Long Wavelength Cosmological Perturbations and Preheating

by

James Peter Zibin

B.Sc., The University of Victoria, 1989
M.Sc., Simon Fraser University, 1995

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Department of Physics and Astronomy)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

March 2004

Abstract

This thesis is concerned with the evolution of long wavelength cosmological perturbations in the very early inflationary and post-inflationary stages of the universe. I first provide a thorough review of the relevent theoretical background. This material is presented in a completely original manner, with essentially all of the required results exposed together. Emphasis is made throughout on elucidating the physical meaning of the results. I next perform a study of a particular inflationary model for which there can be explosive growth of long wavelength perturbations due to the process of parametric resonance, and I try to determine whether the backreaction of small scale perturbations is sufficient to save the standard inflationary predictions. I conclude that, for certain parameter values, it is not. Then I describe in considerable detail general aspects of the evolution of long wavelenth modes. I provide a careful link between the evolution of a set of homogeneous background scalar fields, treated as a dynamical system, and the evolution of physical, long wavelength modes. I show that in general we expect several physical modes which cannot be gauged away, and whose evolution depends on the behaviour of the background system. In parametric resonance the resonance can be seen as the instability of a periodic orbit in the background phase space. Finally I demonstrate that another type of background instability, dynamical chaos, can similarly lead to the rapid growth of long wavelength modes.

Contents

List of Tables

List of Figures

Notation

Tensor indices

Lower-case Greek $(\alpha, \beta, \ldots, \mu, \nu, \ldots)$ run from 0 to 3 (spacetime tensors).
Lower-case Latin $(a, b, \ldots, i, j, \ldots)$ run from 1 to 3 (spatial tensors).
$T_{(\mu\nu)} \equiv 1/2(T_{\mu\nu} + T_{\nu\mu})$ symmetrized tensor
$T_{[\mu\nu]} \equiv 1/2(T_{\mu\nu} - T_{\nu\mu})$ antisymmetrized tensor

Sign conventions

The metric signature is $(-,+,+,+)$.
The Riemann and Einstein tensors follow the sign convention of Misner *et al.* [16], so the Einstein equation reads

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \equiv G_{\mu\nu} = 8\pi G T_{\mu\nu}.$$

Coordinates

$x^\mu = (x^0, x^i)$	coordinate four-vector
t	coordinate proper time
τ	exact proper time
$\eta = \int dt/a$	coordinate conformal time
x	rescaled conformal time [Chapter 4 only; see Eq. (4.27)]

Metric

${}^0 g_{\mu\nu}$	homogeneous background metric of the four-dimensional spacetime
γ_{ij}	homogeneous background metric of a constant-curvature spatial hypersurface
$g_{\mu\nu}$	exact space-time metric
$\delta g_{\mu\nu}$	space-time metric perturbation
$h_{\mu\nu}$	spatial projection tensor [see Eq. (3.63)]

$g = -\det g_{\mu\nu}$

Derivatives

$T_{;\mu}$	covariant derivative with respect to $g_{\mu\nu}$
$\Box T = T^{;\mu}{}_{\mu}$	covariant d'Alembertian
$T_{\|i}$	covariant derivative with respect to γ_{ij}
$\nabla^2 T = T^{\|i}{}_{i}$	covariant Laplacian
$T_{,\mu} = \partial T/\partial x^\mu$	partial derivative with respect to x^μ
$\dot{T} = \partial T/\partial t$	proper-time derivative
$T' = \partial T/\partial \eta$	conformal time derivative
$V_{,\varphi_A} = \partial V/\partial \varphi_A$	partial derivative with respect to scalar field φ_A

Superscripts and subscripts

${}^0T(t)$	homogeneous background quantity (superscript dropped when unnecessary)
$T^{(\mathrm{s})}_{\mu\nu}$	part of a tensor derived from spatial scalar quantities
$T^{(\mathrm{v})}_{\mu\nu}$	part of a tensor derived from spatial transverse vector quantities
$T^{(\mathrm{t})}_{\mu\nu}$	part of a tensor derived from spatial transverse and traceless tensor quantities
${}^{(3)}R$	geometrical quantity calculated for a three-dimensional metric
v_i^{tr}	transverse spatial vector
Σ_t	constant-coordinate-time hypersurface
Σ_r	hypersurface on which $r = 0$ (or $\delta r = 0$)
p_r	matter or metric perturbation variable p in gauge $r = 0$ (or $\delta r = 0$)

Homogeneous background variables

a	scale factor
$H = \dot{a}/a$	Hubble parameter
$\mathcal{H} = a'/a$	conformal-time Hubble parameter
$\mathcal{K} = -1, 0, 1$	spatial curvature signature
Λ	explicit cosmological constant
ρ	energy density in comoving frame
P	pressure in comoving frame
φ_A	N-component scalar field ($A = 1, \ldots, N$)
$V(\varphi_A)$	scalar field potential

Perturbed metric variables

ϕ, ψ, B, E	arbitrary gauge scalar metric functions [see Eq. (3.8)]
Φ, Ψ	longitudinal gauge scalar metric functions
ζ_{MFB}	comoving curvature perturbation ψ_q for case $\Pi = 0$; subscript MFB normally dropped [see Eq. (5.21)]
σ	arbitrary gauge shear scalar [see Eq. (3.55)]
$\delta\theta$	arbitrary gauge perturbed expansion [see Eq. (3.53)]
S_i, F_i	vector metric functions [see Eq. (3.9)]
σ_i	shear vector [see Eq. (3.59)]
$h^i{}_j$	tensor metric function [see Eq. (3.10)]

Perturbed matter variables

$\delta\rho$	arbitrary gauge perturbed energy density [see Eq. (3.68)]
δP	arbitrary gauge perturbed pressure [see Eq. (3.68)]
q	arbitrary gauge scalar momentum density [see Eq. (3.71)]
Π	scalar anisotropic stress [see Eq. (3.72)]
v_i	vector momentum density [see Eq. (3.71)]
π_i	vector anisotropic stress [see Eq. (3.72)]
$\tau^i{}_j$	tensor anisotropic stress [see Eq. (3.72)]
$\delta\varphi_A$	arbitrary gauge N-component scalar field perturbation ($A = 1, \ldots, N$)

Units

I set $c = 1$ and define the Planck mass by $m_{\mathrm{P}} = G^{-1/2}$, where G is the gravitational constant.

Frequentia

For convenience I present here a list of frequently referred to results. The equation numbers are as they appear in the text, where all derivations can be found. Components are given in the chart specified by the corresponding metric. All perturbation expressions are presented in arbitrary gauge. The gauge transformations are only given for the quantities that change only under a temporal transformation T, and hence which are the only quantities required to write the dynamical equations.

Homogeneous background

Metric

$$\boxed{ds^2 = -dt^2 + a^2(t)\gamma_{ij}\, dx^i dx^j \qquad (2.51)}$$

Energy-momentum tensor

$$T^0{}_0 = -\rho(t) \qquad (2.57)$$

$$T^0{}_i = 0 \qquad (2.58)$$

$$T^i{}_j = P(t)\delta^i{}_j \qquad (2.59)$$

Energy-momentum conservation

$$\boxed{\dot{\rho} + 3H(\rho + P) = 0 \qquad (2.65)}$$

Einstein's equation

$$\text{0-0:} \qquad H^2 = \frac{8\pi G}{3}\rho - \frac{\mathcal{K}}{a^2} + \frac{\Lambda}{3} \qquad (2.83)$$

$$i\text{-}i\text{:} \quad -2\dot{H} - 3H^2 = 8\pi G P + \frac{\mathcal{K}}{a^2} - \Lambda \qquad (2.84)$$

$$\dot{H} = -4\pi G(\rho + P) + \frac{\mathcal{K}}{a^2} \qquad (2.87)$$

Scalar field energy-momentum tensor

$$\rho = \tfrac{1}{2}\dot{\varphi}\cdot\dot{\varphi} + V(\varphi_A) \qquad (2.128)$$

$$P = \tfrac{1}{2}\dot{\varphi}\cdot\dot{\varphi} - V(\varphi_A) \qquad (2.130)$$

Klein-Gordon equation

$$\ddot{\varphi}_A + 3H\dot{\varphi}_A + V_{,\varphi_A} = 0 \qquad (2.126)$$

Linear scalar perturbations

Metric

$$ds^2 = a^2(\eta)\left\{-(1+2\phi)d\eta^2 + 2B_{|i}dx^i d\eta + \left[(1-2\psi)\gamma_{ij} + 2E_{|ij}\right]dx^i dx^j\right\} \quad (3.8)$$

Energy-momentum tensor

$$\delta T^0_{\ 0} = -\delta\rho \tag{3.68}$$

$$\delta T^0_{\ i} = \frac{q_{|i}}{a} \tag{3.71}$$

$$\delta T^i_{\ j} = \left(\delta P - \frac{1}{3}\nabla^2\Pi\right)\delta^i_{\ j} + \frac{1}{a^2}\Pi^{|i}_{\ \ j} \tag{3.72}$$

Energy-momentum conservation

$$0:\quad \delta\dot\rho + 3H(\delta\rho + \delta P) + (\rho + P)\left(-3\dot\psi + \frac{1}{a^2}\nabla^2\sigma\right) + \frac{1}{a^2}\nabla^2 q = 0 \tag{3.84}$$

$$i:\quad \dot q + 3Hq + (\rho + P)\phi + \delta P + \frac{2}{3a^2}\nabla^2\Pi = 0 \tag{3.91}$$

Gauge transformations

$$\tilde\phi = \phi + \dot T \tag{3.147}$$

$$\tilde\psi = \psi - HT \tag{3.148}$$

$$\tilde\sigma = \sigma + T \tag{3.149}$$

$$\delta\tilde\theta = \delta\theta + \left(3\dot H + \nabla^2/a^2\right)T \tag{3.150}$$

$$\delta\tilde\rho = \delta\rho + \dot\rho T \tag{3.151}$$

$$\delta\widetilde{P} = \delta P + \dot P T \tag{3.152}$$

$$\tilde q = q - (\rho + P)T \tag{3.153}$$

$$\delta\tilde\varphi_A = \delta\varphi_A + \dot\varphi_A T \tag{3.328}$$

Einstein's equation

$$\text{0-0:}\quad 3H(\dot{\psi}+H\phi)-\frac{1}{a^2}\nabla^2(\psi+H\sigma)=-4\pi G\delta\rho \tag{3.279}$$

$$\text{0-}i\text{:}\quad \dot{\psi}+H\phi=-4\pi Gq \tag{3.282}$$

$$i\neq j\text{:}\quad \psi-\phi+\dot{\sigma}+H\sigma=8\pi G\Pi \tag{3.283}$$

$$i\text{-}i\text{:}\quad \ddot{\psi}+H(3\dot{\psi}+\dot{\phi})+(3H^2+2\dot{H})\phi=4\pi G\left(\delta P+\frac{2}{3a^2}\nabla^2\Pi\right) \tag{3.284}$$

Scalar field energy-momentum tensor

$$\delta\rho=\dot{\varphi}\cdot\delta\dot{\varphi}-\dot{\varphi}\cdot\dot{\varphi}\phi+V_{,\varphi}\cdot\delta\varphi \tag{3.324}$$

$$q=-\dot{\varphi}\cdot\delta\varphi \tag{3.325}$$

$$\delta P=\dot{\varphi}\cdot\delta\dot{\varphi}-\dot{\varphi}\cdot\dot{\varphi}\phi-V_{,\varphi}\cdot\delta\varphi \tag{3.326}$$

$$\Pi=0 \tag{3.327}$$

Klein-Gordon equation

$$\Box\delta\varphi_A+\left(\dot{\phi}+3\dot{\psi}-\frac{1}{a^2}\nabla^2\sigma\right)\dot{\varphi}_A-2\phi V_{,\varphi_A}=V_{,\varphi_A\varphi}\cdot\delta\varphi \tag{3.346}$$

Acknowledgements

I would like to thank my supervisor, Douglas Scott, for his eternal encouragement and heroic patience. Also many thanks to Robert Brandenberger for directing me on a large part of this work. As well I thank my supervisory committee, Paul Hickson, Bill Unruh, and Eric Zhitnitsky, for many helpful comments and suggestions. In addition I acknowledge the Natural Sciences and Engineering Research Council and the Li Tze Fong Fellowship for financial support. Finally, endless thanks to Christine and to my family.

"Be persistent and you will win."

—Far East Fortune Cookie Co. Ltd.

Chapter 1

Introduction

In the study of the early universe, an important issue is the origin and early evolution of the small perturbations from homogeneity that eventually formed the exceedingly rich structure we observe today. While the evolution at late times, and in particular after Hubble radius re-entry, is well understood, much remains to be worked out regarding the earlier behaviour. In this context, the theory of inflation provides the most promising scenario to elucidate those early moments. Inflation was originally proposed as a means of resolving several issues with the standard hot big bang theory, which involve the apparently mysterious initial conditions required by that theory. In essence, it appears that the Hubble length at late times is much smaller than the true causal horizon size. Inflation provides a well-defined dynamical process during which the comoving Hubble length decreases, hence resolving the horizon issue and setting the stage for the standard hot big bang.

As an unexpected bonus, it was soon realized that inflation also provided a mechanism for the generation of cosmological perturbations. This involves quantum scalar field fluctuations being stretched outside the Hubble length. While many problems remain, most notably a concrete particle physics realization of inflation, inflationary model building has become an industry, and the predicted spectrum of perturbations is a fundamental tool linking a particular model to observable quantities, such as the Cosmic Microwave Background (CMB) radiation and large-scale structure.

In the simplest single-scalar-field inflationary models, this linkage is very straightforward, as a conserved curvature perturbation exists for long wavelengths which allows inflation-generated perturbations just after Hubble radius exit to be trivially propagated across the vast gulf of time until Hubble re-entry. However, it has long been known that the situation can be very different when two or more scalar fields are present. Then no conservation law exists in general, and therefore the evolution during the entire super-Hubble era must be carefully followed.

One example where such non-trivial evolution can occur is during the period of reheating, which immediately follows inflation and is characterized by an essentially homogeneous scalar field oscillating and decaying into small scale fluctuations which thermalize and serve as the initial state for the hot big bang. It has been realized relatively recently that during this period it is possible, in multi-field models, for long wavelength perturbations to be amplified tremendously through the process of parametric resonance.

The work in this thesis began by addressing the issue of how important this amplification can be in a specific model, once the backreaction of small scale perturbations is taken into account. The work then turned to more general questions. In particular, what are the general conditions under which long wavelength modes can grow? Are there specific kinds of dynamics other than parametric resonance which occur in realistic inflationary models, and which also may violate the conservation law? The practical importance of these questions lies in the fact that the violation of the standard inflationary predictions in any such model can be used to filter out that model as a viable description of the early universe.

This thesis begins with a thorough review of the theoretical background required for the later results. While most of this material is not new, it is presented in a completely original manner, with essentially all of the required results exposed together. It is intended also as a reference to be queried in later work. Emphasis is made throughout on elucidating the physical meaning of the results, and on the important techniques and approximations that allow a tractable treatment of general relativistic perturbations. Chapter 2 reviews the general relativity of a *homogeneous* universe, describing the metric, energy-momentum tensor, and solutions to Einstein's equation, and then describes the shortcomings of the hot big bang theory, before introducing inflation.

Chapter 3 reviews a wide range of material regarding cosmological perturbations. Fundamental techniques are introduced, such as the decomposition of quantities into scalar, vector, and tensor parts. Gauge invariance is discussed at length, emphasizing the physical meaning of the various results. A novel approach to the derivation of the arbitrary-gauge linearized Einstein's equation is made, and results are presented wherever possible in such arbitrary-gauge form.

In Chapter 4 I describe a particular inflationary model which is known to exhibit parametric resonance. I then perform a study of the growth of long wavelength perturbations in this model, and try to determine whether the backreaction of small scale perturbations is sufficient to save the standard inflationary predictions. I conclude that, for certain parameter values, it is not.

Finally, in Chapter 5 I describe in considerable detail general aspects of the evolution of long wavelenth modes. I begin with a discussion of the adiabatic conservation law. I next provide a careful link between the evolution of a set of homogeneous background scalar fields, treated as a dynamical system, and the evolution of physical, long wavelength modes. I show that in general we expect an adiabatic mode which looks locally like a time-translation of the backgrounds, and several physical modes which cannot be gauged away, and whose evolution depends on the behaviour of the background system. I revisit parametric resonance with this picture in mind, according to which, for long wavelengths, the resonance can be seen as the instability of a periodic orbit in the background phase space. Finally I demonstrate that another type of background instability, dynamical chaos, can similarly lead to the rapid growth of long wavelength modes.

In Appendix A, I present some early work in collaboration with Martin White regarding cosmological parameter estimation for the tensor contribution to the CMB. This work is not presented as a chapter because it lies somewhat outside the scope of the main part of the thesis and would require significant further background material.

More thorough, but still concise, descriptions of the contents of each chapter can be found in their opening pages.

The major original contributions in this work include Sections 4.3, 4.4, and 4.5, which were published in [58]. In addition, parts of Sections 5.2 and 5.3 have been submitted for publication. This work appeared in an earlier form as [59]. Appendix A was published in [57]. While the references [58] and [57] were co-authored, in both cases essentially all of the calculations and most of the writing was done by myself. As I mentioned previously, much of the material on cosmological perturbations is presented in a novel manner, and it is hoped that this exposition can serve a pedagogical purpose.

Chapter 2

Homogeneous Cosmological Background

The study of the very early universe is based upon two main foundations: that of Einstein's theory of gravitation and that of the *cosmological principle*, which states that the universe at the largest scales can be well approximated as homogeneous and isotropic. In this chapter, I will review these two foundations. First, in Section 2.1 I will provide a very brief "derivation" of Einstein's field equation and an elementary description of general relativity including the "conservation" of energy-momentum. My goal is to show how the field equation (and indeed all the results to come) follow from a very simple variational principle, and perhaps to motivate that principle somewhat.

