{(j)} a^{-\nu_j - 3/2}\,, \tag{19.30}$$

with $j = 1, 2$ in the two time intervals. By matching $\delta\varphi_k$ and $\delta\dot{\varphi}_k$ at $t = t_1$ we obtain

$$A_k^{(2)} \simeq [(\nu_1 + \nu_2)/2\nu_2] A_k^{(1)}\,. \tag{19.31}$$

Thus, if $\nu_2 \ll 1$ the amplitude of the dominant mode is amplified. This leads to a spectrum in which modes which leave the Hubble radius before t_1 have a larger amplitude than those which leave after t_1 (see fig. 19.1). In addition, the short-wavelength part of the spectrum will not be scale-invariant.

Let us now consider an example in which the mass evolves as in fig. 19.2 (solid line). From an initial value m_1 with $|m_1| \ll H$ the mass jumps to a larger value m_2 at a time t_1 [the scale factor at this time is $a(t_1) \equiv a_1$]. At a later time t_2, the quantity m_{eff}^2 abruptly decreases to a negative value m_3^2, and finally at time t_3 it returns to a value m_4^2 with $|m_4| \ll H$. The scale factor at time t_2 (t_3) is a_2 (a_3). In this case, the junction conditions can be applied repeatedly at a_1, a_2 and a_3, following the procedure explained in the previous example. For wavelengths which leave the Hubble radius between a_1 and a_2, the spectrum will

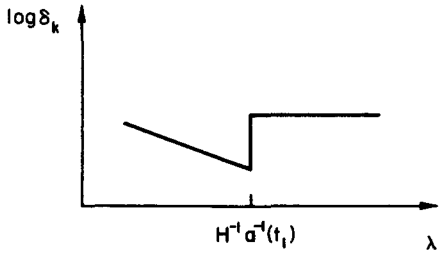

Fig. 19.1. The power spectrum in a model in which the effective mass jumps from some small initial value m_1 to a final value m_2 with $|\nu_2| \ll 1$.

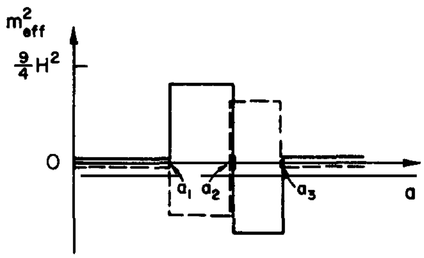

Fig. 19.2. The time dependence of the effective mass in toy models for mountain (solid line) and valley (dashed line) spectra.

rise as k increases, whereas on scales which leave between a_2 and a_3, the spectrum will decrease. If the parameters are fine-tuned, the amplitudes of $|\delta_k|$ on scales which leave the Hubble radius before a_1 or after a_3 can be made equal. One possible fine tuning which achieves this is $\nu_2 + \nu_3 = 3$ together with $a_3/a_2 = a_2/a_1$. The resulting spectrum is

$$|\delta_k| \simeq \begin{cases} A(Ha_1)^{\nu_1-\nu_2}(Ha_2)^{\nu_2-\nu_3}(Ha_3)^{\nu_3-\nu_4}k^{3/2-\nu_1}\,, & Ha(t_i) \ll k \ll Ha_1\,, \\ B(Ha_2)^{\nu_2-\nu_3}(Ha_3)^{\nu_3-\nu_4}k^{3/2-\nu_2}\,, & Ha_1 \ll k \ll Ha_2\,, \\ C(Ha_3)^{\nu_3-\nu_4}k^{3/2-\nu_3} & Ha_2 \ll k \ll Ha_3\,, \\ Dk^{3/2-\nu_4}\,, & Ha_3 \ll k \ll Ha(t_r)\,, \end{cases} \tag{19.32}$$

where A, B, C and D are constants of order unity [which include the jumps in the spectrum analogous to (19.1) in the previous example]. This spectrum is sketched in fig. 19.3 (solid line). It is called a "mountain spectrum".

To obtain a "valley spectrum" we take a model in which m_{eff}^2 first decreases to some negative value and later jumps to a positive value larger than the initial one, as illustrated in fig. 19.2 (dashed line). The resulting spectrum is sketched in fig. 19.3 (dashed line).

19.6. Suppression of long-wavelength perturbations

By introducing a period in the evolution of the universe with $m_{\mathrm{eff}}^2(t) > \frac{9}{4}H^2$ and by tuning parameters carefully, it is possible to obtain a suppression of the long-wavelength part of cosmological perturbation spectra. This is of potential physical relevance when constructing fluctuation spectra from inflationary universe models which give a lot of power on scales of 10–100 Mpc but have little power on the scales which dominate the microwave background.

Let us consider a model in which $m_{\mathrm{eff}}^2(t)$ is almost zero for $a < a_1$ and $a > a_2$ but greater than $\frac{9}{4}H^2$ for $a_1 \leq a \leq a_2$ (see fig. 19.4). A nonzero mass is required to ensure that the energy density in the φ field is dominated by the homogeneous component of φ. This, in turn, is necessary in order that our basic formula (19.18a) for entropy perturbations be valid. There is, obviously, also an upper bound on $m_{\mathrm{eff}}^2(t)$ which stems from requiring that φ does not dominate the energy density. A minimal requirement is $m_{\mathrm{eff}} \ll 1/l$ since the energy density in the inflaton χ is $\varepsilon_\chi = H^2 l^{-2}$ and the energy density in φ is at least $\varepsilon_\varphi > \frac{1}{2}m_{\mathrm{eff}}^2(t)\varphi^2$, where the minimal value of φ from quantum fluctuations [177] is of the order H.

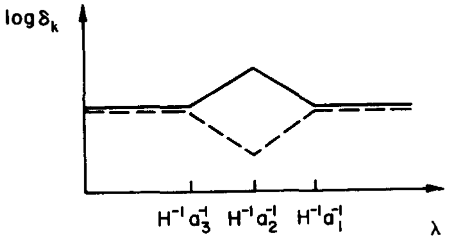

Fig. 19.3. The power spectra in the two models defined by the time-dependent effective masses of fig. 19.2.

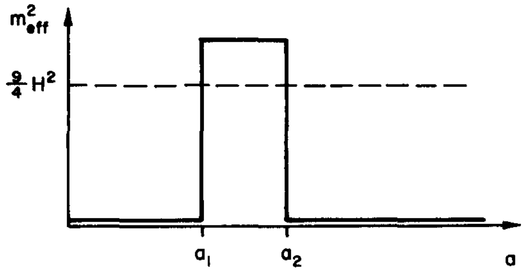

Fig. 19.4. The time dependence of the effective mass in the toy model for suppression of the long-wavelength fluctuations.

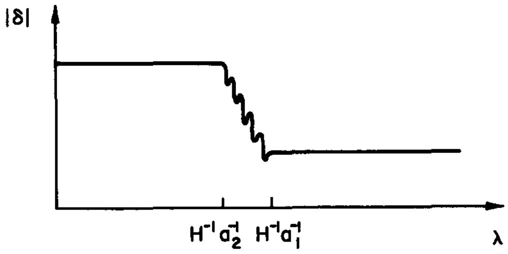

Fig. 19.5. The power spectrum in the model with effective mass sketched in fig. 19.4.