Next, in Section 2.2, I will discuss the cosmological principle and review the standard results of the theory of a precisely homogeneous and isotropic universe. I will derive the forms of the metric and the energy-momentum tensor consistent with exact homogeneity, and find that the metric will be constrained to within one free function, the scale factor a, and the energy-momentum tensor to within two, the comoving energy density ρ and pressure P. Then I will write Einstein's equation, which relates a to ρ and P, present exact solutions in special cases, and describe some properties of the solutions, such as the existence of horizons.

Next I will describe some shortcomings of the standard hot big bang model, and point out that each can be resolved with a very early period when the comoving Hubble length decreases. This will lead directly into an elementary description of scalar fields and the homogeneous dynamics of inflation in Section 2.3.

Most of the material in this chapter is review and can be found in standard texts [16–20].

2.1 Theoretical background

2.1.1 Gravitational action and Eintsein's field equation

In cosmology, as in much of physics, we are interested in the dynamical evolution of various fields on a four-dimensional spacetime manifold. In particular, we are interested in the dynamics of the metric $g_{\mu\nu}$ and any other fields F which represent matter, given some specified conditions. Perhaps the simplest way to attempt to specify the dynamics is to specify $g_{\mu\nu}$ and F on a three-dimensional hypersurface Σ surrounding some four-dimensional spacetime volume in which we wish to determine the evolution (*e.g.* a "slab" between two spacelike hypersurfaces $t = t_1, t_2$ and extending to spatial infinity). If we want our dynamical equations to be generally covariant, *i.e.* invariant under general coordinate transformations $x_\mu \rightarrow \tilde{x}_\mu$, we can begin with a scalar statement of the dynamics. To do this we must first construct the action, S, a scalar-valued functional which maps configurations of $g_{\mu\nu}$ and F inside Σ to the real numbers. The simplest way to then single out one field configuration is to require that it correspond to an extremum of the action, *i.e.*

$$\delta S = 0 \tag{2.1}$$

(subject to $\delta g_{\mu\nu} = \delta F = 0$ on Σ) for the actual dynamical configuration. This configuration will entail the classical evolution of the matter-gravitation system.

The problem then becomes one of finding an action that describes the observed physical world. To start, an action $S_{\rm g}$ that describes the gravitational dynamics in the absence of matter must be constructed from $g_{\mu\nu}$ alone. To do this we can use the Ricci scalar,

$$R = R^{\mu}{}_{\mu}, \tag{2.2}$$

where the Ricci tensor $R_{\mu\nu}$ is a contraction of the Riemann tensor,

$$R_{\mu\nu} = R^{\lambda}{}_{\mu\lambda\nu}. \tag{2.3}$$

The Riemann tensor is related to the connection coefficients $\Gamma^{\mu}_{\nu\lambda}$ by

$$R^{\mu}{}_{\nu\lambda\rho} = 2\Gamma^{\mu}_{\nu[\rho,\lambda]} + 2\Gamma^{\mu}_{\kappa[\lambda}\Gamma^{\kappa}_{\rho]\nu}, \tag{2.4}$$

while the $\Gamma^{\mu}_{\nu\lambda}$ are finally expressed in terms of the metric as

$$\Gamma^{\mu}_{\nu\lambda} = g^{\mu\rho}\left(g_{\rho(\nu,\lambda)} - \frac{1}{2}g_{\nu\lambda,\rho}\right). \tag{2.5}$$

The simplest choice for the gravitational action is the Hilbert action,

$$S_{\mathrm{g}} = \frac{1}{16\pi G}\int R\sqrt{g}\,d^4x, \tag{2.6}$$

where the integral is over the volume bounded by Σ. Also, $g \equiv -\det g_{\mu\nu}$, so that the volume element $\sqrt{g}\ d^4x$ is invariant and thus S_{g} is a manifestly invariant scalar. The gravitational constant G gives the action units of angular momentum, and hence mass dimension zero. This (or any) action may be written in terms of the *Lagrangian density* $\mathcal{L}$, which is defined by

$$S \equiv \int \mathcal{L}\,d^4x. \tag{2.7}$$

Thus the gravitational Lagrangian density is

$$\mathcal{L}_{\mathrm{g}} = \frac{R\sqrt{g}}{16\pi G}. \tag{2.8}$$

Now I will review the derivation of Einstein's equation from the Hilbert action. First define the coefficient in the variation of S_{g} with respect to $g_{\mu\nu}$ to be the Einstein tensor, $G_{\mu\nu}$, *i.e.*

$$\delta S_{\mathrm{g}} \equiv -\frac{1}{16\pi G}\int G^{\mu\nu}\delta g_{\mu\nu}\sqrt{g}\,d^4x. \tag{2.9}$$

To explicitly calculate $G_{\mu\nu}$ we need to evaluate three terms,

$$\delta(\sqrt{g}g^{\mu\nu}R_{\mu\nu}) = \sqrt{g}R_{\mu\nu}\delta(g^{\mu\nu}) + R\delta\sqrt{g} + \sqrt{g}g^{\mu\nu}\delta R_{\mu\nu}. \tag{2.10}$$

The third term can, after some work, be written as an ordinary divergence,

$$\sqrt{g}g^{\mu\nu}\delta R_{\mu\nu} = 2\left(\sqrt{g}g^{\mu\nu}\delta\Gamma^{\lambda}_{\mu[\nu}\right)_{,\lambda]}. \tag{2.11}$$

Thus by Gauss's theorem we may transform this term into a surface integral over Σ which promptly vanishes by virtue of the boundary conditions. Next, for the first term in (2.10), note that

$$0 = \delta(g^{\mu\nu}g_{\nu\lambda}) = \delta(g^{\mu\nu})g_{\nu\lambda} + g^{\mu\nu}\delta g_{\nu\lambda} \tag{2.12}$$

so that

$$\delta(g^{\mu\nu}) = -g^{\mu\rho}g^{\lambda\nu}\delta g_{\rho\lambda}. \tag{2.13}$$

Also, we can calculate for the second term in (2.10)

$$\delta\sqrt{g} = \frac{1}{2}\sqrt{g}g^{\mu\nu}\delta g_{\mu\nu}. \tag{2.14}$$

Combining these results and reading off the coefficient of $\delta g_{\mu\nu}$ we finally obtain

$$G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R. \tag{2.15}$$

In order to complete this "derivation" of Einstein's equation, I will need to discuss the matter action S_{m} and construct the energy-momentum tensor. The total action for the system $(g_{\mu\nu}, F)$ is the sum

$$S = S_{\mathrm{g}} + S_{\mathrm{m}}. \tag{2.16}$$

S_{m} will be a functional of the fields F as well as $g_{\mu\nu}$, expressing the fact that matter couples to gravity. Thus when we determine the dynamics of the combined system through Eq. (2.1), the variation δS_{m} must involve variations in both $g_{\mu\nu}$ and F. However, the part of δS_{m} involving variations in F must be zero, since this condition determines the matter dynamics in the presence of a given gravitational background. Thus we have

$$\delta S_{\mathrm{g}} + \delta S_{\mathrm{m}} = 0 \tag{2.17}$$

subject to $\delta g_{\mu\nu} = 0$ on Σ and $\delta F = 0$ everywhere. Now define the coefficient of the variation of δS_{m} with respect to $g_{\mu\nu}$ to be the energy-momentum tensor $T_{\mu\nu}$, *i.e.*

$$\delta S_{\mathrm{m}} \equiv \frac{1}{2}\int T^{\mu\nu}\delta g_{\mu\nu}\sqrt{g}\, d^4x. \tag{2.18}$$

Note that any antisymmetric part to $T_{\mu\nu}$ will not contribute to δS_{m} by the symmetry of $\delta g_{\mu\nu}$, so we can always take the energy-momentum tensor to be symmetric (and similarly for $G_{\mu\nu}$). Finally, inserting the definitions (2.9) and (2.18) into Eq. (2.17), and allowing arbitrary variations $\delta g_{\mu\nu}$ inside Σ, we obtain Einstein's field equation,

$$G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R = 8\pi G T^{\mu\nu}. \tag{2.19}$$

Note that we can modify the gravitational dynamics in a simple way by adding a constant term to the Ricci scalar in S_{g}, *i.e.*

$$S_{\mathrm{g}} = \frac{1}{16\pi G}\int (R - 2\Lambda)\sqrt{g}\, d^4x. \tag{2.20}$$

Then we have the additional term

$$\delta(-2\Lambda\sqrt{g}) = -2\Lambda\delta\sqrt{g} \tag{2.21}$$

in the variation of S_{g}, which changes the field equation to

$$G^{\mu\nu} + \Lambda g^{\mu\nu} = 8\pi G T^{\mu\nu}. \tag{2.22}$$

Since $G^{\mu\nu}$ involves second order spacetime derivatives, the effect of such a *cosmological constant* Λ will be most noticable at the largest scales. Also notice that the inclusion of the cosmological term is somewhat a matter of taste, since it is always possible to absorb any term $\Lambda g^{\mu\nu}$ into a redefined energy-momentum tensor.

2.1.2 Energy-momentum "conservation"

The action formalism allows us to transparently relate symmetries of the dynamical system to corresponding conserved quantities. For example, gauge invariance in electrodynamics leads to conservation of charge. In a closely analogous way, the general covariance of Einstein's gravitational theory leads to a covariant law which generalizes the special relativistic conservation of the energy-momentum tensor.

To demonstrate that $T_{\mu\nu}$ satisfies such a covariant law, consider a special type of variation of $g_{\mu\nu}$, namely that due to a coordinate change

$$x^\mu \rightarrow \tilde{x}^\mu = x^\mu - \xi^\mu, \tag{2.23}$$

where $\xi^\mu = 0$ on Σ. In this case $g_{\mu\nu}$ will change, according to the tensor transformation law, as

$$g_{\mu\nu}(x) \rightarrow \tilde{g}_{\mu\nu}(\tilde{x}) = g_{\lambda\rho}(x)\frac{\partial x^\lambda}{\partial \tilde{x}^\mu}\frac{\partial x^\rho}{\partial \tilde{x}^\nu}. \tag{2.24}$$

The general covariance of the theory implies that S_{m} will not change under the replacements (2.23) and (2.24) (recall that S_{m} does not change under $F \rightarrow \widetilde{F}$ by the matter equation of motion). However, the coordinate $\tilde{x}$ is a dummy integration variable in S_{m}, so S_{m} also does not change under

$$g_{\mu\nu}(x) \rightarrow \tilde{g}_{\mu\nu}(x). \tag{2.25}$$

For an infinitesimal transformation ξ^μ, this change in $g_{\mu\nu}$ becomes a Lie derivative (see Section 3.4.1),

$$\delta g_{\mu\nu} = \tilde{g}_{\mu\nu}(x) - g_{\mu\nu}(x) = \mathcal{L}_\xi g_{\mu\nu} = \xi_{(\mu;\nu)}. \tag{2.26}$$

Thus, according to the definition of $T_{\mu\nu}$, Eq. (2.18),

$$0 = \delta S_{\mathrm{m}} = \int T^{\mu\nu}\xi_{(\mu;\nu)}\sqrt{g}\, d^4x \tag{2.27}$$

$$= \int T^{\mu\nu}\xi_{\mu;\nu}\sqrt{g}\, d^4x \tag{2.28}$$

$$= \int (T^{\mu\nu}\xi_\mu)_{;\nu}\sqrt{g}\, d^4x - \int T^{\mu\nu}{}_{;\nu}\xi_\mu\sqrt{g}\, d^4x. \tag{2.29}$$

With the aid of the identity

$$V^{\mu}{}_{;\mu} = \frac{1}{\sqrt{g}}(\sqrt{g}V^{\mu})_{,\mu} \tag{2.30}$$

we can write

$$\int (T^{\mu\nu}\xi_{\mu})_{;\nu}\sqrt{g}\, d^4x = \int (\sqrt{g}T^{\mu\nu}\xi_{\mu})_{,\nu}\, d^4x. \tag{2.31}$$

Thus the first term in (2.29) can be rewritten as a surface integral over Σ which vanishes, leaving

$$\int T^{\mu\nu}{}_{;\nu}\xi_{\mu}\sqrt{g}\, d^4x = 0. \tag{2.32}$$

Then, by the arbitrariness of ξ^{μ}, we obtain the covariant law

$$T^{\mu\nu}{}_{;\nu} = 0. \tag{2.33}$$

Note that by a completely analogous calculation we obtain

$$G^{\mu\nu}{}_{;\nu} = 0, \tag{2.34}$$

so that the contracted Bianchi identity can also be seen as a consequence of general covariance.

We can obtain insight into the interpretation of the various components of $T_{\mu\nu}$ in a particular chart by considering the special case of Cartesian coordinates $x^{\mu} = (t, x^i)$ in flat or Minkowsky spacetime (I label spatial components with latin indices). If we constrain ξ^{μ} to lie along a (constant) timelike direction, but let it be otherwise arbitrary, *i.e.* $\xi^{\mu} = (\xi^0, 0)$, we conclude from Eq. (2.32) that

$$T^{0\nu}{}_{,\nu} = \frac{\partial T^{00}}{\partial t} + \frac{\partial T^{0i}}{\partial x^i} = 0. \tag{2.35}$$

This implies via Gauss's theorem that

$$\oint_{\Sigma} T^{0\nu} dS_{\nu} = 0, \tag{2.36}$$

where the integral is over an arbitrary closed hypersurface Σ. In particular, if we let Σ enclose the slab between $t = t_1$ and $t = t_2$ and extending to spatial infinity, and if $T_{\mu\nu}$ vanishes at infinity, then

$$\int T^{00} d^3x = \text{const.} \tag{2.37}$$

In words, time-translation invariance implies energy conservation, with T^{00} the energy density and T^{0i} the energy flux vector. Similary, for $\xi^{\mu} = (0, \xi^i)$, we obtain

$$T^{i\nu}{}_{,\nu} = 0, \tag{2.38}$$

which is the statement of momentum conservation. Again, T^{i0} is the density of the ith component of momentum and T^{ij} is the momentum flux tensor, also called the stress tensor. Thus we recover the familiar connection between translational symmetries and energy-momentum conservation in Minkowsky spacetime.

Note, however, that the covariant "conservation law", Eq. (2.33), does not in general imply the existence of conserved quantities in the sense of Eq. (2.37). This is because the tensor identity

$$T^{\mu\nu}{}_{;\mu} = \frac{1}{\sqrt{g}}(\sqrt{g}T^{\mu\nu})_{,\mu} + \Gamma^{\nu}_{\mu\lambda}T^{\mu\lambda} \tag{2.39}$$

implies instead that

$$\oint_{\Sigma} \sqrt{g}T^{\mu\nu}dS_{\mu} = -\int \Gamma^{\nu}_{\mu\lambda}T^{\mu\lambda}\sqrt{g}\,d^4x, \tag{2.40}$$

and we cannot in general set the connection coeficient in the "source term" on the rhs to zero everywhere. Nevertheless, in special cases where the spacetime exhibits symmetries it may still be possible to write a strict conservation law. For example, if we can find a vector field u^{μ} for which $u_{(\mu;\nu)} = 0$ (which implies an isometry of the spacetime; see the following section), then

$$(T^{\mu\nu}u_{\nu})_{;\mu} = T^{\mu\nu}{}_{;\mu}u_{\nu} = 0, \tag{2.41}$$

which, with the identity (2.30), implies that the vector $T^{\mu\nu}u_{\nu}$ is strictly conserved. Also, by the equivalence principle, we can always consider small enough regions of spacetime that are as near as we wish to Minkowsky, so conservation will be enforced locally in such regions. On the other hand, techniques exist to construct effective energy-momentum pseudotensors which attempt to take into account the contribution to energy and momentum of the gravitational field itself [16, 18, 19]. However, this cannot be done unambiguously, because it is impossible to define a unique covariant *local* gravitational energy-momentum density.

2.2 Exactly homogeneous and isotropic spacetimes

2.2.1 The cosmological principle

The study of the early universe is based upon the foundations of general relativity, described briefly in the previous section, and the *cosmological principle*. This principle states that on sufficiently large scales, the universe is spatially homogeneous and isotropic, except for small perturbations which can be treated with linear perturbation theory. It is strongly motivated by galaxy surveys and measurements of the CMB. A crucial part of the cosmological principle is the tenet of antianthropocentrism, which states that we are in a typical place in the universe (at least insofar as large-scale structure is concerned). Thus while galaxy surveys, for example, only reach so far, accepting the cosmological principle means assuming that the poorly or completely unobserved regions are largely the same as ours. A theoretical basis for better understanding this homogeneity will be provided by the inflationary scenario.

Motivated by the cosmological principle, we can decompose the metric and all matter fields into precisely homogeneous and isotropic *backgrounds* and arbitrary but "small" perturbations. Throughout this thesis, background quantities will be indicated by a superscript ^{0}, perturbations with a δ, and exact quantities will be unadorned, unless otherwise noted. Thus, *e.g.*, a scalar field F is decomposed as

$$F(x^\mu) = {}^0F(t) + \delta F(x^\mu). \tag{2.42}$$

For the remainder of this chapter, I will consider only the exactly homogeneous backgrounds, and thus I will drop the background superscript ^{0}. Perturbations will be the subject of Chapter 3.

In order to state the cosmological principle in the language of general relativity, I will introduce two tensor properties. An *isometry* of spacetime is defined as a coordinate transformation $x^\mu \rightarrow \tilde{x}^\mu$ for which

$$g_{\mu\nu}(x) = \tilde{g}_{\mu\nu}(x). \tag{2.43}$$

Similarly, any scalar, vector, or tensor field F representing matter is said to be *form invariant* if

$$F(x) = \widetilde{F}(x). \tag{2.44}$$

(Both of these conditions can be expressed by the vanishing of the Lie derivative, $\mathcal{L}_\xi g_{\mu\nu} = 0$ and $\mathcal{L}_\xi F = 0$, where $\xi^\mu = x^\mu - \tilde{x}^\mu$; see Section 3.4.1.)