The evolution of $\delta\varphi_k$ is determined by (19.27) and (19.28). The crucial difference compared to the previous examples is that for $a_1 < a < a_2$, the value ν is imaginary. Hence, for perturbations which are outside the Hubble radius at $a = a_1$, we have

$$\delta\varphi_k(a) \simeq \text{const}; \qquad a < a_1, \quad a > a_2, \tag{19.33}$$

$$\delta\varphi_k(a) \simeq A_k a^{-3/2} \sin(|\nu| \ln a + \alpha_k), \quad a_1 < a < a_2 \tag{19.34}$$

$$|\nu| = (m_{\text{eff}}^2/H^2 - \tfrac{9}{4})^{1/2}. \tag{19.35}$$

where α_k is a phase. Thus, perturbations which are outside the Hubble radius at $a = a_1$ oscillate in amplitude between a_1 and a_2. This can lead to a suppression of the long-wavelength part of the spectrum. Scales which leave the Hubble radius between a_1 and a_2 experience a partial damping. The precise form of the spectrum follows from the above solutions, taking into account the junction conditions. The result is

$$|\delta_k| \sim \begin{cases} (a_1/a_2)^{3/2} \sin[\alpha + |\nu| \ln(a_2/a_1)], & Ha(t_i) \ll k \ll Ha_1, \\ (k/Ha_2)^{3/2}[\cosh(\pi|\nu|) - \cos(\beta + 2|\nu| \ln k)]^{1/2}, & Ha_1 \ll k \ll Ha_2, \\ 1, & Ha_2 \ll k \ll Ha(t_r), \end{cases} \tag{19.36}$$

where α and β are constants of order unity. The spectrum is sketched in fig. 19.5. There are evidently two effects which contribute to the suppression of the long-wavelength part of the spectrum. First, between a_1 and a_2 the amplitude of $\delta\varphi_k$ decreases as $a^{-3/2}$ for scales outside the Hubble radius, leading to the suppression factor $(a_1/a_2)^{3/2}$. Second, there is a resonance effect which may lead to additional suppression if the argument of the sine function is close to πn for some integer n.

19.7. Modulation of the spectrum in double inflation models

Let us now return to the double inflation [162] model of section 18.5 in which the Hubble parameter decreases from an initial value H_1 for $a < a_1$ to a final value H_2 for $a > a_2$ [see (18.40)]. For the sake of simplicity we assume $|m_{\text{eff}}(t)| \ll \frac{3}{2}H$. In this case, the equation of motion (19.19) for $\delta\varphi_k$ is identical to the equation (18.41) for gravitational waves. Thus, the analysis of section 18.5 applies and it is possible to immediately write down the spectrum of perturbations of φ (see also ref. [178]). From (18.50) and (18.51) it follows that

$$|\delta_k| \sim \begin{cases} H_1 l\,, & H_1 a_i \ll k \ll H_2 a_2\,, \\ H_2 l (H_1 a_1/k)^{n/(n-1)} |\cos\{[nk/(n-1)H_2 a_2] + \beta\}|\,, & H_2 a_2 \ll k \ll H_1 a_1\,, \\ H_2 l\,, & H_1 a_1 \ll k \ll H_2 a_f\,, \end{cases} \tag{19.37}$$

where β is a constant phase and a_i and a_f are the values of the scale factor at the beginning of the first and the end of the second period of inflation, respectively. For a sketch of this spectrum see fig. 18.7.

The modulation of the spectrum for k values between $H_2 a_2$ and $H_1 a_1$ is similar to Sakharov modulation [7]. The slope of the spectrum in this interval depends on the equation of state in the time interval between the two periods of inflation, whereas the frequency of the modulation depends on the ratio of the Hubble parameters. For example, if the universe is matter-dominated between a_1 and a_2 (i.e., $n = \frac{3}{2}$), then if $H_1/H_2 = 20$, the spectrum has four distinct maxima.

Note that the effect discussed in this section can arise for both adiabatic and entropy perturbations. However, for the distinctive effects to have relevance on cosmologically interesting scales, extreme fine tuning of the model is required. Generically, the nontrivial features in the spectrum will arise on scales much larger than present Hubble radius. This criticism applies not only to the double inflation model, but to all models with particular features in the spectrum. In order for these features to arise on scales of cosmological interest, the models must be fine-tuned such that they occur about 50 expansion times before the end of the final period of inflation.

Note that nontrivial spectra with mountains and valleys can also be obtained from adiabatic perturbations. In fact, we can fix an arbitrary spectrum and construct a family of potentials $V(\varphi)$ which leads to such a spectrum [179]. It is also possible to generate non-Gaussian fluctuations in models with nontrivial potentials [180] (see also ref. [181] for a discussion of non-Gaussian inhomogeneities from higher-order peturbative effects). However, this procedure is extremely unappealing since it implies a complete loss of predictability. We must conclude by stressing that without fine-tuning models, inflation will generically produce a flat spectrum on scales of cosmological interest.

20. Statistical fluctuations

Here we shall investigate how it is possible to define statistical fluctuations of the energy density in an expanding universe with hydrodynamical matter in a consistent way. Obviously, the definition must be independent of the gauge used in the analysis. To be specific, a universe dominated by radiation with equation of state $p = \frac{1}{3}\varepsilon$ will be considered. In this case, there is no phonon creation and thus the amplitude of the statistical fluctuations should be independent of the time when these perturbations are defined.

In the original work by Lifshitz [3], statistical fluctuations were defined in synchronous gauge by

$$(\delta\varepsilon/\varepsilon)(k) \sim \mathcal{N}(k)^{1/2}\,, \tag{20.1}$$

where $\mathcal{N}(k)$ is the number of particles in a box of radius k^{-1}. Since in synchronous gauge, $\delta\varepsilon/\varepsilon$ increases in time as conformal time η on scales outside the Hubble radius, whereas the number of particles per crossing volume does not, the above definition is not independent of the time when the initial conditions are set. In addition, this definition is gauge-dependent. Thus, (20.1) does not form a consistent definition. Note that if (as in the original article [3]), the fluctuations are established at the time of nuclear density, then the resulting amplitude at the present time is much smaller than what is required for galaxy formation, whereas if the Planck time is chosen as the initial time, then the amplitude of fluctuations at late times is much too large.

In the following, we shall use the quantum theory developed in part II of the review to provide a consistent definition of statistical fluctuations [63] (see ref. [182] for a different analysis). This will be based on the gauge-invariant field v on which the quantization scheme of chapters 10 and 11 was based. In terms of the variable v, the fact that in a radiation-dominated universe no photons are produced is reflected in the conformal invariance of the equation of motion of the operator $\hat{v}$ associated with v [see (12.20) and (12.22)],

$$\hat{v}'' - \tfrac{1}{3}\Delta\hat{v} = 0 . \tag{20.2}$$

In particular, conformal invariance implies that the occupation numbers n_k do not depend on time.

Our definition of statistical fluctuations states that the occupation number n_k of the mode with wavevector k is given by the Bose–Einstein distribution at temperature T,

$$n_k = [\exp(\omega/aT) - 1]^{-1} , \quad \omega = k/\sqrt{3} . \tag{20.3}$$

Note that the right-hand side of this equation is time-independent (ω is the comoving frequency).