Homogeneity implies that there exists a *foliation* of spacetime, *i.e.* a choice of the time coordinate t, for which the constant-time hypersurfaces Σ_t are homogeneous. That is, there exists a three-parameter family of spatial translational isometries of the metric tensor on the Σ_t, and all matter fields F are form invariant under the same family of translations.

Isotropy means that there exists a timelike vector field u^μ such that any observer with worldline tangent to u^μ can observe no preferred direction. Thus there is a three-parameter family of spatial rotations that are isometries of $g_{\mu\nu}$ and render all matter fields form invariant. Symmetry implies that the vector field u^μ is everywhere normal to the Σ_t. The field u^μ defines the *comoving* reference frame: if the matter is an ordinary fluid, the comoving frame coincides with the rest frame of the fluid.

To summarize, when the cosmological principle is satisfied exactly, then at a fixed time the metric tensor and all matter fields have the same functional form regardless of where we look or how we orient our spatial coordinates.

It is very natural in a homogeneous and isotropic spacetime to decompose tensors into space and time components. In doing this, an obvious choice for the time coordinate t is one that foliates spacetime into homogeneous hypersurfaces Σ_t. Lines of constant spatial coordinate x^i can be chosen to be integral curves of the comoving vector field u^μ. Unless I state otherwise I will always make these choices. It will be very important to understand the transformation properties of these space and time components under the restricted class of spatial coordinate transformations on the Σ_t,

$$x^0 \to \tilde{x}^0 = x^0, \qquad x^i \to \tilde{x}^i = \tilde{x}^i(x^j). \tag{2.45}$$

From the general transformation properties of a second-rank tensor $S_{\mu\nu}$,

$$\widetilde{S}_{\mu\nu}(\tilde{x}^\rho) = \frac{\partial x^\kappa}{\partial \tilde{x}^\mu}\frac{\partial x^\lambda}{\partial \tilde{x}^\nu} S_{\kappa\lambda}(x^\rho), \tag{2.46}$$

it is clear that under (2.45) S_{00} transforms as a scalar while S_{0i} and S_{ij} transform as three-vectors and second-rank three-tensors, respectively. (For example,

$$\widetilde{S}_{0i}(\tilde{x}) = \frac{\partial x^\kappa}{\partial \tilde{x}^0}\frac{\partial x^\lambda}{\partial \tilde{x}^i} S_{\kappa\lambda}(x) = \frac{\partial x^j}{\partial \tilde{x}^i} S_{0j}(x), \tag{2.47}$$

so that S_{0i} transforms like a spatial vector.) Similarly, the components V_0 and V_i of a four-vector V_μ transform like spatial scalars and vectors, respectively. Thus in a homogeneous and isotropic background, the rotational isometry (or form invariance) implies that all spatial vectors S_{0i} and V_i must vanish in a comoving frame, since the zero vector is the only vector that is invariant with respect to arbitrary rotations. Conversely, if such a vector were not zero, it would define a prefered direction. Similarly, a spatial tensor S_{ij} can only be constructed from the "generic" tensors g_{ij} and ϵ_{ijk}. Also, homogeneity implies that any spatial scalar must depend only on the time.

2.2.2 The Friedmann-Robertson-Walker metric

The high degree of symmetry in a spacetime satisfying the cosmological principle greatly simplifies its treatment. In fact, these symmetries alone will enable us to determine to a great extent the form of the metric, without any use of Einstein's field equation.

To begin our determination of the metric, define t such that $t_2 - t_1$ is the lapse of proper time between hypersurfaces Σ_{t_1} and Σ_{t_2} in a comoving frame, so that $g_{00} = -1$. Isotropy implies that the spatial vector g_{0i} must vanish in the comoving frame, by the argument of the preceeding subsection. Thus in the comoving frame the line element has the form

$$ds^2 = -dt^2 + g_{ij}(x^\mu)\, dx^i dx^j. \tag{2.48}$$

Note that this form of the metric also follows from the orthogonality of the field u^μ to the Σ_t. For, if X^μ is an arbitrary vector in a hypersurface Σ_t, then

$$0 = X_\mu u^\mu = g_{0i} X^i u^0 \tag{2.49}$$

for all X^i implies $g_{0i} = 0$ in a chart in which $u^\mu = (u^0, 0)$, *i.e.* in a comoving chart. A metric of the form (2.48), with $g_{00} = -1$ and $g_{0i} = 0$, defines a *synchronous* coordinate system. The name derives from the fact that in such a system it is possible to globally synchronize clocks [18]. In the present case the hypersurfaces Σ_t provide that synchronization.

Next, consider the time dependence in the spatial metric g_{ij}. To be compatible with the homogeneity and isotropy of the Σ_t for all time, g_{ij} must involve only an overall scale factor $a(t)$, so that

$$g_{ij}(x^\mu) = a^2(t)\gamma_{ij}(x^k). \tag{2.50}$$

That is, any time dependence apart from a uniform expansion or contraction will distort the spatial geometry and destroy the symmetries. Thus the full metric is

$$ds^2 = -dt^2 + a^2(t)\gamma_{ij}(x^k)\, dx^i dx^j. \tag{2.51}$$

The form of the constant spatial metric γ_{ij} is very tightly constrained by the six independent translational and rotational isometries. In fact, the Σ_t are "maximally symmetric" constant-curvature spaces, and can take only three distinct forms, depending on whether the curvature is positive, zero, or negative. Although I will not need the explicit form of the spatial metric, I will state it here. We can choose spherical coordinates such that [19]

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1 - \mathcal{K}r^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right], \tag{2.52}$$

where $\mathcal{K}$ can take the values 1, 0, or -1. For $\mathcal{K} = 1$, each Σ_t is a space of positive curvature, namely the closed three-sphere. For $\mathcal{K} = 0$, the Σ_t are Euclidean. Finally, for $\mathcal{K} = -1$, each Σ_t is an open space of negative curvature. The metric (2.52) is known as the Friedmann-Robertson-Walker (FRW) metric. In the case $\mathcal{K} = 0$, if we choose Cartesian coordinates, γ_{ij} is simply the unit matrix. Also in this case, the normalization of $a(t)$ is arbitrary, and any physical prediction must depend only on ratios of the scale factor.

We can put the metric and upcoming calculations into a more symmetrical form by defining the *conformal time* η through

$$a(t)\,d\eta = dt. \tag{2.53}$$

Then the metric becomes

$$ds^2 = a^2(\eta)(-d\eta^2 + \gamma_{ij}(x^k)\,dx^i dx^j). \tag{2.54}$$

This form of the metric is very useful for elucidating the causal structure of the spacetime, as we will see in Section 2.2.6.

Recall that the spatial coordinates x^i in the FRW metric written in the form (2.51) or (2.54) are comoving coordinates, *i.e.* worldlines of constant x^i are the trajectories of observers to whom the universe appears isotropic. As I mentioned in Section 2.2.1, for an ordinary fluid this means the comoving observers are at rest relative to the fluid. At late times, for example, the comoving coordinates of galaxies are (roughly) constant. Often, on the other hand, it is useful to discuss the *physical* distances between events. The symmetry of the FRW metric makes it unambiguous and easy to write the physical separation between two events on the same constant-time hypersurface Σ_t. Namely, if two such events are separated by comoving distance Δr along the x^i direction, then their physical separation Δl is

$$\Delta l(t) = \int a(t)\,dx^i = a(t)\int dx^i = a(t)\Delta r. \tag{2.55}$$

(Of course no such simple relation connects the proper- and conformal-time separations between events—instead Eq. (2.53) must be integrated.) The proper time derivative of the physical separation is

$$\frac{d\Delta l}{dt} = \dot{a}(t)\Delta r \equiv H(t)\Delta l(t), \tag{2.56}$$

where $\dot{a} \equiv da/dt$ and $H(t) \equiv \dot{a}/a$ is the *Hubble parameter.* Eq. (2.56) is the famous Hubble's Law, expressing the linear relationship between recession speed and physical distances.

2.2.3 Energy-momentum tensor and conservation

Just as the cosmological principle enabled us to determine the form of the metric up to a single arbitrary function $a(\eta)$, the background symmetries will allow us to deduce much about the form of the energy-momentum tensor and its conservation law before we apply Einstein's equation. First, choose comoving synchronous coordinates as specified by the FRW metric (2.51). Homogeneity implies that T_{00} depends only on time. The arguments below Eq. (2.47) imply that $T_{0i} = 0$ and that T_{ij} is a time-dependent multiple of g_{ij}. That is,

$$T_{00} \equiv \rho(t), \tag{2.57}$$

$$T_{0i} = 0, \tag{2.58}$$

$$T_{ij} \equiv P(t) g_{ij}. \tag{2.59}$$

We can combine these components into the unified tensor

$$T_{\mu\nu} = \rho(t) u_\mu u_\nu + P(t) h_{\mu\nu}, \tag{2.60}$$

where

$$h_{\mu\nu} \equiv g_{\mu\nu} + u_\mu u_\nu \tag{2.61}$$

projects orthogonal to u^μ, which is the timelike unit vector field which defines the comoving frame. From the discussion of the energy-momentum tensor in Section 2.1, we can interpret ρ as the energy density in the comoving frame, and P as the "isotropic stress", *i.e.* the pressure, in the same frame. Note that in order that expression (2.60) for $T_{\mu\nu}$ be covariant, ρ and P must be four-scalars. Thus ρ and P are defined to be scalar fields which happen to take on the values of the energy density and the pressure at each event in the comoving frame.

Expression (2.60) is in precisely the form of the energy-momentum tensor for a homogeneous *perfect fluid.* A perfect fluid is defined by the condition that at each point in spacetime there exists a reference frame (the comoving frame) such that the matter in the neighbourhood of the point appears isotropic. For an ordinary fluid this will be the case if mean free paths are much shorter than the distances over which the fluid parameters vary significantly, and viscosity and heat dissipation are negligible.

Recall from Section 2.1 that in general the law $T^{\mu\nu}{}_{;\mu} = 0$ does not imply a strict conservation law. However, the symmetries of a homogeneous and isotropic spacetime will enable us to write a conservation equation. I will use the comoving synchronous chart defined by the FRW metric (2.51). The components $T^{\mu i}{}_{;\mu}$ must be zero because they form a spatial vector. That is, momentum conservation is trivially satisfied in the comoving chart. To calculate the time component, we first have from Eq. (2.60)

$$T^{\mu\nu}{}_{;\mu} = (\rho + P)_{,\mu} u^\mu u^\nu + (\rho + P)\left(u^\mu{}_{;\mu} u^\nu + u^\mu u^\nu{}_{;\mu}\right) + P_{,\mu} g^{\mu\nu} = 0. \tag{2.62}$$

Thus the component parallel to u_ν is

$$u_\nu T^{\mu\nu}{}_{;\mu} = -\rho_{,\mu} u^\mu - (\rho + P) u^\mu{}_{;\mu} = 0. \tag{2.63}$$

However, $\rho_{,\mu}u^\mu = \partial\rho/\partial t \equiv \dot{\rho}$, and

$$u^\mu{}_{;\mu} = \Gamma^\mu_{\mu\lambda}u^\lambda = \Gamma^\mu_{\mu 0} = 3H \tag{2.64}$$

[see Eq. (2.72) in the next section]. Thus we have finally

$$\dot{\rho} = -3H(\rho + P). \tag{2.65}$$

This energy conservation equation can be rewritten

$$\frac{d}{dt}(\rho a^3) = -P\frac{da^3}{dt}, \tag{2.66}$$

which amounts to the first law of thermodynamics for the energy in a volume proportional to a^3, under conditions of constant entropy (adiabatic expansion). The simple energy conservation equation (2.65) is very powerful, since it is independent of the presence of a cosmological term Λ or spatial curvature $\mathcal{K}$.

Further progress requires specifying a relationship between ρ and P, namely the equation of state. If we define the parameter w, also called the equation of state, by

$$w \equiv \frac{P}{\rho}, \tag{2.67}$$

then, in the case $w =$ const, we can readily integrate the energy conservation equation to obtain

$$\rho a^{3(w+1)} = \text{const.} \tag{2.68}$$

The value of w (if it is even a constant) must be determined by the particular form of the energy-momentum tensor derived from the matter action via Eq. (2.18). However, I can summarize three important special cases here. The case $w = 0$, which corresponds to pressureless "dust", or a "matter-dominated" universe, gives

$$\rho a^3 = \text{const.} \tag{2.69}$$

In words, the energy within a given comoving volume is constant. For the radiation-dominated equation of state, $w = 1/3$,

$$\rho a^4 = \text{const.} \tag{2.70}$$

This expresses the constancy of the number of relativistic particles within a comoving volume. Finally, for $w = -1$,

$$\dot{\rho} = 0. \tag{2.71}$$

For such an equation of state, the energy density does not decay, as the negative pressure does work on a volume as it expands. Note that had we absorbed a cosmological constant term $\Lambda g^{\mu\nu}$ into the energy-momentum tensor, then in vacuum we would have $\rho = \Lambda/8\pi G$ and $P = -\Lambda/8\pi G$, so that $w = -1$. Thus this equation of state is referred to as a cosmological-constant- or vacuum-dominated equation of state.

2.2.4 Einstein's equation

I have now derived the general form of the metric imposed by the cosmological principle, namely the FRW metric (2.51) or (2.54). Also, I have determined the form of the energy-momentum tensor consistent with homogeneity and isotropy, namely the homogeneous perfect fluid form (2.60). It is now time, finally, to apply Einstein's field equation (2.22) to the FRW metric in the presence of the cosmological $T_{\mu\nu}$ and thus determine the dynamical evolution of the spacetime. In order to proceed, we first need to calculate the connection coefficients using their definition, Eq. (2.5). Very straightforward calculations give

$$\Gamma^i_{j0} = H\delta^i_j, \qquad \Gamma^0_{ij} = a^2 H\gamma_{ij}, \qquad \Gamma^i_{jk} = {}^{(3)}\Gamma^i_{jk} \tag{2.72}$$

for the proper-time, comoving chart defined by Eq. (2.51). Using instead conformal time, I obtain

$$\Gamma^0_{00} = \mathcal{H}, \qquad \Gamma^i_{j0} = \mathcal{H}\delta^i_j, \qquad \Gamma^0_{ij} = \mathcal{H}\gamma_{ij}, \qquad \Gamma^i_{jk} = {}^{(3)}\Gamma^i_{jk}, \tag{2.73}$$

where $\mathcal{H} \equiv a'/a$. The space-space-space components are equal to the connection coefficients on the constant-time hypersurfaces,

$${}^{(3)}\Gamma^i_{jk} = \gamma^{il}\left(\gamma_{l(j,k)} - \frac{1}{2}\gamma_{jk,l}\right), \tag{2.74}$$

although I will not need their explicit form. Note though that these purely spatial components vanish for Cartesian coordinates in a spatially flat ($\mathcal{K} = 0$) universe. All Γ's not shown here are either zero or related by the symmetry

$$\Gamma^\lambda_{\mu\nu} = \Gamma^\lambda_{\nu\mu}. \tag{2.75}$$

Next, we can calculate the symmetric Ricci tensor from the contracted version of Eq. (2.4),

$$R_{\mu\nu} = 2\Gamma^\lambda_{\mu[\nu,\lambda]} + 2\Gamma^\lambda_{\kappa[\lambda}\Gamma^\kappa_{\nu]\mu}. \tag{2.76}$$

We know that the vector R_{0i} is zero in a comoving chart. Thus we only need to calculate the two distinct components R_{00} and R_{ij}. From the expression (2.76), it is apparent that R_{ij} can be written as the sum

$$R_{ij} = R_{ij}|_{\mathcal{K}=0} + {}^{(3)}R_{ij}, \tag{2.77}$$

where ${}^{(3)}R_{ij}$ is the spatial Ricci tensor calculated from ${}^{(3)}\Gamma^i_{jk}$ alone. The only tensor that can be used to construct ${}^{(3)}R_{ij}$ on the maximally symmetric hypersurfaces Σ_t is γ_{ij}, and hence we must have

$${}^{(3)}R_{ij} = 2\mathcal{K}\gamma_{ij}, \tag{2.78}$$

where the proportionality constant is determined by consistency with the FRW metric, Eq. (2.52). Thus, using Eq. (2.76) I find the non-zero components of the Ricci tensor to be

$$R_{00} = -3\frac{\ddot{a}}{a}, \qquad R_{ij} = (\ddot{a}a + 2\dot{a}^2 + 2\mathcal{K})\gamma_{ij} \tag{2.79}$$

in the proper-time chart, or

$$R_{00} = -3\mathcal{H}', \qquad R_{ij} = (\mathcal{H}' + 2\mathcal{H}^2 + 2\mathcal{K})\gamma_{ij} \tag{2.80}$$

in conformal time. Contracting the Ricci tensor gives the Ricci scalar,

$$R = R^{\mu}{}_{\mu} = \frac{6}{a^2}\left(\ddot{a}a + \dot{a}^2 + \mathcal{K}\right) = \frac{6}{a^2}\left(\frac{a''}{a} + \mathcal{K}\right). \tag{2.81}$$

Collecting these expressions, together with the energy-momentum tensor (2.60), we can now write the two distinct components of Einstein's equation. (Notice that the forms of the Einstein and energy-momentum tensors we have arrived at *are* consistent—indeed we could have deduced the form of the cosmological $T_{\mu\nu}$ by demanding it equal the FRW-derived $G_{\mu\nu}$.) The time-time component

$$R_{00} - \frac{1}{2}g_{00}R + \Lambda g_{00} = 8\pi G T_{00} \tag{2.82}$$

gives

$$\frac{\dot{a}^2}{a^2} = \frac{a'^2}{a^4} = H^2 = \frac{8\pi G}{3}\rho - \frac{\mathcal{K}}{a^2} + \frac{\Lambda}{3}, \tag{2.83}$$

while the space-space components give

$$-2\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} = -2\frac{a''}{a^3} + \frac{a'^2}{a^4} = 8\pi G P + \frac{\mathcal{K}}{a^2} - \Lambda. \tag{2.84}$$

Eq. (2.83) is sometimes called the Friedmann equation. We can use the time-time equation to eliminate the first time derivative from the space-space equation, giving

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) + \frac{\Lambda}{3} \tag{2.85}$$

or

$$\frac{a''}{a^3} = \frac{4\pi G}{3}(\rho - 3P) - \frac{\mathcal{K}}{a^2} + \frac{2}{3}\Lambda. \tag{2.86}$$

Another useful way to write these equations is

$$\dot{H} = -4\pi G(\rho + P) + \frac{\mathcal{K}}{a^2}. \tag{2.87}$$

These equations are consistent with the covariant energy-momentum conservation law, as the Bianchi identity demands: Taking the time derivative of Eq. (2.83) and substituting Eq. (2.84), we recover the energy conservation equation (2.65).