To find the power spectrum of statistical metric fluctuations, we calculate the two-point function of the gauge-invariant potential Φ in a state with occupation numbers given by n_k. Recall that the expansion of $\hat{v}$ in terms of creation and annihilation operators is

$$\hat{v} = \frac{1}{\sqrt{2}} \int \frac{d^3k}{(2\pi)^{3/2}} [e^{ik\cdot x} v_k^*(\eta)\hat{a}_k^- + e^{-ik\cdot x} v_k(\eta)\hat{a}_k^+] . \tag{20.4}$$

From (20.2), it follows that the solutions for $v_k(\eta)$ are $\omega^{-1/2}\exp(i\omega\eta)$, where the normalization factor is a consequence of the initial conditions discussed after (11.28). Thus

$$\hat{v} = \frac{1}{\sqrt{2}} \int \frac{d^3k}{(2\pi)^{3/2}} \frac{1}{\sqrt{\omega}} (e^{i(k\cdot x - \omega\eta)}\hat{a}_k^- + e^{-i(k\cdot x - \omega\eta)}\hat{a}_k^+) . \tag{20.5}$$

Substituting the above mode functions into the equation (12.19) for the power spectrum $\delta_\Phi(\eta, k)$ and multiplying by $(2n_k + 1)^{1/2}$ yields the result

$$|\delta_\Phi(\eta, k)| = \frac{3}{4\pi} \frac{1}{3^{1/4}} \frac{l}{t} [\coth(\omega/2aT)]^{1/2} \frac{(1 + \omega^2\eta^2)^{1/2}}{\omega\eta} , \tag{20.6}$$

where the proportionality $z \sim a$ has been used. Note that the full two-point function (12.4) is multiplied by $2n_k + 1$. Thus, the power spectrum is multiplied by the square root.

The power spectrum (20.6) contains both the contributions from vacuum and from statistical fluctuations. When $n_k \ll 1$ [i.e., when $\lambda \ll (aT)^{-1}$], quantum fluctuations dominate, whereas when $n_k \gg 1$ [i.e., $\lambda \gg (aT)^{-1}$], statistical perturbations are dominant. Evaluating (20.6) for a scale k at the time $t_{\mathcal{H}}(k)$ when k enters the Hubble radius, and expressing the result in terms of the number $\mathcal{N}$ of particles in a sphere of radius $k^{-1} = \lambda$,

$$\mathcal{N} \sim a^3 T^3 \lambda^3 , \tag{20.7}$$

yields the result

$$|\delta_\Phi(t_{\mathcal{H}}(k), k)| \sim \mathcal{N}^{-1/2} . \tag{20.8}$$

From the above, we can relate the definition of statistical fluctuations given here with the original one [3]. Since at the time of Hubble-radius crossing $\delta\varepsilon^{(\mathrm{gi})}/\varepsilon \sim \Phi$, our definition implies that inside the Hubble radius

$$\delta\varepsilon^{(\mathrm{gi})}/\varepsilon \sim \mathcal{N}^{-1/2} . \tag{20.9}$$

On scales inside the Hubble radius, energy-density perturbations do not grow in the radiation-dominated phase as was shown in section 5.3 [see in particular eq. (5.44)]. Also, the leading term in $\delta\varepsilon/\varepsilon$ when expanding in $(\eta k)^{-1}$ is independent of gauge and equals $\delta\varepsilon^{(\mathrm{gi})}/\varepsilon$. Hence, while being consistent and gauge-invariant, our definition of statistical fluctuations reduces to the naive definition, but only on scales smaller than the Hubble radius.

Finally, we shall show that statistical fluctuations in models without inflation do not give large enough density perturbations for galaxy formation. The amplitudes $\delta_\varepsilon(k)$ of energy-density perturbations in the case of statistical and quantum fluctuations are related as follows:

$$|\delta_{\varepsilon\ \mathrm{statis}}(k)| = (2n_k + 1)^{1/2} |\delta_{\varepsilon\ \mathrm{quantum}}(k)| . \tag{20.10}$$

Hence, from (12.37) we find

$$|\delta_{\varepsilon\ \mathrm{statis}}(k)| \propto k n_k^{1/2} \propto k^{1/2} . \tag{20.11}$$

Multiplying the right-hand side of (12.42) by $(1 + 2n_k)^{1/2}$, the following maximal value of the amplitude $\delta_\varepsilon(t_{\mathrm{rec}})$ of energy-density perturbations at the time of recombination can be derived:

$$\delta_\varepsilon(t_{\mathrm{rec}}, \lambda) \sim \frac{a(t_{\mathrm{rec}})}{a(t_{\mathrm{m}})} \frac{t_{\mathrm{m}}}{t_{\mathrm{r}}} \frac{t_{\mathrm{r}}}{\lambda} \left(\frac{\lambda}{\lambda_\gamma}\right)^{1/2} (\varepsilon_{\mathrm{m}} l^4)^{1/2} (1 - c_{\mathrm{s}}^2) < 10^{-8} , \tag{20.12}$$

where λ_γ is the characteristic wavelength of the microwave background, and the other symbols were defined in chapter 12. For galaxy formation, an amplitude of $\delta_\varepsilon \sim 10^{-4}$ is required. Hence, statistical fluctuations in a model without inflation are too small for galaxy formation.

21. Conclusions

In part III of this review article we have considered four important physical examples for which the gauge-invariant approach to cosmological perturbation theory developed in the previous two parts can easily be applied. The four topics were the derivation of the theory of microwave background anisotropies, the generation and evolution of gravitational radiation, entropy perturbations, and statistical fluctuations.

There are quite obvious advantages of using the gauge-invariant approach. For example, the derivation of the Sachs–Wolfe formula relating density perturbations at last scattering to the CBR anisotropies becomes quite simple. Also, a consistent definition of statistical fluctuations becomes possible. Gravitational wave perturbations are gauge-invariant ab initio. However, the analysis of the generation and evolution of gravitational waves is a straightforward extension of the techniques developed for scalar metric perturbations. The analysis of entropy perturbations in chapter 19 demonstrates that the formalism developed here can be applied in many physically interesting cases.

Acknowledgments

We are grateful to Dick Bond, Ruth Durrer, George Ellis, Jai-chan Hwang, Lev Kofman, David Lyth, Silvia Mollerach, Tomislav Prokopec, Bharat Ratra and Alexei Starobinsky for useful comments on the first draft of this review article. One of us (V.M.) acknowledges stimulating discussions over the years with G. Chibisov, V. Ginsburg, A. Linde, M. Markov and Ya. B. Zel'dovich.

V.M. would like to thank the Brown University Physics Department for hospitality during the time when this review article was written. He also thanks Los Alamos National Laboratory and the Canadian Institute for Theoretical Astrophysics (CITA) for their hospitality. This work has been supported at Brown in part by an Alfred P. Sloan Foundation fellowship to R.B., and in part by the U.S. Department of Energy under contract DE-AC02-76ER03130, Tasks K & A. H.F. was supported in part by the Natural Science and Engineering Research Council of Canada.

Appendix A. Notation

The appendix is organized as follows. First we explain some general rules and describe the units, then we define variables which are used throughout the article. The variables are alphabetized and are divided into four sections: background and perturbed variables and Latin and Greek symbols. We present tables A.1 and A.2 which summarize the main equations which may be useful when applying the theory.

A.1. General comments

A.1.1. Basics

Greek indices α, β, μ, ν, etc. run over the four space–time general coordinate labels $0, 1, 2, 3$ or t, x, y, z.

Latin indices i, j, k, etc. run over the three spatial coordinate labels $1, 2, 3$ or x, y, z.

Einstein summation convention (repeated indices are summed) is employed throughout.

The signature for the metric used throughout is $+1$, -1, -1, -1.

We use the same Riemann and Einstein tensor convention as in Misner et al. [183].

A.1.2. Units

The units employed throughout are $c = \hbar = 1$ where c is the speed of light in vacuum, $\hbar$ is Planck's constant.

G is Newton's gravitational constant.