Note that the time-time component, Eq. (2.83), does not contain second time derivatives. This is a general feature of Einstein's equation—the time-time and time-space components contain only first time derivatives of the metric. This follows from the contracted Bianchi identity, Eq. (2.34), which can be written in components

$$G^{0\mu}{}_{;0} = -G^{i\mu}{}_{;i}. \tag{2.88}$$

The rhs contains at most second time derivatives, hence so does $G^{0\mu}{}_{;0}$. But this expression contains an explicit time derivative, so $G^{0\mu}$ must contain at most first time derivatives. Thus only the space-space components of Einstein's equation are "dynamical", in the sense that they are second-order in time. The time components instead act as constraints on the metric components which must be satisfied when posing an initial-value problem. In the present case, if we fix the energy density ρ at some initial time, the scale factor is constrained by Eq. (2.83).

2.2.5 Solutions of Einstein's equation

Solutions to Einstein's equation in the homogeneous cosmological context are covered in detail in standard texts [16, 19, 20]. Here I will only briefly describe some qualitative properties of the solutions and present explicit solutions in a few special cases.

Perhaps the best way to understand the qualitative behaviour of the various solutions is to reexpress the energy constraint equation (2.83) by multipling it by $(a/a_0)^2$, where a_0 is the value of the scale factor at some standard reference time (*e.g.* today). The result is

$$\left(\frac{\dot{a}}{a_0}\right)^2 + V(a/a_0) = -\frac{\mathcal{K}}{a_0^2}, \tag{2.89}$$

where

$$V(a/a_0) \equiv -\frac{8\pi G\rho}{3}\left(\frac{a}{a_0}\right)^2 - \frac{\Lambda}{3}\left(\frac{a}{a_0}\right)^2. \tag{2.90}$$

Eq. (2.89) looks just like an ordinary energy equation for a mechanical system, with kintic term $(\dot{a}/a_0)^2$, effective potential $V(a/a_0)$, and constant effective energy $-\mathcal{K}/a_0^2$. For matter- or radiation-dominated evolution (with $\rho \propto a^{-3}$

and a^{-4}, respectively), the matter density term in the effective potential dominates at small values of the scale factor, while the Λ term dominates at large values of a/a_0. For $\Lambda = 0$, we see immediately that the system is "bound" for a spatially closed universe ($\mathcal{K} = 1$), and the universe will reach a maximum size and then recontract, while an open universe ($\mathcal{K} = 0$ or -1) is "unbound" and will always expand (or always contract!). For $\Lambda > 0$, it is possible for the universe to begin in an ordinary compact expanding state, make it over the "potential hill", (perhaps with "hesitation"), and then accelerate down the cosmological constant slope at large a/a_0. Similarly, a universe at large a/a_0 may contract, "bounce" off the potential hill, and then return down the slope. If the curvature term is precisely balanced by Λ and ρ, it is possible for a homogeneous universe to be stationary at the unstable fixed point at the top of the potential hill (this is the "Einstein universe").

We can conveniently express the balance between the matter density, curvature, and cosmological constant by rewriting the energy constraint equation (2.83) yet again. Dividing it by H^2, we obtain

$$\Omega_{\rm m} + \Omega_{\mathcal{K}} + \Omega_\Lambda = 1, \tag{2.91}$$

where

$$\Omega_{\rm m} \equiv \frac{\rho}{\rho_{\rm c}}, \qquad \Omega_{\mathcal{K}} \equiv -\frac{\mathcal{K}}{a^2H^2}, \qquad \Omega_\Lambda \equiv \frac{\Lambda}{3H^2}. \tag{2.92}$$

The parameter

$$\rho_{\rm c} \equiv \frac{3H^2}{8\pi G} \tag{2.93}$$

is called the *critical matter density.* This name derives from the fact that if $\Lambda = 0$, then $\rho = \rho_{\rm c}$ implies that $\mathcal{K} = 0$, *i.e.* the universe is spatially flat. Thus the critical density value separates the qualitatively different cosmological solutions of eternal expansion and eventual recollape. In general, allowing nonzero Λ, we have

$$\Omega_{\rm m} + \Omega_\Lambda \begin{cases} < 1 & \text{if } \mathcal{K} < 0 \\ = 1 & \text{if } \mathcal{K} = 0 \\ > 1 & \text{if } \mathcal{K} > 0 \end{cases} \tag{2.94}$$

so that spatial flatness corresponds to $\Omega_{\rm m} + \Omega_\Lambda = 1$.

To close this subsection, I will derive explicit solutions to Einstein's equation in a few special cases. As was the case in solving the energy conservation equation, progress can only be made by specifying a relationship between ρ and P, namely the equation of state. In addition, in solving the dynamics we must also specify the spatial curvature $\mathcal{K}$ and any explicit cosmological constant Λ. I will assume here that the spatial curvature is zero (which is well justified in the real universe, as we will see), and I will consider the case of a positive Λ by deriving the solution for the case $P = -\rho$. Again, the matter action must ultimately determine the equation of state.

The equations of motion can be readily integrated on the assumption of a constant equation of state $w = P/\rho$. In this case, the energy conservation equation implied that $\rho a^{3(w+1)}$ is constant [see Eq. (2.68)]. Thus the energy constraint equation (2.83) implies

$$\dot{a}^2 a^{3w+1} = \frac{8\pi G}{3}\rho_0 a_0^{3(w+1)}, \tag{2.95}$$

where ρ_0 and a_0 are evaluated at some arbitrary reference time. Integrating this equation for the case $w \neq -1$, I obtain the power law solution

$$\frac{a}{a_0} = (\widetilde{w}\tilde{t})^{1/\widetilde{w}}, \tag{2.96}$$

where

$$\widetilde{w} \equiv \frac{3}{2}(w+1) \tag{2.97}$$

and

$$\tilde{t} \equiv \sqrt{\frac{8\pi G}{3}\rho_0}\, t \tag{2.98}$$

is a rescaled dimensionless time coordinate. Note that I have suppressed an additional arbitrary constant which determines the origin of the time coordinate.

For the case $w = 0$, which describes dust, Eq. (2.96) becomes

$$\frac{a}{a_0} = \left(\frac{3}{2}\tilde{t}\right)^{2/3}. \tag{2.99}$$

For radiation, $w = 1/3$, and Eq. (2.86) immediately gives

$$a' = \dot{a}a = \text{const.} \tag{2.100}$$

The explicit solution is, from Eq. (2.96),

$$\frac{a}{a_0} = \sqrt{2\tilde{t}}. \tag{2.101}$$

Finally, for $w = -1$, which, as I discussed at the end of Section 2.2.3, is equivalent to the case of a pure cosmological constant, Eq. (2.95) must be integrated separately. The result is

$$\frac{a}{a_0} = e^{\tilde{t}} = e^{Ht} \tag{2.102}$$

with

$$H = \sqrt{\frac{8\pi G}{3}\rho_0} = \sqrt{\frac{\Lambda}{3}} = \text{const.} \tag{2.103}$$

In this case the metric becomes, in polar comoving coordinates and recalling that $\mathcal{K} = 0$,

$$ds^2 = -dt^2 + a_0^2 e^{2Ht}\left[dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right]. \tag{2.104}$$

This metric describes a spatially flat slicing of *de Sitter spacetime.* This spactime is of fundamental importance in general relativity as it is a maximally symmetric (constant Ricci curvature R) spacetime [21, 22]. It is also possible to write the de Sitter metric using spatially closed or open slicings, or in a static form.

2.2.6 Horizons

A very important property of the homogeneous and isotropic solutions to Einstein's equation is the existence of *horizons.* Horizons delimit regions of spacetime that can have been in causal contact with a particular observer or can ever be in contact with the observer. While in Minkowsky spacetime all regions could have been or eventually can be in causal contact with any observer, this will turn out not to be the case in important cosmological models. In fact, this will lead to a serious problem with the standard hot big bang model.

It will be easiest to discuss horizons using the conformal-time form of the metric, Eq. (2.54). The name "conformal" is indicative of the fact that the metric (2.54), for $\mathcal{K} = 0$, is manifestly a *conformal transformation* of the flat Minkowsky metric, *i.e.* $g_{\mu\nu}^{\text{FRW}} = \Omega^2(x^\lambda) g_{\mu\nu}^{\text{Mink}}$ for $\Omega = a$. The conformal metric is very important because it preserves the causal structure of Minkowsky spacetime. Namely, a null light ray is characterized by $d\eta^2 = dx_i dx^i$, so that in an x^i-η spacetime diagram, the light ray follows a straight diagonal line and all of the usual causal structure of light cones, timelike or spacelike curves, etc. can be simply read off the diagram just as in the flat case.

The conformal time η is related to the proper time and the scale factor by

$$\eta = \int \frac{dt}{a} = \int \frac{da}{\dot{a}a}. \tag{2.105}$$

If this integral converges at its lower limit as $a \longrightarrow 0$, then all of the spacetime will lie above some value $\eta = \eta_0$, and hence will be conformally related to only the portion of Minkowsky spacetime above a constant time hyperplane. Thus a spacetime diagram using η and a comoving spatial coordinate will only consist of the half-plane above the singularity at η_0 (see Fig. 2.1). Clearly an

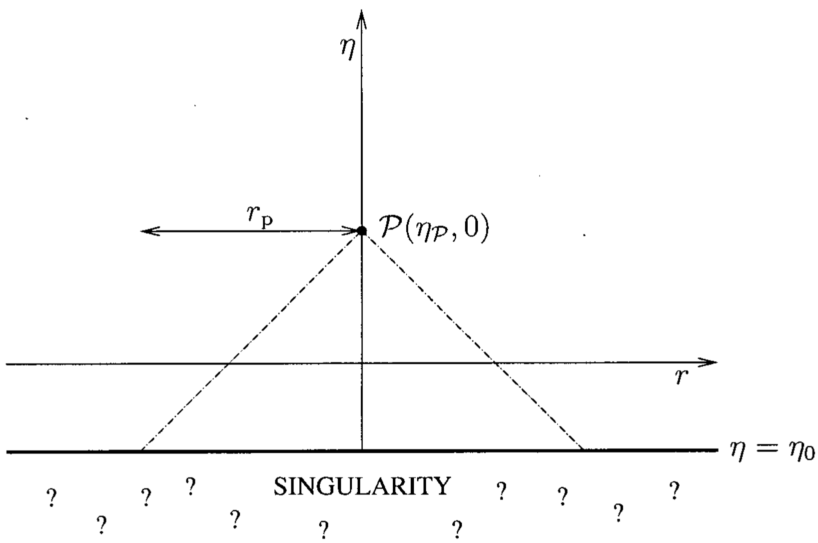

Figure 2.1: A particle horizon exists when the entire spacetime is conformally related to the portion of Minkowsky spacetime above $\eta = \eta_0$. In this case an observer at $\mathcal{P}$ cannot have yet been in causal contact with particles outside of a comoving radius r_p.

observer at some event $\mathcal{P}$ at coordinates $(\eta_\mathcal{P}, 0)$ can only have been in causal contact with comoving particles that have been inside $\mathcal{P}$'s past light cone, *i.e.* particles within a comoving radius r_p of $\mathcal{P}$, given by

$$r_\mathrm{p} = \eta_\mathcal{P} - \eta_0 = \int_0^{a(\eta_\mathcal{P})} \frac{da}{\dot{a}a}. \tag{2.106}$$

Particles outside of this radius will only be able to contact the observer at later times. The radius r_p defines the *particle horizon.* The comoving radius can be translated into a physical distance, namely the spacelike distance between $\mathcal{P}$ and the event at comoving coordinates $(\eta_\mathcal{P}, r_\mathrm{p})$. Using Eq. (2.55), we have

$$l_\mathrm{p} \equiv a(\eta_\mathcal{P}) r_\mathrm{p} \tag{2.107}$$

as the physical distance to the particle horizon.

Similarly, if the integral (2.105) converges at its upper limit as $t \longrightarrow \infty$ (or as $t \longrightarrow t_\mathrm{max}$ if the universe recollapses to a singularity at t_max) then all of the spacetime will be conformally related to the portion of Minkowsky spacetime below a constant time surface at some $\eta = \eta_\mathrm{f}$ (see Fig. 2.2). In this case an observer at $\mathcal{P}$ will clearly never be able to receive a signal from particles

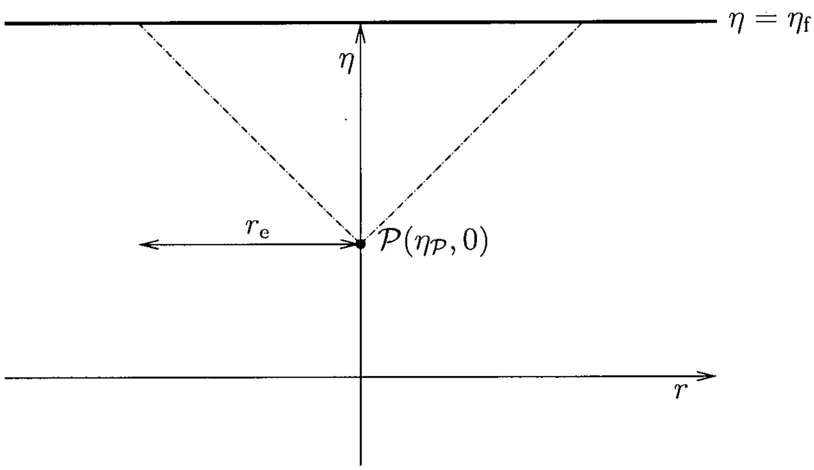

Figure 2.2: An event horizon exists when the entire spacetime is conformally related to the portion of Minkowsky spacetime below $\eta = \eta_{\mathrm{f}}$. In this case an observer at $\mathcal{P}$ will never be in causal contact with particles outside of a comoving radius r_{e}.

outside a comoving radius

$$r_{\mathrm{e}} = \eta_{\mathrm{f}} - \eta_{\mathcal{P}} = \int_{a(\eta_{\mathcal{P}})}^{\infty} \frac{da}{\dot{a}a} \tag{2.108}$$

(this expression is written for the case $a \to \infty$ as $t \to \infty$). Here the radius r_{e} defines the *event horizon.* Once again, we can define a corresponding physical distance

$$l_{\mathrm{e}} \equiv a(\eta_{\mathcal{P}})r_{\mathrm{e}}. \tag{2.109}$$

It is quite easy to specify general conditions on the function $a(t)$ that must be satisfied in order that horizons exist. If a particle horizon exists, the integral (2.106) must converge, as I explained. Thus we must have $\dot{a} \to \infty$ as $a \to 0$. (If $\dot{a}$ approaches a non-negative constant as $a \to 0$, then clearly the integral diverges logarithmically or worse.) Similarly, if an event horizon exists, then $\dot{a} \to \infty$ as $a \to \infty$ (in the expand-forever case). Thus we can immediately conclude that the spatially flat matter- and radiation-dominated solutions (2.99) and (2.101) posses no event horizons, and the Λ-dominated solution (2.102) contains no particle horizon. Indeed, in an expanding universe, whenever $\ddot{a} > 0$ as $a \to 0$ then $\dot{a}$ cannot diverge as $a \to 0$, so no particle horizon can exist.

If $\dot{a}$ increases at least as quickly as $a^{-\alpha}$ as $a \to 0$, for some positive α, then the integral (2.106) will converge and a particle horizon will exist. Similarly, if $\dot{a}$ increases at least as quickly as a^α as $a \longrightarrow \infty$, for some positive α, then an event horizon will exist. For the case of a constant equation of state $w = P/\rho$ and $\mathcal{K} = \Lambda = 0$, Eq. (2.95) implies that $\dot{a} \propto a^{-(3w+1)/2}$. Thus a particle horizon (and no event horizon) will exist for $w > -1/3$, in particular for the matter- and radiation-dominated solutions, while an event horizon (and no particle horizon) will exist for $w < -1/3$, in particular for the de Sitter solution (and indeed any expanding universe with $\ddot{a} > 0$ will contain an event horizon). I should stress that for these results to hold the appropriate behaviour of $\dot{a}$ must persist as $a \longrightarrow 0$ (for the particle horizon) or as $a \longrightarrow \infty$ (for the event horizon). This will provide a loophole to resolve the horizon problem which I will discuss in the next subsection.

It is very staightforward to evaluate the integrals (2.106) or (2.108) for the case of constant w using Eq. (2.95). The result for the physical horizon distance can be summarized as

$$l_x = \frac{2H^{-1}}{|3w+1|}, \tag{2.110}$$

where $l_x = l_{\mathrm{p}}$ (particle horizon) for $w > -1/3$, and $l_x = l_{\mathrm{e}}$ (event horizon) for $w < -1/3$. In particular, for the matter-dominated case, the physical particle horizon is

$$l_{\mathrm{p}} = 2H^{-1}, \tag{2.111}$$

while for the radiation-dominated case

$$l_{\mathrm{p}} = H^{-1}. \tag{2.112}$$

For the de Sitter case, the physical event horizon is

$$l_{\mathrm{e}} = H^{-1}. \tag{2.113}$$

It is not surprizing that the results are each of order the *Hubble radius* H^{-1}, as this is a fundamental length scale in the universe. Note also that, for this reason, the term "horizon" is often conflated with the term "Hubble radius" in the literature. However, in general (and in the particular case of inflation, as we will see), the correspondence between horizons and H^{-1} may not exist (indeed, for $w \simeq -1/3$, Eq. (2.110) shows that $l_x \gg H^{-1}$). Thus, in order to avoid ambiguity, in this thesis I will always clearly indicate whether I am referring to the particle or event horizon or to the Hubble length.

2.2.7 Shortcomings of the hot big bang[1]

The standard hot big bang model is extraordinarily successful at describing the thermal history of the early radiation-dominated phase and later matter-dominated phase of the universe. Standard texts (*e.g.* [20]) spell out its

[1] In this subsection the subscript $_0$ indicates a present value.

achievements, most notably regarding nucleosynthesis. In this subsection, I will highlight a few deep issues that the standard hot big bang model cannot address. These questions are outside of the domain of the big bang theory because they involve the form of the initial conditions used in that theory. Indeed it is not surprizing that there is more to the story than the big bang tells us, since we expect fundamentally new physics to operate as $a \to 0$. Two central issues can be summarized by the questions *why is the universe so big* (the flatness problem) and *why is the universe so smooth* (the horizon problem). A related problem is the very high density of relic particles predicted by theories of particle physics. The horizon problem is compounded by the fact that the universe is not *perfectly* smooth—inhomogeneities *do* exist, and their origin must be explained. I will demonstrate that each of the problems covered in this subsection will turn out to arise from the single fact that the Hubble length H^{-1} appears to be much smaller than the causal horizon size. Thus each will be amenable to the same solution provided by inflation and described in Section 2.3.2.

The flatness problem

Recent years have seen dramatic progress in measuring the matter and vacuum density parameters Ω_{m} and Ω_Λ. Results from the Wilkinson Microwave Anisotropy Probe (WMAP) satellite combined with Type Ia supernova observations and the Hubble Space Telecope Key Project measurement of H_0 imply that $\Omega_{\mathrm{m}} + \Omega_\Lambda = 1.02 \pm 0.02$ today [23]. Thus the Friedmann equation (2.91) demands that the curvature parameter is in the 2σ range

$$-0.04 < \Omega_{\mathcal{K}} < 0. \tag{2.114}$$

This is especially interesting because, subject to a certain restriction, the value $\Omega_{\mathcal{K}} = 0$ is an unstable fixed point in an expanding universe. To demonstrate this, I will calculate the time derivative $\dot{\Omega}_{\mathcal{K}}$ from the definition (2.92). The result is

$$\dot{\Omega}_{\mathcal{K}} = -2\Omega_{\mathcal{K}}\frac{\ddot{a}}{\dot{a}} \tag{2.115}$$

$$= \frac{2\Omega_{\mathcal{K}}}{3H}[4\pi G(\rho + 3P) - \Lambda], \tag{2.116}$$

where I have substituted Eq. (2.85) to obtain the second line. Thus if the curvature parameter $\Omega_{\mathcal{K}}$ is initially zero it will remain so, as expected. However, as long as the expanding universe decelerates in the sense that $\ddot{a} < 0$, any non-zero $\Omega_{\mathcal{K}}$ will grow in magnitude. (This is also clear from the fact that $|\Omega_{\mathcal{K}}| = \mathcal{K}/\dot{a}^2$.)

The statement that the value $\Omega_{\mathcal{K}} = 0$ is an unstable fixed point appears to be temporally asymmetric, which may seem to conflict with the symmetry of Einstein's equation. However, as Eq. (2.115) shows, the stability or instability of $\Omega_{\mathcal{K}}$ is determined by the sign of $\dot{a}$, in addition to that of $\ddot{a}$. Consider, *e.g.*, a closed universe which undergoes a big bang and a big crunch, so that $\ddot{a} < 0$ always (regardless of how we choose the direction of time!). Then the value $\Omega_{\mathcal{K}} = 0$ is unstable as we move in time away from *either* singularity, which is a time-symmetric situation. This is why I always qualify above that $\Omega_{\mathcal{K}} = 0$ is unstable in an *expanding* universe. (Formally, Eq. (2.115) contains an odd number of time derivatives on each side, so is invariant under $t \to -t$.)

I should emphasize that the statement that the curvature parameter value $\Omega_{\mathcal{K}} = 0$ is unstable in an expanding universe, and hence that any non-zero $|\Omega_{\mathcal{K}}|$ will increase, does *not* mean that the spatial curvature will increase. From Eq. (2.78) we can calculate the Ricci curvature scalar for the spatial hypersurfaces,

$$^{(3)}R = {}^{(3)}R_{ij}g^{ij} = \frac{6\mathcal{K}}{a^2}. \tag{2.117}$$

Thus in an expanding universe, the spatial curvature *always* decreases, which should be geometrically obvious. The statement that $|\Omega_{\mathcal{K}}|$ increases instead means that the contribution of spatial curvature to the rhs of the Friedmann equation (2.83) *decreases more slowly than that of the energy density.* That is, as the universe expands, any non-zero spatial curvature will have a greater and greater effect on the dynamics, relative to the effect of the matter density.

Returning to the problem this issue presents to the standard big bang theory, I have now shown that $\Omega_{\mathcal{K}} = 0$ is unstable in an expanding universe when $\ddot{a} < 0$. We can see from Eq. (2.116) above that the condition $\ddot{a} < 0$ requires a sufficiently small Λ, or, if $\Lambda = 0$, it requires $P > -\rho/3$. These conditions are certainly satisfied during the "standard" matter- or radiation-dominated phases of evolution (before any late-time domination of Λ). In fact, during matter domination the solution (2.99) gives $\dot{a} \propto a^{-1/2}$, so that $\Omega_{\mathcal{K}} \propto a$. Thus at matter-radiation equality, when the scale factor was $a_{\rm eq} \simeq a_0/3500$, this proportionality together with the observational limit (2.114) imply that the curvature parameter was $|\Omega_{\mathcal{K}}| \lesssim 10^{-5}$. Similarly, during radiation domination we have $\Omega_{\mathcal{K}} \propto a^2$. Assuming radiation domination and adiabatic evolution (temperature $T \propto a^{-1}$) all the way back to the Planck time ($T_{\rm P} \sim 10^{19}$ GeV) gives $|\Omega_{\mathcal{K}}| \lesssim 10^{-61}$ at this very early time. This clearly entails a fine-tuning problem.

This *flatness problem* can be restated in a number of ways. Corresponding to the spatial curvature parameter $\Omega_{\mathcal{K}}$ we can define a curvature length $R_{\rm curv} = (H\sqrt{|\Omega_{\mathcal{K}}|})^{-1} = a/|\mathcal{K}|$. Then at the Planck time, $R_{\rm curv} \gtrsim 10^{30} H^{-1}$, *i.e.* the curvature scale far exceeded any natural length scale. Similarly, the current

Hubble radius is $H_0^{-1} \sim 10^{28}$ cm. At the Planck time, the comoving volume corresponding to today's Hubble radius was roughly 1 μm across, or 10^{29} times the Planck length. Also, the current entropy within the Hubble radius and the total mass relative to the Planck mass are extremely large. For any of these comparisons, it is at first sight a complete mystery where such huge (or miniscule) numbers come from. Without exceedingly fine tuned initial conditions, after emerging from a quantum gravity era the universe should be expected to either recollapse on the order of the Planck time if closed, or expand rapidly to an essentially empty open universe.

Note, however, that Eq. (2.116) hints at a resolution to the flatness problem: If we can arrange to have $P < -\rho/3$ (*i.e.* $w < -1/3$) at some very early stage of evolution, then $\Omega_{\mathcal{K}} = 0$ becomes a stable fixed point and the spatial curvature parameter will be driven towards zero. Equivalently, if $\ddot{a} > 0$ at an early stage, then $\dot{a}$ increases and hence $|\Omega_{\mathcal{K}}| = \mathcal{K}/\dot{a}^2$ decreases. [Note again that this *is* a time-symmetric statement: regardless of what initial (or final!) conditions we set for $\Omega_{\mathcal{K}}$, if $w < -1/3$ then $\Omega_{\mathcal{K}}$ will be driven to zero in an *expanding* universe (and conversely will grow in a *contracting* universe).] Indeed, for the equation of state $P = -\rho$ the explicit solution (2.102) gives $\Omega_{\mathcal{K}} \propto a^{-2}$. This, in fact, is precisely how the inflationary scenario solves the flatness problem, as we will see in the next section.

The horizon problem

As I demonstrated in Section 2.2.6, both matter- and radiation-dominated evolution exhibit particle horizons. In other words, if the universe had a dust or radiation equation of state to arbitrarily early times, there are (and always were) regions that cannot have had causal contact with us. Thus at any particular time, there is no reason to expect the universe to be homogeneous on scales greater than the horizon distance at that time, $l_{\rm p} \sim H^{-1}$. Sub-horizon regions, on the other hand, *have* been in causal contact and hence could have been homogenized by microphysical processes.

While this poses no problem for our local cosmological neighbourhood, we can observe (apparently) causally disconnected regions at early enough times. For example, the cosmic microwave background (CMB) radiation was emitted at around the time of matter-radiation decoupling at an age of roughly 370 kyr. We observe the CMB temperature to be highly uniform ($\Delta T/T < 10^{-4}$) even in opposite directions in the sky. Opposing patches of the CMB are separated, in comoving coordinates, by a few times our *current* comoving Hubble radius $(a_0 H_0)^{-1}$ (see Fig. 2.3). At the decoupling time $t_{\rm dec}$, the comoving Hubble radius (and by assumption the causal length scale) was of course much smaller than it is today. Thus we are led to the serious *horizon problem*: The universe

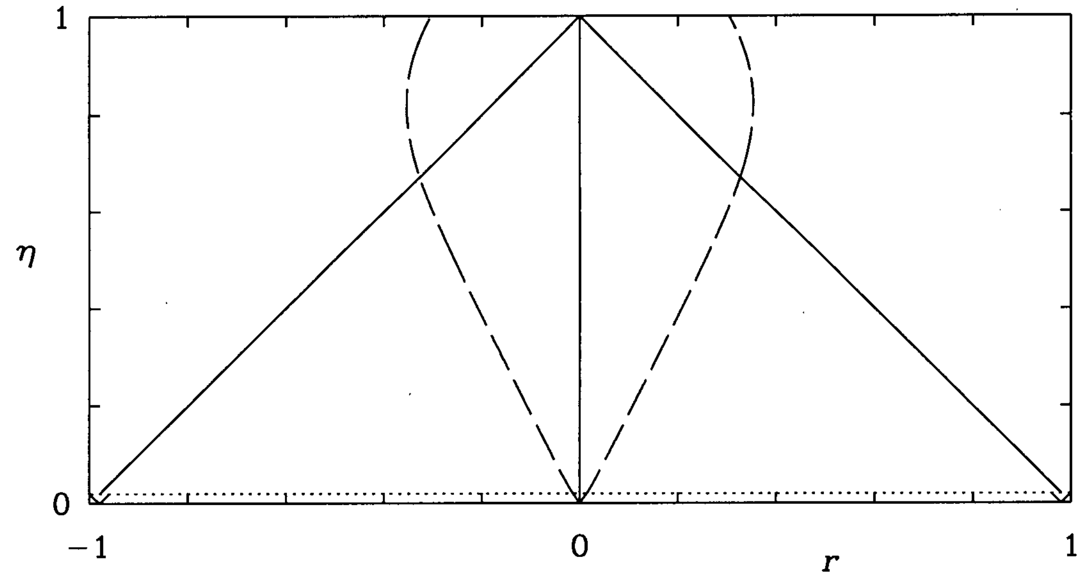

Figure 2.3: Conformal spacetime diagram extending from today ($\eta = 1$) to the singularity ($\eta = 0$), assuming radiation domination as $a \to 0$. The coordinate r is *comoving,* and our worldline is the vertical line at $r = 0$. The long diagonal lines indicate our past light cone to $t_{\rm dec}$ (indicated by the dotted line), when the CMB was released. The small cones indicate the (apparent) horizon size at that time. This diagram is *to scale*, so we can directly see how much larger the observed smoothness scale is than the (apparent) horizon size at $t_{\rm dec}$. The dashed curves are the numerically calculated comoving Hubble length, $1/\dot{a}$, which increases during matter domination but decreases at late times due to the vacuum density (I used $\Omega_{\rm m} = 0.29$, $\Omega_\Lambda = 0.71$ today). Since $|\Omega_\mathcal{K}| = \mathcal{K}/\dot{a}^2$, the diagram also illustrates the extreme fine tuning required to give a small $\Omega_\mathcal{K}$ today.

was apparently not nearly old enough at $t_{\rm dec}$ to allow what we observe as opposing patches to have interacted and equilibrated to the same temperature. Indeed it is straightforward to calculate the angle subtended at earth by the horizon size at $t_{\rm dec}$ (assuming radiation domination as $a \to 0$). Using scale factors at the decoupling and equality times of $a_0/a_{\rm dec} = 1090$ and $a_0/a_{\rm eq} = 3500$, respectively [23], I find that the causal horizon at the emission of the CMB subtends only $1^\circ.3$. Why does the cosmological principle apply on scales that are apparently causally disconnected?

This problem is only amplified if we consider earlier times still. For example, at the time of nucleosynthesis ($a_{\rm nuc}/a_0 \sim 10^{-9}$), the causal horizon was roughly 10^{-11} times the comoving width of today's Hubble radius (assuming, again, that $w = 1/3$ as $a \to 0$). And, as mentioned above, at the Planck time the corresponding length ratio was 10^{-29}, so that something like 10^{87} apparently causally disconnected regions eventually expanded to form our strikingly homogeneous universe!

I have carefully stated a crucial assumption in discussing the horizon problem, namely that of a radiation dominated equation of state as $a \to 0$. Recall from Section 2.2.6 that if the equation of state instead approaches a constant $w < -1/3$ as $a \to 0$, then there will be no particle horizon. Thus such a period of evolution before the standard hot big bang phase could enable contact between regions previously assumed to be causally disconnected. Note from above that the condition $w < -1/3$ is also required to solve the flatness problem. Inflation, as we will see, fulfills this requirement and can thus solve both problems.

Relic abundances

Particle physicists expect that the strong and electroweak interactions were unified at an energy scale above roughly 10^{14} GeV. As the universe cooled below the corresponding temperature, a unified gauge symmetry was spontaneouly broken. As a result, point topological defects (magnetic monopoles) are generically expected to be produced [20]. Because the fields undergoing symmetry breaking should not be correlated on length scales greater than the particle horizon, it is expected that the number of monopoles generated this way (via the *Kibble mechanism*) is of order unity per horizon volume. Their mass is expected to be very high, of order 10^{16} GeV, and they are not expected to annihilate quickly. Since the comoving horizon length at the symmetry breaking time was of order 10^{-24} times the current comoving horizon length (assuming radiation domination as $a \to 0$), the monopoles are expected to contribute a "somewhat significant" density today, namely approximately 10^{11} times the critical density! Equivalently, the energy density of heavy particles like monopoles should decay like a^{-3}, so that they would quickly dominate over the radiation background. Other types of particles are similarly predicted to be produced and to contribute too large relic abundances.

A solution is provided, again, by proposing that *after* the symmetry breaking production of monopoles there was a period during which the scale factor accelerated, $\ddot{a} > 0$ (*i.e.* $w < -1/3$). This condition is equivalent to the condition that the comoving Hubble radius $(aH)^{-1} = \dot{a}^{-1}$ *decreases*. Thus, after a sufficiently long period of accelerated expansion, the number of relics per comoving Hubble volume will become as small as is needed. Equivalently, for

$w < -1/3$, the total density must decay more slowly than a^{-2} [see Eq. (2.68)], and hence will quickly dominate over any heavy particles. As long as the temperature after such a period is not high enough to restore the symmetry, the problem can be solved.

A common solution

I have indicated that each of the three problems I have discussed here can be solved if, before the standard radiation-dominated phase, there was a period during which $w < -1/3$, or equivalently $\ddot{a} > 0$, *i.e.* the comoving Hubble radius $(aH)^{-1}$ decreased. This suggests that there is essentially a single physical source to these problems, namely that the true causal horizon size is much greater than the current Hubble radius H^{-1}. In other words, it appears that when we examine our current Hubble volume, we are actually seeing only a small fraction of the current causally connected volume. Hence our Hubble volume can be extremely smooth, causally connected, and very dilute of heavy relics.

I illustrate the situation in Fig. 2.4. This is a conformal spacetime diagram, like Fig. 2.3, so the causal structure is manifest. The figure is identical to Fig. 2.3, except that I have added a very early period during which $w = -1$. The new period ended just after $\eta = 0$, when a/a_0 was extremely small (but non-zero), so the singularity has been "pushed down" to a time earlier than the plot shows. It should be clear from the diagram that if this new early period was long enough, then it provided plenty of spacetime volume to causally connect opposing patches of CMB.

Fig. 2.4 also shows the comoving Hubble length, $1/\dot{a}$, which *decreased* during the new early period. Since $|\Omega_\mathcal{K}| = \mathcal{K}/\dot{a}^2$, the plot shows how the new period solves the flatness problem: At some extremely early time (*e.g.* $\eta = -1$), we suppose that $\Omega_\mathcal{K}$ was small but not exceedingly so (*e.g.* $\Omega_\mathcal{K} \sim 0.01$). Then $\Omega_\mathcal{K}$ was driven exceedingly close to zero while $w = -1$ (recall that $\Omega_\mathcal{K} = 0$ is *stable* for $w < -1/3$). Later, during the familiar radiation- and matter-dominated stages, $\Omega_\mathcal{K}$ grew back to some value consistent with observations. The need to explain exceedingly fine-tuned values of the curvature parameter near $\eta = 0$ is thus apparently removed.