The Planck length is $l = (\frac{8}{3}\pi G)^{1/2} = 4.7 \times 10^{-33}$ cm.

Table A.1

	Hydrodynamical matter $\mathscr{K}=0, \pm 1$	Scalar-field matter $\mathscr{K}=0$	Higher-derivative gravity
Action	$-\int \sqrt{-g}\, \mathrm{d}^4x \left(\frac{R}{16\pi G} - \varepsilon\right)$	$-\int \sqrt{-g}\, \mathrm{d}^4x \left(\frac{R}{16\pi G} - [\frac{1}{2}\varphi_{,\alpha}\varphi^{,\alpha} - V(\varphi)]\right)$	$-\frac{1}{16\pi G}\int f(R)\sqrt{-g}\, \mathrm{d}^4x$
Metric	$\mathrm{d}s^2 = a^2(\eta)\{(1+2\phi)\,\mathrm{d}\eta^2 - 2B_{\vert i}\,\mathrm{d}x^i\,\mathrm{d}\eta - [(1-2\psi)\gamma_{ij} + 2E_{\vert ij}]\,\mathrm{d}x^i\,\mathrm{d}x^j\}$		
Gauge-invariant metric variables	$\Phi = \phi + (1/a)[(B-E')a]', \qquad \Psi = \psi - (a'/a)(B-E')$		
Gauge-invariant matter variables	$\delta\varepsilon^{(\mathrm{gi})} = \delta\varepsilon + \varepsilon_0'(B-E')$ $\delta p^{(\mathrm{gi})} = \delta p + p_0'(B-E')$ $\delta u_i^{(\mathrm{gi})} = \delta u_i + a(B-E')_{\vert i}$	$\delta\varphi^{(\mathrm{gi})} = \delta\varphi + \varphi_0'(B-E')$	$\delta R^{(\mathrm{gi})} = \delta R + R_0'(B-E')$
Gauge-invariant EoM for perturbations	$\Delta\Psi - 3\mathscr{H}(\mathscr{H}\Phi + \Psi') + 3\mathscr{K}\Psi = 4\pi G a^2\, \delta T_0^{(\mathrm{gi})0}$ $(\mathscr{H}\Phi + \Psi')_{,i} = 4\pi G a^2\, \delta T_i^{(gi)0}$ $[\Psi'' + 2\mathscr{H}\Psi - \mathscr{K}\Psi + (2\mathscr{H}' + \mathscr{H}^2)\Phi + \mathscr{H}\Phi' + \frac{1}{2}\Delta D]\delta_j^i$ $-\frac{1}{2}\gamma^{ik}D_{\vert kj} = -4\pi G a^2\, \delta T_j^{(\mathrm{gi})i}$		same as scalar-field matter with $\Phi \to \tilde{\Phi}, \Psi \to \tilde{\Psi}$, $a \to \tilde{a}, \quad \mathscr{H} \to \tilde{\mathscr{H}}$ where $\tilde{\Phi} = \Phi + (\partial \ln F_0^{1/2}/\partial R)\,\delta R^{(\mathrm{gi})}$ $\tilde{\Psi} = \Psi - (\partial \ln F_0^{1/2}/\partial R)\,\delta R^{(gi)}$
$\delta T_0^{(\mathrm{gi})0}$ $\delta T_i^{(\mathrm{gi})0}$ $\delta T_j^{(\mathrm{gi})i}$	$\delta\varepsilon^{(\mathrm{gi})}$ $(\varepsilon_0 + p_0)\,\delta u_i^{(\mathrm{gi})}$ $-\delta p^{(\mathrm{gi})}\delta_j^i$	$(1/a^2)[\varphi_0'\,\delta\varphi^{(\mathrm{gi})\prime} - \varphi_0'^2\Phi + a^2 V_{,\varphi}\,\delta\varphi^{(\mathrm{gi})}]$ $(1/a^2)\varphi_0'\,\delta\varphi_{,i}^{(\mathrm{gi})}$ $(1/a^2)[-\varphi_0'\,\delta\varphi^{(\mathrm{gi})\prime} + \varphi_0'^2\Phi + a^2 V_{,\varphi}\,\delta\varphi^{(\mathrm{gi})}]$	$\tilde{a} = F^{1/2}a,\ \tilde{\mathscr{H}} = \tilde{a}'/\tilde{a}$ $\varphi = \sqrt{3/16\pi G}\ln F$ $V(\varphi) = (1/6l^2)(f - RF)/F^2$ $F = \partial f/\partial R$

Table A.2

	Hydrodynamical matter $\mathscr{K}=0, \pm 1$	Scalar-field matter $\mathscr{K}=0$
Reduced classical EoM	$u'' - c_s^2\Delta u - (\theta''/\theta)u = a^2(\varepsilon_0 + p_0)^{-1/2}\tau\,\delta S$	$u'' - \Delta u - (\theta''/\theta)u = 0$
u	$(4\pi G)^{-1}\Phi(\varepsilon_0 + p_0)^{-1/2}$	$(a/\varphi_0')\Phi$
θ	$(\mathscr{H}/a)[\frac{2}{3}(\mathscr{H}^2 - \mathscr{H}')]^{-1/2}$	$\mathscr{H}/a\varphi_0'$
Action $\delta_2 S$	$\frac{1}{2}\int\left(v'^2 - c_s^2\gamma^{ij}v_{,i}v_{,j} + \frac{z''}{z}v^2\right)\sqrt{\gamma}\,\mathrm{d}^3x\,\mathrm{d}\eta$	$\frac{1}{2}\int\left(v'^2 - v_{,i}v_{,i} + \frac{z''}{z}v^2\right)\mathrm{d}^4x$
Reduced quantum EoM	$v'' - c_s^2\Delta v - (z''/z)v = 0$	$v'' - \Delta v - (z''/z)v = 0$
v	$(1/\sqrt{6}l)(\delta\varphi_v^{(\mathrm{gi})} - 2z\Phi)$	$a[\delta\varphi^{(\mathrm{gi})} + (\varphi_0'/\mathscr{H})\Phi]$
z	$1/c_s\theta$	$1/\theta$
Relationship between u and v	$\Delta u(\eta) = (z/c_s)[v(\eta)/z]'$	$\Delta u(\eta) = -4\pi G z(v(\eta)/z)'$
Relationship between Φ and v	$\Delta\Phi = -\sqrt{\frac{3}{2}}l\,[(\mathscr{H}^2 - \mathscr{H}' + \mathscr{K})/\mathscr{H}c_s^2](v/z)'$	$\Delta\hat{\Phi} = 4\pi G(\varphi_0'^2/\mathscr{H})(\hat{v}/z)'$

A.1.3. Derivatives

$q_{;\mu}$ covariant derivative.
$q_{\vert i}$ covariant derivative on the background hypersurface of constant time.
$q_{,\mu} \equiv \partial q/\partial x^\mu$ partial derivative.
$\dot{q}$ time derivative of q.
q' conformal time derivative of q.
$V_{,\varphi} = \mathrm{d}V/\mathrm{d}\varphi$.

A.1.4. General variables

$g_{\mu\nu}$ metric.
$G_{\mu\nu}$ Einstein tensor.
$T_{\mu\nu}$ energy–momentum tensor.
$R_{\mu\nu}$ Ricci tensor.
R Ricci scalar
In general for any variable q its background unperturbed value is denoted q_0, its perturbed value is δq and the gauge-invariant form of the perturbation variable is denoted $\delta q^{(\mathrm{gi})}$.

A.1.5. General coordinates

t physical time.
η conformal time, $\mathrm{d}\eta = a^{-1}\,\mathrm{d}t$.
x four-space–time coordinate vector (t, x, y, z).
$\boldsymbol{x}$ cartesian three-vector.

A.1.6. Subscripts

q_r, q_m denote variables in the radiation and matter dominated periods, respectively.
q_rec, q_eq denote variables at the time of recombination and at equal matter and radiation, respectively.
q_ph denotes quantities measured in physical coordinates.
$\delta_n q$ denotes the nth order perturbation term of q.

A.2. Variables used throughout the text

A.2.1. Latin – background

$a = a(t)$ the scale factor of the universe.
c_s the sound velocity for hydrodynamical matter.
$H = \dot{a}/a$ the Hubble parameter in physical time.
${}^{(0)}g_{\mu\nu}$ the background metric.
$g \equiv \det(g_{\mu\nu})$ the determinant of the metric.
$\mathscr{L}$ the Lagrangian.
p, ε unperturbed pressure and energy density, respectively.
S the action.
u^i three-velocity field.
u^α four-fluid-velocity field.
v the gauge-invariant potential. In part II this variable is the single gauge-invariant variable from which we derive all necessary equations to find the power spectra.
$\tilde{v}$ the conformal gauge-invariant potential.
$V(\varphi)$ the scalar-field potential.
$w = p_0/\varepsilon_0$ variable describing the equation of state of matter.

A.2.2. Latin – perturbations

B, E, ϕ, ψ scalar perturbation variables.
$\delta g_{\mu\nu}$ perturbation of the metric.

h_{ij} tensor perturbation variables.
δp, $\delta\varepsilon$ perturbed pressure and energy density, respectively.

A.2.3. Greek – background

γ_{ij} the background metric of a constant time hypersurface.
$\gamma \equiv \det(\gamma_{ij})$.
ε_0, p_0 unperturbed energy density and pressure, respectively.
$\mathcal{H} = a'/a = aH$.
$\mathcal{K}$ spatial curvature constant ($\mathcal{K} = \pm 1, 0$).
Λ cosmological constant.
$\xi = \eta/\eta_{eq}$ where η_{eq} the time at equal matter and radiation.
φ scalar field (usually the inflaton field).

A.2.4. Greek – perturbations

$\delta\varepsilon$, δp perturbed energy density and pressure, respectively.
$\xi_\varepsilon(\boldsymbol{r})$ the two-point energy density correlation function.
ϕ, ψ, B, E metric scalar perturbations.
Φ, Ψ gauge-invariant metric scalar perturbations, respectively.
$\delta\varphi$ perturbed part of the scalar field.

References

[1] N. Mandolesi and N. Vittorio, eds., The Cosmic Microwave Background: 25 Years Later (Kluwer, Dordrecht, 1990).
[2] R. Sachs and A. Wolfe, Astrophys. J. 147 (1967) 73.
[3] E. Lifshitz, Zh. Eksp. Teor. Fiz. 16 (1946) 587.
[4] E. Lifshitz and I. Khalatnikov, Adv. Phys. 12 (1963) 185.
[5] S. Weinberg, Gravitation and Cosmology (Wiley, New York, 1972).
[6] P.J.E. Peebles, The Large-Scale Structure of the Universe (Princeton Univ. Press, Princeton, 1980).
[7] Ya.B. Zel'dovich and I. Novikov, The Structure and Evolution of the Universe, (Chicago Univ. Press, Chicago, 1983).
[8] W. Press and E. Vishniac, Astrophys. J. 239 (1980) 1.
[9] J. Bardeen, Bull. Am. Astron. Soc. 73 (1968).
[10] J. Bardeen, Phys. Rev. D 22 (1980) 1882.
[11] U. Gerlach and U. Sengupta, Phys. Rev. D 18 (1978) 1789.
[12] S. Hawking, Astrophys. J. 145 (1966) 544.
[13] G. Field and L. Shepley, Astrophys. Space Sci. 1 (1968) 309.
[14] E. Harrison, Rev. Mod. Phys. 39 (1967) 862.
[15] D. Olson, Phys. Rev. D 14 (1976) 327.
[16] R. Brandenberger, R. Kahn and W. Press, Phys. Rev. D 28 (1983) 1809.
[17] M. Sasaki, Prog. Theor. Phys. 70 (1983) 394;
H. Kodama and M. Sasaki, Prog. Theor. Phys. Suppl. No. 78 (1984) 1.
[18] D. Lyth, Phys. Rev. D 31 (1985) 1792.
[19] L. Abbott and M. Wise, Nucl. Phys. B 237 (1984) 226.
[20] V. Mukhanov, Zh. Eksp. Teor. Fiz. 94 (1988) 1.
[21] M. Sasaki, Prog. Theor. Phys. 76 (1986) 1036.
[22] A. Goncharov, A. Linde and V. Mukhanov, Inter. J. Mod. Phys. A 2 (1987) 561.
[23] D. Salopek, J. Bond and J. Bardeen, Phys. Rev. D 40 (1989) 1753;
D. Salopek and J. Bond, Phys. Rev. D 43 (1991) 1005.
[24] H. Feldman and R. Brandenberger, Phys. Lett. B 227 (1989) 359.
[25] H. Feldman, Phys. Lett. B 249 (1990) 200.
[26] N. Gouda and M. Sasaki, Prog. Theor. Phys. 76 (1986) 1016.
[27] T. Moody, Phys. Lett. B 149 (1984) 328.