Finally, Fig. 2.4 also illustrates how the early period of accelerated expansion can solve the relic problem. Suppose, for example, that defects were produced near the time $\eta \simeq -1/3$. At this time, the plot shows that the comoving Hubble length was comparable to today's value. Thus we immediately conclude that, if roughly one defect was produced per Hubble volume, then roughly one would lie in our *current* Hubble volume (or fewer, if the defects annihilate).

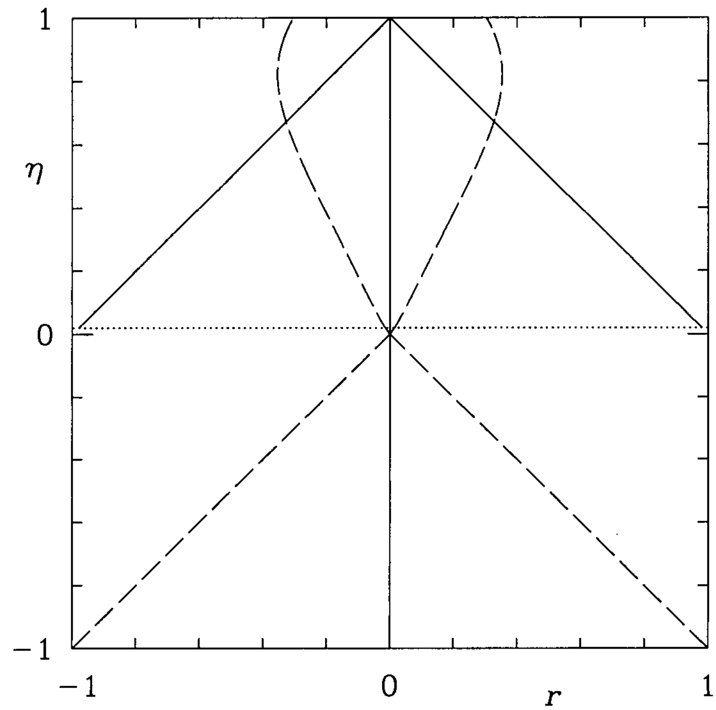

Figure 2.4: Conformal spacetime diagram, identical to Fig. 2.3, except with a very early period (prior to $\eta = 0$) when $w = -1$. Again I plot our past light cone to t_{dec}, but now there is plenty of volume at earlier times to account causally for the smoothness scale. The dashed curves indicate the comoving Hubble length, $1/\dot{a}$, which *decreases* during the new early period. (Again I used $\Omega_{\mathrm{m}} = 0.29$ and $\Omega_{\Lambda} = 0.71$ today and plotted the diagram *to scale*.)

It should now be clear that our only hope to understand the apparently very baffling initial conditions required by the big bang is by extending the spacetime as I did in going from Fig. 2.3 to Fig. 2.4. A very early period when the comoving Hubble length decreased accomplishes this task. Inflation, as we will see next, provides a *dynamical* realization of such a period.

2.3 Scalar fields and inflation

Finally it is time to write down the fundamental action of a matter field and deduce its possible cosmological consequences. In current theories of particle physics, scalar fields play a fundamental role. They occur in the process of spontaneous symmetry breaking, during which some gauge group is broken and some fields gain mass. An example is the Higgs field which is expected to facilitate the electroweak transition. In addition, many scalar fields are expected to be associated with the compactified dimensions in string theory. Aside from any particle physics plausibility, a scalar field at very early times satisfying certain relatively weak constraints will provide, as we will see, a scenario which can solve each of the problems with the standard hot big bang model described in the previous section. This *inflationary* scenario involves a dynamical, scalar-field driven period of evolution during which the comoving Hubble radius decreases by many e-folds, setting the stage for the traditional hot big bang. The tremendous success of inflation is qualified by new challenges, such as the need to explain very small coupling constants and to work out the details of the transition between inflation and the standard hot big bang. Of course the greatest challenge is in finding a convincing particle physics realization of inflation. The flip side is the potential for cosmological observations to elucidate physics beyond the standard model. In this section I will briefly describe the *homogeneous* and *classical* dynamics of inflation.

2.3.1 Scalar fields

I will consider a system of N scalar fields φ_A, $A = 1, \ldots, N$. The action for this system coupled to gravity takes the general form

$$S_{\mathrm{m}} = -\frac{1}{2}\int \left(\varphi^{;\mu}\cdot\varphi_{;\mu} + \xi R\varphi\cdot\varphi + 2V(\varphi_A)\right)\sqrt{g}\, d^4x. \tag{2.118}$$

Here, in order to reduce index clutter, I have defined

$$\varphi\cdot\varphi \equiv \varphi_A\varphi^A \tag{2.119}$$

(I will assume a Euclidean "field space" metric, so $\varphi_A = \varphi^A$). The kinetic term $\varphi^{;\mu}\cdot\varphi_{;\mu}$ ensures that the scalar field equation of motion will be a second order differential equation. The parameter ξ couples the scalar fields to gravity via the Ricci scalar curvature, and is in principle free, although there are two important special values. For $\xi = 1/6$, the scalar field dynamics is *conformally invariant*, in the sense that, for $V = 0$, φ_A is a solution to the field equations with metric $g_{\mu\nu}$ if and only if $\Omega^{-1}\varphi_A$ is a solution with metric $\Omega^2 g_{\mu\nu}$ [17]. The case $\xi = 0$ is called *minimal coupling*, as it results in the simplest action. I will assume minimal coupling throughout this thesis. The potential function $V(\varphi_A)$ enables the fields to interact and have masses.

The classical scalar field equation of motion is very straightforward to derive by varying the action S_{m} with respect to φ_A or equivalently by writing

Lagrange's equations. The result is the N-component equation

$$\Box\varphi_A - V_{,\varphi_A} = 0, \tag{2.120}$$

where $V_{,\varphi_A} \equiv \partial V/\partial\varphi_A$. This is known as the *Klein-Gordon* equation.

The scalar field energy-momentum tensor is almost as simple to derive from the fundamental definition (2.18). Varying the scalar Lagrangian density with respect to the metric, I find

$$\delta\mathcal{L} = -\frac{1}{2}[\delta\sqrt{g}(g^{\mu\nu}\varphi_{;\mu}\cdot\varphi_{;\nu} + 2V) + \sqrt{g}\delta(g^{\mu\nu})\varphi_{;\mu}\cdot\varphi_{;\nu}]. \tag{2.121}$$

Using relations (2.13) and (2.14) for the variations $\delta(g^{\mu\nu})$ and $\delta\sqrt{g}$, this becomes

$$\delta\mathcal{L} = -\frac{1}{4}[g^{\mu\nu}(\varphi_{;\lambda}\cdot\varphi^{;\lambda} + 2V) - 2\varphi^{;\mu}\cdot\varphi^{;\nu}]\sqrt{g}\delta g_{\mu\nu}. \tag{2.122}$$

Thus we can simply read off the energy-momentum tensor from the definition (2.18),

$$T^{\mu\nu} = \varphi^{;\mu}\cdot\varphi^{;\nu} - \frac{1}{2}g^{\mu\nu}\left(\varphi_{;\lambda}\cdot\varphi^{;\lambda} + 2V(\varphi_A)\right). \tag{2.123}$$

Notice that the equation of motion (2.120) is also contained within the covariant energy-momentum conservation law, $T^{\mu\nu}{}_{;\nu} = 0$. For

$$\begin{aligned} 0 = T^{\mu\nu}{}_{;\nu} &= \varphi^{;\mu}{}_{\nu}\cdot\varphi^{;\nu} + \varphi^{;\mu}\cdot\varphi^{;\nu}{}_{\nu} - \frac{1}{2}g^{\mu\nu}(\varphi_{;\lambda\nu}\cdot\varphi^{;\lambda} + \varphi_{;\lambda}\cdot\varphi^{;\lambda}{}_{\nu} + 2V_{,\varphi}\cdot\varphi_{;\nu}) \\ &= (\Box\varphi - V_{,\varphi})\cdot\varphi^{;\mu}. \end{aligned} \tag{2.124}$$

This must hold for each arbitrary component $\varphi^{;\mu}_A$, hence energy-momentum conservation implies the Klein-Gordon equation (2.120). Conversely, the argument clearly runs in reverse, so that the Klein-Gordon equation implies conservation of $T^{\mu\nu}$.

The preceeding results for scalar fields are completely general (apart from the minimal coupling assumption). In this chapter, however, I am considering the dynamics of a precisely homogeneous and isotropic universe. As discussed in Section 2.2.1, precise homogeneity means there exists a time coordinate t such that all matter fields are form-invariant under all spatial translations on the constant-time hypersurfaces. This implies that a scalar field depends only on t (which follows also from isotropy, which demands that there exists a frame in which $\varphi^A_{,i} = 0$). Thus I will now restate the general results above for the case of homogeneous fields φ_A in a spatially flat FRW background.

The d'Alembertian becomes (suppressing the scalar field index A)

$$\Box\varphi = \varphi^{;\mu}{}_{\mu} + \Gamma^{\mu}_{\mu\lambda}\varphi^{;\lambda} = \begin{cases} \varphi^{;\mu}{}_{\mu} + 3H\varphi^{;0} & t \text{ chart} \\ \varphi^{;\mu}{}_{\mu} + 4\mathcal{H}\varphi^{;0} & \eta \text{ chart} \end{cases} \tag{2.125}$$

using the FRW connection coefficients (2.72) or (2.73). Setting all spatial gradients to zero, the Klein-Gordon equation becomes

$$\ddot{\varphi}_A + 3H\dot{\varphi}_A + V_{,\varphi_A} = 0 \tag{2.126}$$

or

$$\frac{1}{a^2}(\varphi''_A + 2\mathcal{H}\varphi'_A) + V_{,\varphi_A} = 0. \tag{2.127}$$

There are only two distinct components of the homogeneous scalar field energy-momentum tensor. In the proper time chart, they are

$$T^{00} = \rho = \frac{1}{2}\dot{\varphi}\cdot\dot{\varphi} + V(\varphi_A) \tag{2.128}$$

and

$$T^{ij} = \frac{1}{2}g^{ij}(\dot{\varphi}\cdot\dot{\varphi} - 2V(\varphi_A)), \tag{2.129}$$

which gives for the isotropic pressure

$$P = \frac{1}{3}T^i_{\ i} = \frac{1}{2}\dot{\varphi}\cdot\dot{\varphi} - V(\varphi_A). \tag{2.130}$$

Note that, as a consistency check, the energy-momentum tensor we have obtained for the homogeneous scalar field is of the general form (2.60) imposed by the cosmological principle. Note also that for a homogeneous and static field ($\dot{\varphi}_A = 0$) we have a constant equation of state $w = P/\rho = -1$, which corresponds to a cosmological-constant-like energy-momentum tensor. This is the familiar result that a constant scalar field merely amounts to a "restructuring" of the vacuum; the energy-momentum tensor affects only gravitational dynamics.

2.3.2 Homogeneous inflationary dynamics

The presence of a scalar field at very early times can dramatically change the evolutionary history of the universe. Indeed, the equation of state of (approximately) $w = -1$ for a (nearly) constant scalar field mentioned above implies that the scale factor will increase (nearly) exponentially [recall Eq. (2.102)]. Such an approximately de Sitter phase of evolution is called a period of *inflation*. Crucially, an inflationary equation of state satisfies the condition $w < -1/3$, *i.e.* the comoving Hubble length decreases, which as I explained in Section 2.2.7 allows us to resolve each of the problems with the hot big bang theory described there. The goal of this section is to make plausible such an inflationary phase, given an appropriate scalar field, and to determine the conditions that allow inflation. I will, for definiteness, mainly discuss the special case of a massive, non-interacting scalar field. Later in this section I will briefly generalize to arbitrary potentials, emphasizing the *physical* requirements for inflation. *Warning:* This section will, I am afraid, involve some hand-waving.

Initiating chaotic inflation

There are very many scalar field models that realize inflation [24]. Here I will only describe the simplest, that of *chaotic inflation* (see [21] for an excellent review), for the case of a single-component non-interacting massive scalar field φ, *i.e.* a field with the potential

$$V(\varphi) = \frac{1}{2}m^2\varphi^2. \tag{2.131}$$

Such a scalar field that leads to inflation is called an *inflaton field.* Suppose that during a very early period the matter sector is dominated by the scalar field φ. At some extremely early time, the energy density in φ will be of order m_{P}^4, where $m_{\mathrm{P}} = G^{-1/2}$ is the Planck mass. Prior to this time we expect to require a quantum mechanical description of the spacetime. Consider a time soon after this, when it has become roughly sensible to speak of a classical spacetime. If we cavalierly suppose that at this time the scalar field is randomly distributed due to the quantum fluctuations of the preceeding era (hence the origin of the name "chaotic" inflation), then we should expect the energy density to be randomly distributed spatially according to

$$\rho \sim m_{\mathrm{P}}^4. \tag{2.132}$$

Decomposing into potential and kinetic terms, we also expect them to be randomly distributed according to

$$V(\varphi) \sim m_{\mathrm{P}}^4, \tag{2.133}$$

$$\varphi_\mu\varphi_\nu \sim m_{\mathrm{P}}^4. \tag{2.134}$$

Similarly, invariants R constructed from the Riemann tensor should satisfy

$$R \sim m_{\mathrm{P}}^2. \tag{2.135}$$

Now consider at this early time a region of physical size on the order of the Planck length or larger, in which the spatial gradients of φ and the curvature invariants which contribute to spatial inhomogeneity and anisotropy are several times below their averages. That is, consider a relatively spatially smooth region in the "chaotic sea". In this region, we can consider the metric to be approximately FRW in form, and hence the energy constraint equation (2.83) becomes

$$H^2 \simeq \frac{4\pi}{3m_{\mathrm{P}}^2}(\dot{\varphi}^2 + m^2\varphi^2) - \frac{\mathcal{K}}{a^2}. \tag{2.136}$$

Here I have neglected spatial gradients and inserted the energy density for the homogeneous scalar field, Eq. (2.128). Similarly, the Klein-Gordon equation approximately takes the homogeneous form

$$\ddot{\varphi} + 3H\dot{\varphi} + m^2\varphi = 0. \tag{2.137}$$

This equation of motion is precisely of the form of a damped harmonic oscillator with frequency m and time-dependent damping proportional to $H(t)$. Thus, for $H(t)$ sufficiently slowly varying, we expect either overdamped decay of φ or underdamped oscillations, depending on the relative strength of the damping to the oscillation frequency, *i.e.*

$$\frac{3}{2}H \begin{cases} < m & \text{underdamped oscillations,} \\ > m & \text{overdamped decay.} \end{cases} \tag{2.138}$$

Homogeneous dynamics

First consider the behaviour in the strongly overdamped regime. We will soon see what conditions on φ this implies. We can write the solution to the homogeneous Klein-Gordon equation as a sum of two decaying modes in the adiabatic approximation,

$$\begin{aligned} \varphi = {} & C_1 \exp\left\{-\int \left[3H/2 - \sqrt{(3H/2)^2 - m^2}\right] dt\right\} \\ & + C_2 \exp\left\{-\int \left[3H/2 + \sqrt{(3H/2)^2 - m^2}\right] dt\right\} + \mathcal{O}(\dot{H}/H^2). \end{aligned} \tag{2.139}$$

I will make the assumption

$$|\dot{H}| \ll H^2 \tag{2.140}$$

and drop the $\mathcal{O}(\dot{H}/H^2)$ terms, and check at the end that this is a consistent assumption. In the strongly overdamped regime,

$$H \gg m \tag{2.141}$$

and the mode proportional to C_2 decays much more quickly than the other mode. Thus after a brief transient the field becomes

$$\varphi \simeq C_1 \exp\left(-\frac{m^2}{3}\int H^{-1}\, dt\right), \tag{2.142}$$

and hence

$$\dot{\varphi} \simeq -\frac{m^2 \varphi}{3H}. \tag{2.143}$$

Therefore, recalling the form of the Klein-Gordon equation (2.137), we see that in the strongly overdamped regime we can ignore the second time derivative $\ddot{\varphi}$ (*i.e.* the "force" due to the potential matches that due to the damping). Also, Eq. (2.143) and the condition (2.141) imply that

$$|\dot{\varphi}| \ll m|\varphi|, \tag{2.144}$$

so that the energy density is potential dominated and the Hubble parameter (2.136) becomes simply

$$H^2 \simeq \frac{4\pi}{3m_\mathrm{P}^2} m^2\varphi^2 - \frac{\mathcal{K}}{a^2}. \tag{2.145}$$

Taking the time derivative gives

$$\dot{H} \simeq -\frac{m^2}{3} + \frac{\mathcal{K}}{a^2}. \tag{2.146}$$

Now $m^2 \ll H^2$ by the strongly overdamped assumption, and the spatial curvature term decays approximately exponentially, as we will see in a moment. Thus if the spatial curvature is not too great initially, then $|\dot{H}| \ll H^2$ rapidly holds, and the earlier assumption of adiabaticity is consistent.