[28] K. Tomita and M. Den, Phys. Rev. D 34 (1986) 3570;
M. Den, Prog. Theor. Phys. 77 (1987) 653.
[29] G. Chibisov and V. Mukhanov, Theory of relativistic potential: cosmological perturbations, Preprint No. 154, P.N. Lebedev Physical Institute (1983).
[30] G. Ellis and M. Bruni, Phys. Rev. D 40 (1989) 1804;
G. Ellis, J. Hwang and M. Bruni, Phys. Rev. D 40 (1989) 1819;
G. Ellis, M. Bruni and J. Hwang, Phys. Rev. D 42 (1990) 1035;
J. Hwang and E. Vishniac, Astrophys. J. 353 (1990) 1.
[31] D. Lyth and M. Mukherjee, Phys. Rev. D 38 (1988) 485.
[32] D. Lyth and E. Stewart, Astrophys. J. 361 (1990) 343.
[33] W. Zimdahl, Classical Quantum Gravity 8 (1991) 677.
[34] R. Durrer and N. Straumann, Helv. Phys. Acta 61 (1988) 1027.
[35] J. Bardeen, in: Particle Physics and Cosmology, ed. A. Zee (Gordon and Breach, London, 1989).
[36] B. Ratra, Phys. Rev. D 38 (1988) 2399.
[37] P. D'Eath, Ann. Phys. 98 (1976) 237.
[38] J. Stewart and M. Walker, Proc. R. Soc. London A 341 (1974) 49.
[39] M. Bruni, P. Dunsby and G. Ellis, Cosmological perturbations and the physical meaning of gauge-invariant variables, SISSA preprint 138/91/A (1991).
[40] P. Dunsby, M. Bruni and G. Ellis, Covariant perturbations in a multi-fluid cosmological medium, SISSA preprint 139/91/A (1991);
M. Bruni, G. Ellis and P. Dunsby, Gauge-invariant perturbations in a scalar field dominated universe, SISSA preprint 77/91/A (1991).
[41] A. Guth, Phys. Rev. D 23 (1981) 347.
[42] G. Chibisov and V. Mukhanov, Galaxy formation and phonons, P.N. Lebedev Physical Institute Preprint No. 162 (1980); Mon. Not. R. Astron. Soc. 200 (1982) 535.
[43] V. Lukash, Pis'ma Zh. Eksp. Teor. Fiz. 31 (1980) 631.
[44] A. Sakharov, Zh. Eksp. Teor. Fiz. 49 (1965) 245.
[45] W. Press, Phys. Scr. 21 (1980) 702.
[46] K. Sato, Mon. Not. R. Astron. Soc. 195 (1981) 467.
[47] E. Harrison, Phys. Rev. D1 (1970) 2726.
Ya. Zel'dovich, Mon. Not. R. Astron. Soc. 160 (1972) 1P.
[48] V. Mukhanov and G. Chibisov, Pis'ma Zh. Eksp. Teor. Fiz. 33 (1981) 549.
[49] A. Starobinsky, Phys. Lett. B 91 (1980) 99.
[50] A. Starobinsky, Phys. Lett. B 117 (1982) 175.
[51] S. Hawking, Phys. Lett. B 115 (1982) 295.
[52] A. Guth and S.-Y. Pi, Phys. Rev. Lett. 49 (1982) 1110.
[53] J. Bardeen, P. Steinhardt and M. Turner, Phys. Rev. D 28 (1983) 679.
[54] V. Mukhanov and G. Chibisov, Zh. Eksp. Teor. Fiz. 83 (1982) 475.
[55] R. Brandenberger and R. Kahn, Phys. Rev. D 29 (1984) 2172;
J. Frieman and M. Turner, Phys. Rev. D 30 (1984) 265.
[56] R. Brandenberger, Rev. Mod. Phys. 57 (1985) 1.
[57] V. Mukhanov, Pis'ma Zh. Eksp. Teor. Fiz. 41 (1985) 402;
L. Kofman and V. Mukhanov, Pis'ma Zh. Eksp. Teor. Fiz. 44 (1986) 481.
[58] A. Vilenkin, Nucl. Phys. B 226 (1983) 527.
[59] E. Mottola, Phys. Rev. D 27 (1983) 2294.
[60] J. Kripfganz and H. Perlt, Classical Quantum Gravity 3 (1986) 417.
[61] R. Kahn and R. Brandenberger, Phys. Lett. B 141 (1984) 317.
[62] J. Stewart, Classical Quantum Gravity 7 (1990) 1169.
[63] V. Mukhanov, Evolution of primordial fluctuations in the universe, Dissertation, USSR Academy of Sciences (1989).
[64] H. Feldman, The Dynamics of Scalar Fields in the Early Universe, Ph.D. thesis, SUNY, Stony Brook (1989).
[65] B. Bednarz, Phys. Rev. D 31 (1985) 2674.
[66] R. Durrer, Astron. Astrophys. 208 (1989) 1.
[67] R. Durrer, Phys. Rev. D 42 (1990) 2533.
[68] J. Hwang, Large Scale Evolution of a Perturbed Universe, Preprint, University of Texas (1990).
[69] J. Hwang and E. Vishniac, Astrophys. J. (1991) in press.
[70] J. Hwang, Classical Quantum Gravity 7 (1990) 1613; Phys. Rev. D 42 (1990) 2601.
[71] M. Kasai and K. Tomita, Phys. Rev. D 33 (1986) 1576;
U. Kraemmer and A. Rebhan, Phys. Rev. Lett. 67 (1991) 793.
[72] H. Kodama and M. Sasaki, Inter. J. Mod. Phys. A 1 (1986) 265; A 2 (1987) 491.
[73] D. Kompaneetz, V. Lukash and I. Novikov, Astron. Zh. 59 (1982) 424.

[74] M. Turner, F. Wilczek and A. Zee, Phys. Lett. B 125 (1983) 35;
M. Axenides, R. Brandenberger and M. Turner, Phys. Lett. B 126 (1983) 178.
[75] D. Seckel and M. Turner, Phys. Rev. D 32 (1985) 3178.
[76] A. Linde, Pis'ma Zh. Eksp. Teor. Fiz. 40 (1984) 496; Phys. Lett. B 158 (1985) 375.
[77] L. Kofman and A. Linde, Nucl. Phys. B 282 (1987) 555.
[78] V. Mukhanov, L. Kofman and D. Pogosyan, Phys. Lett. B 193 (1987) 427.
[79] L. Kofman, Phys. Lett. B 173 (1986) 400;
L. Kofman and D. Pogosyan, Phys. Lett. B 214 (1988) 508;
D. Salopek, J. Bond and J. Bardeen, Phys. Rev. 40 (1989) 1753.
[80] V. Mukhanov and M. Zel'nikov, Phys. Lett. B 263 (1991) 169.
[81] S. Mollerach, Phys. Rev. D 42 (1990) 313; Phys. Lett. B 242 (1990) 158.
[82] T.W.B. Kibble, J. Phys. A 9 (1976) 1387.
[83] T. Vachaspati and A. Vilenkin, Phys. Rev. D 30 (1984) 2036;
A. Albrecht and N. Turok, Phys. Rev. Lett. 54 (1985) 1868.
[84] Ya. Zel'dovich, Adv. Astron. Astrophys. 3 (1965) 241;
C. Hogan, N. Kaiser and M. Rees, Philos. Trans. R. Soc. London A 307 (1982) 97;
J. Traschen, Phys. Rev. D 29 (1984) 1563;
L. Abbott and J. Traschen, Astrophys. J. 302 (1986) 39.
[85] J. Traschen, N. Turok and R. Brandenberger, Phys. Rev. D 34 (1986) 919.
[86] H. Kodama, Prog. Theor. Phys. 71 (1984) 946; 73 (1985) 674.
[87] E. Kamke, Differential Equations (Nauka, Moscow, 1980).
[88] L. Grishchuk and Ya. Zel'dovich, Astron. Zh. 55 (1978) 20;
J. Silk and J. Wilson, Phys. Scr. 21 (1980) 708.
[89] B. Ratra, Inflation in an exponential potential model, Preprint PUPT-1102, Princeton University (1989).
[90] G. Chibisov, Physical processes in the universe, thesis, P.N. Lebedev Physical Institute, Moscow (1972).
[91] P. Higgs, Phys. Lett. 12 (1964) 132; Phys. Rev. Lett. 13 (1964) 508;
F. Englert and R. Brout, Phys. Rev. Lett. 13 (1964) 3;
G. Guralnik, C. Hagen and T.W.B. Kibble, Phys. Rev. Lett. 13 (1964) 585.
[92] C. Itzykson and J.-B. Zuber, Quantum Field Theory (McGraw-Hill, New York, 1980).
[93] D. Kirzhnits, Pis'ma Zh. Eksp. Teor. Fiz. 15 (1982) 745;
D. Kirzhnits and A. Linde, Phys. Lett. B 42 (1972) 471;
S. Weinberg, Phys. Rev. Lett. 36 (1976) 294;
A. Linde, Pis'ma Zh. Eksp. Teor. Fiz. 23 (1976) 73.
[94] A. Linde, Pis'ma Zh. Eksp. Teor. Fiz. 19 (1974) 320;
S. Bludman and M. Ruderman, Phys. Rev. Lett. 38 (1977) 255.
[95] M. Markov and V. Mukhanov, Phys. Lett. A 104 (1984) 200;
V. Belinsky, L. Grishchuk, I. Khalatnikov and Ya. Zel'dovich, Phys. Lett. B 155 (1985) 232;
L. Kofman, A. Linde and A. Starobinsky, Phys. Lett. B 157 (1985) 36;
T. Piran and R. Williams, Phys. Lett. B 163 (1985) 331.
[96] J. Kung and R. Brandenberger, Phys. Rev. D 42 (1990) 1008.
[97] R. Brandenberger, H. Feldman and J. Kung, Phys. Scr. T 36 (1991) 64.
[98] D. Goldwirth and T. Piran, Phys. Rev. Lett. 64 (1990) 2852.
[99] V. Arnold, Mathematical Methods of Classical Mechanics (Springer, New York, 1978).
[100] A. Linde, Phys. Lett. B 129 (1983) 177.
[101] V. Arnold, Ordinary Differential Equations (MIT Press, Cambridge, 1973).
[102] L. Abbott and M. Wise, Nucl. Phys. B 244 (1984) 541;
F. Lucchin and S. Matarrese, Phys. Rev. D 32 (1985) 1316; Phys. Lett. B 164 (1985) 282;
J. Halliwell, Phys. Lett. B 185 (1987) 341;
J. Yokoyama and K. Maeda, Phys. Lett. B 207 (1988) 31.
[103] V. Mukhanov, Pis'ma Zh. Eksp. Teor. Fiz. 41 (1985) 402.
[104] A. Linde, Mod. Phys. Lett. A 1 (1986) 81.
[105] R. Brandenberger, Rev. Mod. Phys. 57 (1985) 1; Inter. J. Mod. Phys. A 2 (1987) 77.
[106] D. Blochinzev, Space, Time and Microworld (Nauka, Moscow, 1970).
[107] A. Albrecht, P. Steinhardt, M. Turner and F. Wilczek, Phys. Rev. Lett. 48 (1982) 1437.
L. Abbott, E. Farhi and M. Wise, Phys. Lett. B 117 (1982) 29;
J. Traschen and R. Brandenberger, Phys. Rev. D 42 (1990) 2491.
[108] K. Stelle, Phys. Rev. D 16 (1977) 953.
[109] D. Capper and M. Duff, Nuovo Cimento 23A (1974) 173;