Approximate solutions in the strongly overdamped regime are easy to derive. The scalar field decays according to Eq. (2.142). For times such that $Ht \ll H^2/m^2$, this expression can be approximated by

$$\varphi(t) \simeq \varphi(0) - \frac{m_\mathrm{P} m}{\sqrt{12\pi}}\, t\, \mathrm{sgn}(\varphi). \tag{2.147}$$

The scale factor is given (as always) by the exact expression

$$\frac{a}{a_0} = e^{\int H\, dt}, \tag{2.148}$$

where in H any spatial curvature term rapidly decays and we can write

$$H \simeq \sqrt{\frac{4\pi}{3}} \frac{m|\varphi|}{m_\mathrm{P}}. \tag{2.149}$$

We can easily calculate the number of e-folds of expansion, defined by

$$N \equiv \int H\, dt. \tag{2.150}$$

Combining Eq. (2.149) for the Hubble parameter with Eq. (2.143) for $\dot{\varphi}$, I find

$$N = \frac{2\pi}{m_\mathrm{P}^2}\left(\varphi_0^2 - \varphi_\mathrm{f}^2\right) \tag{2.151}$$

for the number of e-folds between the times that $\varphi = \varphi_0$ and $\varphi = \varphi_\mathrm{f}$. Since $\dot{H} \ll H^2$, the time evolution of the scale factor can be called *quasi-exponential.* Indeed, Eq. (2.143) together with the scalar field expressions for energy density and pressure, Eqs. (2.128) and (2.130), give

$$w + 1 \simeq \frac{\dot{\varphi}^2}{V} \simeq \frac{1}{6\pi}\frac{m_\mathrm{P}^2}{\varphi^2} \ll 1 \tag{2.152}$$

in the overdamped regime. In words, the equation of state is close to that of the exponentially expanding de Sitter space ($w = -1$) and we have achieved inflation. Also note that the neglected spatial gradient terms rapidly redshift, adding weight to their exclusion.

The overdamping condition (2.138) together with the approximate form of H in the overdamped regime, Eq. (2.149), imply that the universe will inflate when

$$|\varphi| \gtrsim \frac{m_{\mathrm{P}}}{\sqrt{3\pi}}. \tag{2.153}$$

The proposed initial condition (2.133) implies that initially

$$|\varphi| \sim \frac{m_{\mathrm{P}}^2}{m}. \tag{2.154}$$

Thus if

$$m \lesssim m_{\mathrm{P}} \tag{2.155}$$

then the universe will begin (at the early time we have been considering) in the overdamped, inflating phase (provided, of course, that inhomogeneities are small at this early time as we have assumed). This condition on m will turn out to be very easily satisfied because the inflationary generation of an appropriate amplitude of perturbations will require a very light inflaton mass. Also, according to Eqs. (2.151) and (2.154), we expect

$$N \sim \frac{m_{\mathrm{P}}^2}{m^2} \tag{2.156}$$

e-folds of inflation in total, which will be a large number for a light inflaton.

The scalar field decays according to Eq. (2.142) during inflation. Therefore at some critical time, the condition (2.153) will become violated and the field will start to undergo underdamped oscillations,

$$\varphi \simeq \left(\frac{a}{a_0}\right)^{-3/2} \left[C_1 \cos(mt) + C_2 \sin(mt)\right] \tag{2.157}$$

for $|\varphi| \ll m_{\mathrm{P}}$. If, as we are doing in this chapter, we ignore spatial inhomogeneities, these damped oscillations will continue indefinitely, leading to a very uninteresting featureless universe. It is only when perturbations of the inflaton and other fields are considered that the process of *reheating* will transform the oscillating homogeneous inflaton into the particle excitations that we see in the actual universe. Aspects of this process will be the subject of Chapter 4.

To summarize this description of the inflationary dynamics of a massive scalar field, in order to obtain chaotic inflation we must begin with two relatively painless assumptions. The first is that at a very early post-Planck time, inflaton and metric quantities are distributed randomly according to Eqs. (2.133) to (2.135), and that in some region spatial inhomogeneities of the inflaton field as well as of the spacetime can be ignored. The second assumption is that the inflaton mass is light, $m \lesssim m_{\mathrm{P}}$. Then, the scalar field will undergo overdamped decay and the spacetime will undergo a near-de Sitter inflationary phase. Once $|\varphi|$ drops below a critical value, the field undergoes underdamped oscillations which decay into inhomogeneities through the process of reheating. The stage has been set for the standard hot big bang.

General single-field potentials

It is of course possible to generalize the conditions that allow inflation to arbitrary single-field potentials $V(\varphi)$. The condition for a strongly inflating, near-de Sitter state can be written in a number of completely equivalent ways:

$$w + 1 \ll 1 \iff \dot{\varphi}^2 \ll V(\varphi) \iff -\frac{\dot{H}}{H^2} \ll 1. \tag{2.158}$$

The first condition tells us immediately that the comoving Hubble length decreases. The middle condition says that in the energy density, the potential dominates over the kinetic term. This is the origin of the expression *slow roll* which is often used to describe a strongly inflating state. The final condition indicates that the expansion is quasiexponential.

In order that inflation last long, we must again have overdamped decay of the inflaton. In the general case, the field's oscillation frequency (or effective mass) is time-dependent,

$$m_\varphi^2 = V_{,\varphi\varphi}. \tag{2.159}$$

The overdamping condition thus becomes

$$V_{,\varphi\varphi} < \frac{9}{4}H^2, \tag{2.160}$$

and when this is strongly satisfied the Klein-Gordon equation again effectively becomes the first-order

$$3H\dot{\varphi} \simeq -V_{,\varphi}. \tag{2.161}$$

When we have both strong inflation and an overdamped scalar field, we can write conditions in terms of the potential alone. The third condition in (2.158) combined with Eq. (2.161) can be written

$$\epsilon \equiv \frac{m_{\mathrm{P}}^2}{16\pi}\left(\frac{V_{,\varphi}}{V}\right)^2 \ll 1. \tag{2.162}$$

Similarly, the overdamping condition (2.160) combined with the middle condition in (2.158) give

$$\eta \equiv \frac{m_{\mathrm{P}}^2}{8\pi}\frac{V_{,\varphi\varphi}}{V} \ll 1. \tag{2.163}$$

While the two "slow-roll conditions" (2.162) and (2.163) are necessary for inflation to occur and to last long, they are not sufficient, since they do not exclude any transient phase where the kinetic term may dominate over the potential. Also note that for the massive scalar field $V = m^2\varphi^2/2$ (and indeed for any power law potential) both slow-roll conditions are equivalent.

What does inflation do for us?

I have now shown that for a cosmological scalar field system, we can obtain a period during which φ undergoes overdamped decay and the equation of state is essentially vacuum-dominated, $w + 1 \ll 1$. This latter condition is sufficient to resolve the flatness, horizon, and relic problems of the standard big bang theory, as I discussed in Section 2.2.7. Indeed, Fig. 2.4 illustrates precisely how inflation solves these problems, as that plot was drawn for the case $w = -1$ before the standard radiation-dominated phase. While for that figure the inflationary stage was simply "tacked on", we now have the dynamics under our belt to do a realistic calculation. In Fig. 2.5 I plot the comoving Hubble length, $1/\dot{a}$, calculated by numerically evolving the homogeneous Klein-Gordon equation, Eq. (2.126), coupled with the flat space Einstein energy constraint, Eq. (2.83), for the case of a quartic chaotic inflation potential,

$$V(\varphi) = \frac{\lambda}{4}\varphi^4. \tag{2.164}$$

I chose an initial value of $\varphi = 1.38m_{\mathrm{P}}$, which gave a reasonable span of inflation, and integrated through a similar span (in conformal time) of underdamped oscillations. The value of λ is irrelevant here since it only sets the time scale, which is arbitrary in the figure. Fig. 2.5 shows the comoving Hubble length decreasing during inflation, as expected, followed by a period of modulated growth. The oscillations do not decay since here I have only integrated the background equations and hence have ignored energy transfer to fluctuations. This plot can be considered an extreme blow-up of the "throat" in Fig. 2.4.

How long must inflation last? As I explained in Section 2.2.7, the temperature after inflation must not exceed roughly 10^{14} GeV in order to prevent defect formation. Assuming radiation domination all the way back, this means that there could have been at most roughly 60 e-folds of expansion since the end of inflation. In addition, to solve the problems discussed in Section 2.2.7,

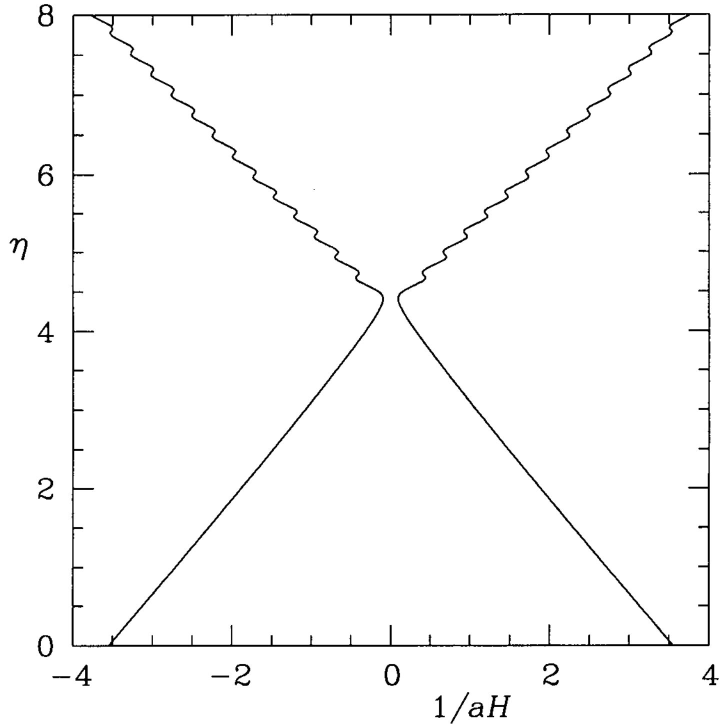

Figure 2.5: Comoving Hubble length as a function of conformal time during the latter part of inflation and into the reheating stage for a quartic scalar field potential, ignoring inhomogeneities. Initially $\varphi = 1.38m_{\rm P}$. This plot is an extreme blow-up of the "throat" in Fig. 2.4.

we required that the comoving Hubble length at the beginning of inflation was as least of the order of the current comoving Hubble length, *i.e.*

$$\dot{a}|_{\rm beginning} \lesssim \dot{a}|_{\rm today}. \tag{2.165}$$

But during inflation, $\dot{a} \propto a$, and during radiation domination, $\dot{a} \propto a^{-1}$. Since we have at most 60 e-folds after inflation, we therefore can solve the flatness, smoothness, and relic problems if inflation lasted at least $N \simeq 60$ e-folds. As I explained above, typically the inflaton must be very light, and so Eq. (2.156) tells us that we can very easily attain at least 60 e-folds. Of course I have neglected many details here, but my purpose was simply to explain how a number of the size $N \simeq 60$ appears as an inflationary requirement.

I will close this chapter with a few comments on the rather mysterious sounding initial conditions required to initiate inflation. While many argue that those initial conditions are reasonable [21], it should be emphasized that it is by no means necessary to attempt to describe the universe near the Planck era, as I have very sketchily done above, in order to understand the inflationary period. Indeed, even if we were to conclude that inflation did occur, but found that it required fine-tuned initial conditions, we still would have accomplished a great deal, namely the description of an extremely early period of the universe. Just as it should not be the job of the big bang theory to explain its initial conditions, it should not (necessarily) be the job of inflation to explain its.

Chapter 3

Cosmological Perturbations

The spatially homogeneous and isotropic universe described in the previous chapter is of course only an approximation. In this chapter I will generalize the results of the previous chapter in order to describe the perturbations from homogeneity that must eventually lead to the structure observed at late times. The general approach in cosmological perturbation theory is to decompose any exact geometrical or matter quantity $S(t, x^i)$ according to

$$S(t, x^i) = {}^0S(t) + \delta S(t, x^i). \tag{3.1}$$

Here ${}^0S(t)$ is a homogeneous and isotropic *background* part, which, as discussed in the previous chapter, must depend only on the time that foliates the background spacetime homogeneously, and $\delta S(t, x^i)$ is the *perturbation*, which encapsulates the departures from homogeneity of the exact quantity.

In this chapter I will introduce three techniques to simplify the treatment of the perturbations. First, in Section 3.1 I will describe a decompostion of the perturbed metric into terms that transform spatially as scalars, vectors, and tensors. This will greatly simplify metric calculations as it will enable us to treat individually the perturbation modes that lead to structure formation and those that correspond to vector modes or gravitational waves. In Section 3.2 I will provide a full geometrical interpretation of the metric perturbation functions, decomposing the curvature into intrinsic and extrinsic parts. In Section 3.3 I will repeat the spatial transformation decomposition on the energy-momentum tensor and generalize the covariant conservation law to linear perturbations.

In general relativity, perturbations contain an inherent ambiguity. This *gauge freedom* is especially important for cosmological perturbations, and in Section 3.4 I will describe it in detail. I will elucidate the origin of the ambiguity, and point out that removing the ambiguity by fixing a gauge will in fact enable a simplification of the form of the metric and the equations of motion. I will show that only temporal gauge transformations are "physical", in that linear spatial transformations do not affect Einstein's equations due to the homogeneity of the background. I will describe several choices of gauge that are physically motivated, and close with some cautionary remarks on the applicability of gauge transformations.

In Section 3.5 I will derive the perturbed dynamical equations under the linear approximation. This is known to be a valid approximation at sufficiently early times, and it will substantially simplify the equations of motion. My derivation is novel: To ease metric manipulations, I derive the equations in a particular gauge, and then generalize them to completely arbitrary gauges by utilizing the known gauge transformation properties of the Einstein tensor. I stress the utility of presenting the dynamical and conservation equations in arbitrary gauge: this enables us to easily specialize to any particular gauge, if that gauge is well-defined.

While in general the perturbed Einstein equations consist of ten metric functions satisfying nonlinear equations, the simplifying techniques I introduce will reduce this to two metric functions satisfying linear equations, for the modes that can form structure. Even more reduction will be possible when the matter is a scalar field, which I discuss in Section 3.6. I derive the general forms of the dynamical equations for scalar field perturbations, and then write them in longitudinal gauge, where only a single metric function appears.

Parts of the material in this chapter are based very loosely on the review articles [25, 26] as well as standard texts, in particular Wald [17].

3.1 Decomposition of the metric

The completely general perturbed line element has the form

$$ds^2 = a^2(\eta)\left[-(1+2\phi)d\eta^2 + 2B_i dx^i d\eta + E_{ij}dx^i dx^j\right], \tag{3.2}$$

where ϕ and B_i are arbitrary functions and E_{ij} is symmetric. (We have ten independent functions altogether, as required for the arbitrary symmetric tensor $\delta g_{\mu\nu}$.) This general metric reduces to the FRW metric (2.54) for $\phi = B_i = 0$ and $E_{ij} = \gamma_{ij}$. I have made no assumption yet about the size of the departures in the general metric from the homogeneous FRW metric. As I discussed in Section 2.2.1, under the restricted class of spatial coordinate transformations on constant-time hypersurfaces

$$x^0 \to \tilde{x}^0 = x^0, \qquad x^i \to \tilde{x}^i = \tilde{x}^i(x^j), \tag{3.3}$$

ϕ transforms as a scalar while B_i and E_{ij} transform as spatial three-vectors and second-rank three-tensors, respectively. Here, the lack of homogeneity means that we cannot in general find coordinates such that all spatial vectors vanish and scalars depend only on the time. In reference to metric perturbation functions, the terms "scalar", "vector", and "tensor" will always refer to these spatial transformation properties, rather than to spacetime transformation properties.

Geometrically, the function ϕ determines the lapse function, which specifies the ratio between proper- and coordinate-time separations between two neighbouring constant-time hypersurfaces. The function $a^2 B_i$ is the shift vector and specifies the rate of deviation of a constant spatial coordinate line from a line normal to a constant-time hypersurface. The function $a^2 E_{ij}$ specifies the spatial metric on constant-time hypersurfaces. The physical interpretation of the metric functions will be discussed more fully in Section 3.2.

The functions B_i and E_{ij} can be decomposed as follows [27]. The shift function (indeed any vector function in a constant-curvature space) can be written as the sum of the gradient of a scalar and a transverse (solenoidal) vector,

$$B_i = B_{|i} + S_i, \tag{3.4}$$

where the transverse condition

$$S_i^{\ |i} = 0 \tag{3.5}$$

says that S_i has no part that transforms like a scalar. (Here the symbol $|i$ designates a covariant derivative with respect to the homogeneous background spatial metric γ_{ij}. Thus in the case of Cartesian coordinates in a spatially flat background, this covariant derivative reduces to a partial derivative. Also, in Eq. (3.5) and henceforth indices on spatial vectors and tensors are raised and lowered with the background spatial metric γ_{ij} or γ^{ij}, as is conventional, rather than with the perturbed spatial metric $a^2 E_{ij}$.) Thus the three components of B_i are decomposed into a single scalar component and two independent transverse vector components. Next the symmetric three-tensor E_{ij} can be written [27]

$$E_{ij} = (1 - 2\psi)\gamma_{ij} + 2E_{|ij} + 2F_{(i|j)} + h_{ij}, \tag{3.6}$$

where ψ and E are scalar functions, F_i is a transverse vector, and the symmetric tensor h_{ij} is transverse and traceless (TT), *i.e.*.

$$h_{ij}^{\ \ |j} = 0, \qquad h_i^{\ i} = 0. \tag{3.7}$$

The conditions (3.7) mean that no parts of h_{ij} transform as vectors or as scalars. Thus the six components of E_{ij} are decomposed into two scalar components, the two components of the transverse vector, and the two independent components of the TT tensor.

We can now make the conventional classification of metric perturbations into parts derived from scalars, vectors, and tensors. First, collecting terms derived from scalar functions, the "scalar" perturbation line element is

$$ds^2_{(\mathrm{s})} = a^2(\eta)\left\{-(1+2\phi)d\eta^2 + 2B_{|i}dx^i d\eta + \left[(1-2\psi)\gamma_{ij} + 2E_{|ij}\right]dx^i dx^j\right\}. \tag{3.8}$$

Note that the (standard) notation is perhaps confusing: the scalar perturbation line element is *derived* from scalar functions alone, but does contain the vector and tensor parts $B_{|i}$ and $E_{|ij}$. The scalar line element contains four independent functions.