J. Dowker and R. Critchley, Phys. Rev. D 13 (1976) 3224;
S. Deser, M. Duff and C. Isham, Nucl. Phys. B 111 (1976) 45;
S. Hawking, Commun. Math. Phys. 55 (1977) 133;
S. Christensen and S. Fulling, Phys. Rev. D 15 (1977) 2088;
L. Brown and J. Cassidy, Phys. Rev. D 16 (1977) 1712;
T. Bunch and P.C.W. Davies, Proc. R. Soc. London A 356 (1977) 569;
A. Duncan, Phys. Rev. D 17 (1978) 964;
S. Mamaev and V. Mostepanenko, Phys. Lett. A 67 (1978) 203;
R. Wald, Phys. Rev. D 17 (1978) 1477;
S. Mamaev, Teor. Mat. Fiz. 42 (1980) 350.
[110] B. Zwiebach, Phys. Lett. B 156 (1985) 315;
A. Tseytlin, Phys. Lett. B 176 (1986) 92;
P. Oh, Phys. Lett. B 166 (1986) 292.
[111] C. Mathiazhagan and V. Johri, Classical Quantum Gravity 1 (1984) L24;
D. La and P. Steinhardt, Phys. Rev. Lett. 62 (1989) 376;
P. Steinhardt and F. Accetta, Phys. Rev. Lett. 64 (1990) 2740.
[112] V. Gurovich and A. Starobinsky, Zh. Eksp. Teor. Fiz. 77 (1979) 1683;
M. Fishetti, J. Hartle and B.-L. Hu, Phys. Rev. D 20 (1979) 1757;
A. Starobinsky, Pis'ma Astron. Zh. 9 (1983) 579;
P. Anderson, Phys. Rev. D 29 (1984) 615;
A. Vilenkin, Phys. Rev. D 32 (1985) 2511.
[113] M. Mijic, M. Morris and W.-M. Suen, Phys. Rev. D 34 (1986) 2934.
[114] B. Whitt, Phys. Lett. B 145 (1984) 176.
[115] L. Kofman and V. Mukhanov, Pis'ma Zh. Eksp. Teor. Fiz. 44 (1986) 481;
L. Kofman, V. Mukhanov and D. Pogosyan, Zh. Eksp. Teor. Fiz. 93 (1987) 769;
V. Mukhanov, L. Kofman and D. Pogosyan, Phys. Lett. B 193 (1987) 427.
[116] J. Barrow and S. Cotsakis, Phys. Lett. 215 B (1988) 515;
K. Maeda, Phys. Rev. D 39 (1989) 3159.
[117] N. Deruelle, C. Gundlach and D. Polarski, On the quantization of perturbations in inflation, DAMTP preprint R-91/7 (1991).
[118] R. Brandenberger, Nucl. Phys. B 245 (1984) 328;
R. Brandenberger and C. Hill, Phys. Lett. B 179 (1986) 30.
[119] R. Brandenberger, in: Physics of the Early Universe, eds. J. Peacock, A. Heavens and A. Davies (Scottish Universities Summer School, Edinburgh, 1989).
[120] V. Mukhanov, Phys. Lett. B 218 (1989) 17.
[121] Y. Nambu and M. Sasaki, Prog. Theor. Phys. 83 (1990) 37;
W. Unruh, Gauge-invariant density perturbations, University of British Columbia preprint (1988), unpublished.
[122] S. Wada, Mod. Phys. Lett. A 3 (1988) 929;
I. Shirai and S. Wada, Nucl. Phys. B 303 (1988) 728.
[123] W. Fischler, B. Ratra and L. Susskind, Nucl. Phys. B 259 (1985) 730.
[124] J. Halliwell and S. Hawking, Phys. Rev. D 31 (1985) 1777.
[125] R. Arnowitt, S. Deser and C. Misner, in: Gravitation: An Introduction to Current Research, ed. L. Witten (Wiley, New York, 1962).
[126] V. Fock, Theory of Space, Time and Gravitation (Pergamon, London, 1959).
[127] A. Grib, S. Mamaev and V. Mostepanenko, Quantum Effects in Strong External Fields (Atomizdat, Moscow, 1980).
[128] N. Birrell and P. Davies, Quantum Fields in Curved Space (Cambridge Univ. Press, Cambridge, 1982).
[129] V. de Lapparent, M. Geller and J. Huchra, Astrophys. J. Lett. 302 (1986) L1;
G. Efstathiou, N. Kaiser, W. Saunders, A. Lawrence and M. Rowan-Robinson, Mon. Not. R. Astron. Soc. 247 (1991) 10P.
[130] G. Smoot, M. Gorenstein and R. Muller, Phys. Rev. Lett. 39 (1977) 898.
[131] R. Partridge, Rep. Prog. Phys. 51 (1988) 647.
[132] L. Grishchuk, A. Doroshkevich and I. Novikov, Zh. Eksp. Teor. Fiz. 28 (1969) 1210;
A. Doroshkevich, V. Lukash and I. Novikov, Sov. Astron. 18 (1975) 554;
L. Grishchuk and Ya. B. Zel'dovich, Sov. Astron. 21 (1978) 125.
[133] G. Daucourt, Mon. Not. R. Astron. Soc. 144 (1969) 255;
A. Anile and S. Motta, Astrophys. J. 207 (1976) 685.
[134] P.J.E. Peebles and J. Yu, Astrophys. J. 162 (1970) 815.
[135] P.J.E. Peebles, Astrophys. J. Lett. 243 (1981) L119; Astrophys. J. 259 (1982) 442; Astrophys. J. Lett. 263 (1982) L1.
[136] J. Silk and M. Wilson, Phys. Scr. 21 (1980) 708; Astrophys. J. Lett. 244 (1981) L37.
M. Wilson and J. Silk, Astrophys. J. 243 (1981) 14.
[137] N. Kaiser, Mon. Not. R. Astron. Soc. 198 (1982) 1033;

R. Fabbri, I. Guidi and V. Natale, Astron. Astrophys. 122 (1983) 151;
K. Tomita and K. Tanabe, Prog. Theor. Phys. 69 (1983) 828.
[138] J. Traschen and D. Eardley, Phys. Rev. 34 (1986) 1665.
[139] A. Starobinsky, Sov. Astron. Lett. 9 (1983) 579; 11 (1985) 323;
L. Abbott and M. Wise, Astrophys. J. Lett. 282 (1984) L47; Phys. Lett. B 135 (1984) 279;
L. Kofman, D. Pogosyan and A. Starobinsky, Sov. Astron. Lett. 12 (1986) 419;
L. Kofman and A. Starobinsky, Sov. Astron. Lett. 11 (1985) 643.
[140] J. Bond and G. Efstathiou, Astrophys. J. Lett. 285 (1984) L45; Mon. Not. R. Astron. Soc. 226 (1987) 635.
[141] G. Efstathiou, in: Physics of the Early Universe, eds. J. Peacock, A. Heavens and A. Davies (Scottish Universities Summer School in Physics, Edinburgh, 1989).
[142] V. Mukhanov and G. Chibisov, Sov. Astron. Lett. 10 (1984) 890.
[143] M. Panek, Phys. Rev. D 34 (1986) 416.
[144] L. Abbott and R. Schaefer, Astrophys. J. 308 (1986) 546.
[145] I.A. Stukov, D.P. Skulachev and A.A. Klypin, in: Large Scale Structure of the Universe eds. J. Audouze et al. (Kluwer, Dordrecht, 1988);
G. Smoot et al., Astrophys. J. Lett. 371 (1991) L1;
S. Meyer, E. Cheng and L. Page, Astrophys. J. Lett. 371 (1991) L7.
[146] N. Vittorio and J. Silk, Astrophys. J. Lett. 285 (1984) L39.
[147] V. Sahni, H.A. Feldman and A. Stebbins, Astrophys. J. 385 (1992) 1.
[148] L. Grishchuk, Zh. Eksp. Teor. Fiz. 67 (1974) 825.
[149] L. Grishchuk and Yu. Sidorov, Classical Quantum Gravity 6 (1989) L161.
[150] Yu. Sidorov, Europhys. Lett. 10 (1989) 415.
[151] L. Grishchuk and Yu. Sidorov, Phys. Rev. D 42 (1990) 3413.
[152] A. Starobinsky, Pis'ma Zh. Eksp. Teor. Fiz. 30 (1979) 719.
[153] V. Rubakov, M. Sazhin and A. Veryaskin, Phys. Lett. B 115 (1982) 189.
[154] R. Fabbri and M. Pollock, Phys. Lett. B 125 (1983) 445.
[155] V. Mukhanov, Ph.D. thesis, Moscow Physical Technical Institute (1982), unpublished.
[156] L. Abbott and M. Wise, Nucl. Phys. B 244 (1984) 541.
[157] L. Abbott and M. Wise, Phys. Lett. B 135 (1984) 279.
[158] L. Abbott and D. Harari, Nucl. Phys. B 264 (1986) 487.
[159] B. Allen, Phys. Rev. D 37 (1988) 2078.
[160] V. Sahni, Phys. Rev. D 42 (1990) 453.
[161] M. Zelnikov and V. Mukhanov, Pis'ma Zh. Eksp. Teor. Fiz. 54 (1991) 201.
[162] L. Kofman, A. Linde and A. Starobinsky, Phys. Lett. B 157 (1985) 361;
A. Starobinsky, Pis'ma Zh. Eksp. Teor. Fiz. 42 (1985) 152;
J. Silk and M. Turner, Phys. Rev. D 35 (1987) 419.
[163] A. Vilenkin, Phys. Lett. B 107 (1982) 47;
C. Hogan and M. Rees, Nature 311 (1984) 109;
R. Brandenberger, A. Albrecht and N. Turok, Nucl. Phys. B 277 (1986) 605.
[164] Ya. B. Zel'dovich, Mon. Not. R. Astron. Soc. 192 (1980) 663;
A. Vilenkin, Phys. Rev. Lett. 46 (1981) 1169;
A. Vilenkin, Phys. Rev. 121 (1985) 263.
[165] R. Brandenberger and N. Turok, Phys. Rev. D 33 (1985) 2182.
[166] B. Allen, Phys. Rev. D 34 (1986) 3670;
B. Allen and M. Turyn, Nucl. Phys. B 292 (1987) 813;
E. Floratos, J. Iliopoulos and T. Tomaras, Phys. Lett. B 197 (1987) 373.
[167] K. Thorne, in: 300 Years of Gravitation, eds. S. Hawking and W. Israel (Cambridge Univ. Press, Cambridge, 1987);
C. Hogan, Mon. Not. R. Astron. Soc. 218 (1986) 629.
[168] M. Turner and F. Wilczek, Phys. Rev. Lett. 65 (1990) 3080.
[169] J. Bond, E. Kolb and J. Silk, Astrophys. J. 255 (1982) 341;
M. Fukugita and V. Rubakov, Phys. Rev. Lett. 56 (1986) 988;
K. Iso, H. Kodama and K. Sato, Phys. Lett. B 169 (1986) 337.
[170] P.J.E. Peebles, Astrophys. J. Lett. 315 (1987) L73;
P.J.E. Peebles, Nature 327 (1987) 210.
[171] J. Yokoyama and Y. Suto, Preprint, University of Tokyo UTAP-110/90 (1990).
[172] M. Sasaki and J. Yokoyama, Initial conditions for the minimal isocurvature scenario, Preprint, University of Tokyo UTAP-115/91 (1991).
[173] Y. Nambu and M. Sasaki, Phys. Rev. D 42 (1990) 3918.
[174] D. Lyth, Phys. Lett. B 236 (1990) 400.
[175] M. Abramowitz and I. Stegun, Handbook of Mathematical Functions (Dover Publ., New York, 1965).

[176] G. Efstathiou and J.R. Bond, Mon. Not. R. Astron. Soc. 218 (1986) 103;
G. Efstathiou and J.R. Bond, Mon. Not. R. Astron. Soc. 227 (1987) 33P.
[177] A. Vilenkin and L. Ford, Phys. Rev. D 26 (1982) 1231;
A. Linde, Phys. Lett. B 132 (1983) 137;
R. Brandenberger, Nucl. Phys. B 245 (1984) 328.
[178] S. Gottloeber, V. Mueller and A. Starobinsky, Phys. Rev. D 43 (1991) 2510.
[179] H. Hodges and G. Blumenthal, Phys. Rev. D 42 (1990) 3329.
[180] H. Hodges, G. Blumenthal, L. Kofman and J. Primack, Nucl. Phys. B 335 (1990) 197.
[181] T. Allen, B. Grinstein and M. Wise, Phys. Lett. B 197 (1987) 66.
[182] M. Morikawa, Prog. Theor. Phys. 77 (1987) 1163.
[183] C. Misner, K. Thorne and J. Wheeler, Gravitation (Freeman, New York, 1970).