Similarly, collecting terms derived from vectors, the vector-perturbed line element is

$$ds^2_{(\mathrm{v})} = a^2(\eta)\left[-d\eta^2 + 2S_i dx^i d\eta + \left(\gamma_{ij} + 2F_{(i|j)}\right) dx^i dx^j\right]. \tag{3.9}$$

Four independent functions (two for each transverse vector) determine the vector line element. Finally, the tensor-perturbed line element is

$$ds^2_{(\mathrm{t})} = a^2(\eta)\left[-d\eta^2 + (\gamma_{ij} + h_{ij}) dx^i dx^j\right], \tag{3.10}$$

which is determined by the two independent components of h_{ij}. These correspond to the two independent polarization modes of gravitational waves. The total number of independent functions for scalar, vector, and tensor perturbations is thus ten, in agreement with the original general line element, Eq. (3.2).

Note that the decompositions (3.4) and (3.6) are purely *mathematical* results. The *physical* importance of this classification derives first from the fact that the decompositions (3.4) and (3.6) are unique. Thus any vector or tensor equation such as

$$G_{ij} = 8\pi G T_{ij} \tag{3.11}$$

can be unambiguously decomposed into an equation where all quantities are derived only from scalars,

$$G^{(\mathrm{s})}_{ij} = 8\pi G T^{(\mathrm{s})}_{ij}, \tag{3.12}$$

and likewise for vector- and tensor-derived equations. The classification is useful secondly because in *linearly* constructing a tensor from another tensor using only covariant derivatives and the spatial metric γ_{ij} the decomposition is maintained [26]. For example, if we construct the *linear* perturbation δG_{ij} from the metric tensor, then

$$\delta G^{(\mathrm{s})}_{ij}(g_{\mu\nu}) = \delta G_{ij}(g^{(\mathrm{s})}_{\mu\nu}). \tag{3.13}$$

In words, the scalar-derived part of δG_{ij} will depend only on the scalar-derived parts of g_{00}, g_{0i}, and g_{ij}. Thus when we write the linearly perturbed Einstein equation

$$\delta G_{\mu\nu} = 8\pi G \delta T_{\mu\nu} \tag{3.14}$$

the equations of motion for the scalar parts of the metric, ϕ, B, ψ, and E, will depend only on the scalar-derived part of $\delta T_{\mu\nu}$. Similarly the vector and tensor parts of $g_{\mu\nu}$ will couple only to the vector and tensor parts of $\delta T_{\mu\nu}$, respectively. In particular, scalar fields will only couple linearly to scalar parts of the metric. Importantly this result does not hold beyond first order in the perturbations. Second order terms will couple scalar and vector terms etc.

3.2 Geometrical interpretation of metric functions

A particular choice of the perturbation functions in the general form of the perturbed metric, Eq. (3.2), implicitly implies a choice of coodinates, and in particular a choice of a time coordinate η. We can visualize such a choice of a time coordinate as a foliation of the spacetime into hypersurfaces Σ_η of constant coordinate time. We can also imagine a set of timelike curves with tangents n^μ which are everywhere normal to the Σ_η (these normal curves will in general differ from the curves of constant spatial coordinates x^i). Then the geometrical properties of the spacetime can be fully described by the *intrinsic* curvature of the Σ_η, encapsulated by the spatial part g_{ij} of the metric, and the *extrinsic* curvature $K_{\mu\nu}$, which, as we will see, describes how the normal curves evolve in time. In this section I will evaluate various intrinsic and extrinsic geometrical properties for our perturbed metric, Eq. (3.2). This will be very useful not only to clarify the geometrical meaning of the perturbation functions, but also, as we will see in Section 3.4, to help us apply simplifying gauge conditions.

3.2.1 Intrinsic curvature

The simplest measure of the intrinsic curvature of the hypersurfaces Σ_η is the three-dimensional Ricci scalar ${}^{(3)}\!R$. This can be calculated in the usual way by constructing the Ricci tensor ${}^{(3)}\!R^i{}_j$ from the spatial metric g_{ij}. However, there is an easier way to calculate ${}^{(3)}\!R$ in the case that the metric perturbations $\delta g_{\mu\nu}$, defined by

$$g_{\mu\nu} = {}^0g_{\mu\nu} + \delta g_{\mu\nu}, \tag{3.15}$$

are "small". Here ${}^0g_{\mu\nu}$ is the homogeneous background metric. By "small" I will mean that their products can be ignored, so that a first-order linearized approximation can be made. For a spatially flat background, the scalar curvature ${}^{(3)}\!R$ is equal to the perturbation $\delta{}^{(3)}\!R$ in the curvature. Thus we can take the linearized result Eq. (3.221) that I will demonstrate in Section 3.5.1 and apply it to the three-dimensional case, giving

$${}^{(3)}\!R = \frac{1}{a^2}\left(-\delta g^{|i}{}_i + \delta g^i{}_{j|i}{}^j\right). \tag{3.16}$$

Here I have used the fact that the background Ricci tensor vanishes for a spatially flat background. Also, the factor $1/a^2$ compensates for the fact that I raise indices on spatial vectors with γ_{ij} rather than with g_{ij}. Next, note that by the argument given at the end of Section 3.1, the linearized curvature

${}^{(3)}R$ is a scalar and hence can only be constructed from the scalar metric functions ϕ, ψ, B, and E. Thus ${}^{(3)}R$ vanishes for purely vector or tensor linear perturbations.

I will now calculate ${}^{(3)}R$ for general (linear) scalar metric perturbations δg_{ij}. From the general scalar line element, Eq. (3.8), we have

$$\delta g_{ij} = -2a^2\psi\gamma_{ij} + 2a^2 E_{|ij}, \tag{3.17}$$

$$\delta g^i{}_j = -2\psi\delta^i{}_j + 2E^{|i}{}_j. \tag{3.18}$$

Therefore the trace of the spatial metric perturbation is

$$\delta g \equiv \delta g^i{}_i = -6\psi + 2\nabla^2 E, \tag{3.19}$$

where $\nabla^2 E \equiv E^{|i}{}_i$. The next term needed for ${}^{(3)}R$ is easily calculated to be

$$\delta g^i{}_{j|i}{}^j = -2\nabla^2\psi + 2\nabla^2\nabla^2 E. \tag{3.20}$$

Combining these terms I find the very compact expression

$${}^{(3)}R = \frac{4}{a^2}\nabla^2\psi. \tag{3.21}$$

Thus the Ricci scalar curvature of the spatial hypersurfaces Σ_η is completely determined (to linear order) by the metric function ψ.

3.2.2 Extrinsic curvature

Above I described the set of *normal curves*, curves with tangents n^μ everywhere normal to the hypersurfaces Σ_η. As I stated briefly, there exists a tensor $K_{\mu\nu}$, called the extrinsic curvature, which completes (with the intrinsic curvature g_{ij}) the description of the curvature of the full spacetime. The extrinsic curvature is defined by

$$K_{\nu\mu} \equiv h^\rho{}_\nu n_{\mu;\rho}, \tag{3.22}$$

where $h^\rho{}_\nu$ is the projection tensor onto the subspace orthogonal to n^μ,

$$h^\mu{}_\nu \equiv \delta^\mu{}_\nu + n^\mu n_\nu. \tag{3.23}$$

A very short calculation gives

$$K_{\nu\mu} = n_{\mu;\nu} + a_\mu n_\nu, \tag{3.24}$$

where the *acceleration* a_μ is defined by

$$a_\mu \equiv n_{\mu;\rho}n^\rho \tag{3.25}$$

and I have assumed that the timelike vector field n^μ is normalized, *i.e.* $n^\mu n_\mu = -1$. Next we can decompose the tensor $K_{\mu\nu}$ (as indeed any tensor) into an antisymmetric part, the *twist*,

$$K_{[\mu\nu]} \equiv \omega_{\nu\mu}; \tag{3.26}$$

and a symmetric part,

$$K_{(\mu\nu)} \equiv \sigma_{\mu\nu} + \frac{1}{3}\theta h_{\mu\nu}. \tag{3.27}$$

Here I have further decomposed the symmetric part into a traceless part, the *shear* $\sigma_{\mu\nu}$, and a trace part proportional to the *expansion* θ defined by

$$\theta \equiv n^\mu{}_{;\mu} = K^\mu{}_\mu. \tag{3.28}$$

Note that when the integral curves of n^μ are everywhere normal to a set of hypersurfaces, as is true in the case we are considering here, then it can be shown [17] that the twist vanishes, *i.e.* $\omega_{\mu\nu} = 0$. Thus we can write the extrinsic curvature as

$$K_{\mu\nu} = K_{\nu\mu} = \sigma_{\mu\nu} + \frac{1}{3}\theta h_{\mu\nu}. \tag{3.29}$$

To decipher the physical meaning of $K_{\mu\nu}$, consider a displacement vector ξ^μ from some particular normal curve. Suppose that ξ^μ is *Lie dragged* along the field n^μ, *i.e.* $\mathcal{L}_n \xi^\mu = 0$ (Lie derivatives are introduced in Section 3.4.1). That is, ξ^μ and n^μ are coordinate vector fields. Then

$$n^\nu \xi^\mu{}_{;\nu} = \xi^\nu n^\mu{}_{;\nu} = \left(K^\mu{}_\nu - a^\mu n_\nu\right)\xi^\nu. \tag{3.30}$$

In words, the rate of change of ξ^μ along the direction n^ν (*i.e.* the failure of ξ^μ to be parallelly transported along n^ν) is determined by the "map" $K^\mu{}_\nu - a^\mu n_\nu$. Consider now a bundle of normal curves near some fiducial normal curve. The Lie dragged displacements from the fiducial curve to the other curves in the bundle evolve according to Eq. (3.30). Thus the bundle's volume expansion rate will be θ, and it will experience shear if $\sigma_{\mu\nu} \neq 0$. The time components of the displacements will change if $a_\mu \neq 0$. (If we had not been considering hypersurface-normal curves, the bundle would twist about the fiducial curve if $\omega_{\mu\nu} \neq 0$.)

Now the task remaining is to evaluate the geometrical quantities of expansion, shear, and acceleration for the general perturbed metric we have been considering. It will be helpful to introduce a few more geometrical objects. First consider a vector field t^μ defined to be tangent to curves of constant spatial coordinates. Thus t^μ has contravariant components in our chart

$$t^\mu = (1, 0). \tag{3.31}$$

Next define the *shift vector* N^μ to have the entirely spatial contravariant components

$$N^\mu = (0, N^i). \tag{3.32}$$

Finally define the *lapse function* N such that

$$t^\mu = Nn^\mu + N^\mu. \tag{3.33}$$

We can easily solve for the lapse and shift in terms of n^μ, with the result

$$N = \frac{1}{n^0}, \qquad N^i = -Nn^i. \tag{3.34}$$

The shift vector gives the spatial deviation between a constant-coordinate curve and a normal curve. The lapse function specifies the ratio between the values of coordinate time (given by t^0) and proper time (given by the 0-component of the normalized n^μ) along a normal worldline between neighbouring hypersurfaces.

Now I will evaluate the extrinsic curvature in terms of the lapse and shift. Very straightforward calculations give

$$K_{\mu\nu} = h^\rho{}_\mu n_{\nu;\rho} = \frac{1}{2}\left(n^\lambda{}_{;\mu} h_{\lambda\nu} + n^\lambda{}_{;\nu} h_{\lambda\mu} + n^\lambda h_{\mu\nu;\lambda}\right) \tag{3.35}$$

$$= \frac{1}{2N}\left[(Nn^\lambda)_{;\mu} h_{\lambda\nu} + (Nn^\lambda)_{;\nu} h_{\lambda\mu} + Nn^\lambda h_{\mu\nu;\lambda}\right] \tag{3.36}$$

$$= \frac{1}{2N}\left(\mathcal{L}_t h_{\mu\nu} - \mathcal{L}_N h_{\mu\nu}\right). \tag{3.37}$$

The first line involves simple algebra while the final follows from the definition of the Lie derivative, Eq. (3.102), and its linearity. Finally, we can write the spatial components of the extrinsic curvature in our chart as

$$K_{ij} = \frac{1}{2N}\left(t^\lambda h_{ij,\lambda} - N^k{}_{,i} h_{kj} - N^k{}_{,j} h_{ki} - N^k h_{ij,k}\right) \tag{3.38}$$

$$= \frac{1}{2N}\left(g'_{ij} - 2N_{(i|j)}\right) \tag{3.39}$$

Here I have used the facts that $h_{ij} = g_{ij}$ and $t^\lambda = (1, 0)$.

At last I am ready to evaluate the extrinsic curvature for linear scalar perturbations in our chart, which is defined by the general scalar metric (3.8). I will need the first-order expressions

$$g^{00} = -\frac{1}{a^2}(1 - 2\phi), \qquad g^{0i} = \frac{1}{a^2}B^{|i}, \qquad g^{ij} = \frac{1}{a^2}\left[(1 + 2\psi)\gamma^{ij} - 2E^{|ij}\right]. \tag{3.40}$$

The defining orthogonality condition

$$n_\mu X^\mu = 0 \tag{3.41}$$

for any contravariant vector $X^\mu = (0, X^i)$ lying in Σ_η implies that $n_i = 0$ in our chart. Thus the relations $n^\mu = g^{\mu\nu} n_\nu$ and $n^\mu n_\mu = -1$ give

$$n_\mu = (-a(1+\phi), 0), \qquad n^\mu = \frac{1}{a}\left((1-\phi), -B^{|i}\right). \tag{3.42}$$

Hence

$$N = a(1+\phi), \qquad N_i = a^2 B_{|i}. \tag{3.43}$$

Inserting the spatial part of the perturbed scalar metric Eq. (3.8) into the expression (3.39) for the extrinsic curvature, I find to lowest order

$$K_{ij} = a\left\{\mathcal{H}\left[(1-\phi-2\psi)\gamma_{ij} + 2E_{|ij}\right] - \psi'\gamma_{ij} + (E'-B)_{|ij}\right\}. \tag{3.44}$$

Calculating the trace and carefully retaining all first-order terms, I find for the expansion

$$\theta = h^{\rho\mu} n_{\mu;\rho} = g^{ij} K_{ij} \tag{3.45}$$

$$= 3H(1-\phi) - 3\dot{\psi} + \frac{1}{a^2}\nabla^2[a(E'-B)]. \tag{3.46}$$

Notice that, as expected, the volume expansion rate is a first-order perturbation from the background rate, $3H$. The spatial components of the shear are

$$\sigma_{ij} = K_{ij} - \frac{1}{3}\theta g_{ij} \tag{3.47}$$

$$= a(E'-B)_{|ij} - \frac{1}{3}\nabla^2[a(E'-B)]\gamma_{ij}. \tag{3.48}$$

To work out the acceleration, I need to calculate one perturbed connection coefficient,

$$\Gamma^0_{i0} = \phi_{|i} + \mathcal{H} B_{|i}. \tag{3.49}$$

Then I find

$$a_i = u_{i;\mu} u^\mu = \phi_{|i}. \tag{3.50}$$

The relation $a^\mu u_\mu = 0$ then leads to

$$a_0 = 0, \tag{3.51}$$

to first order.

To summarize these results, I have considered the normal curves to the constant-time hypersurfaces Σ_η for the completely general scalar-perturbed metric,

$$ds^2_{(\mathrm{s})} = a^2(\eta)\left\{-(1+2\phi)d\eta^2 + 2B_{|i}dx^i d\eta + \left[(1-2\psi)\gamma_{ij} + 2E_{|ij}\right]dx^i dx^j\right\}. \tag{3.52}$$

I found, to first order, a perturbed volume expansion rate

$$\delta\theta = -3H\phi - 3\dot{\psi} + \frac{1}{a^2}\nabla^2\sigma \tag{3.53}$$

and traceless shear tensor

$$\sigma_{ij} = \sigma_{|ij} - \frac{1}{3}\nabla^2\sigma\gamma_{ij}, \tag{3.54}$$

where I define the shear scalar σ by

$$\sigma \equiv a(E' - B). \tag{3.55}$$

The acceleration is

$$a_\mu = (0, \phi_{|i}). \tag{3.56}$$

Repeating these calculations for the general vector line element, Eq. (3.9), I obtain to linear order

$$\delta\theta = a_\mu = 0, \tag{3.57}$$

and shear tensor

$$\sigma_{ij} = \sigma_{(i|j)} \tag{3.58}$$

where

$$\sigma_i \equiv a(F_i' - S_i). \tag{3.59}$$

Finally, for linear tensor perturbations [Eq. (3.10)] I obtain

$$\delta\theta = a_\mu = 0, \tag{3.60}$$

and shear tensor

$$\sigma_{ij} = \frac{a}{2}h_{ij}'. \tag{3.61}$$

3.3 Perturbed energy-momentum tensor

3.3.1 Decomposition of the energy-momentum tensor

Now that I have thoroughly discussed the perturbed metric tensor, it is time to introduce the corresponding perturbed energy-momentum tensor. I will write the completely general arbitrarily perturbed energy-momentum tensor as

$$T_{\mu\nu} = \rho u_\mu u_\nu + Ph_{\mu\nu} + 2q_{(\mu}u_{\nu)} + \pi_{\mu\nu}, \tag{3.62}$$

where $h_{\mu\nu}$ is the spatial projection tensor

$$h_{\mu\nu} \equiv g_{\mu\nu} + u_\mu u_\nu \tag{3.63}$$

and u_μ is some (arbitrary for now) timelike unit vector field. The functions q_μ and $\pi_{\mu\nu}$ are defined to be (and $h_{\mu\nu}$ is by construction) orthogonal to u^μ,

$$q_\mu u^\mu = \pi_{\mu\nu} u^\mu = h_{\mu\nu} u^\mu = 0. \tag{3.64}$$

Also, the tensor $\pi_{\mu\nu}$ is defined to be symmetric and traceless,

$$\pi_{\mu\nu} = \pi_{\nu\mu}, \qquad \pi^\mu{}_\mu = 0. \tag{3.65}$$

Counting degrees of freedom, we have two for ρ and P, three for the four-vector q_μ constrained by Eq. (3.64), and five for the symmetric $\pi_{\mu\nu}$ constrained by Eqs. (3.64) and (3.65). Thus we have a total of ten independent functions, which proves that we indeed can represent an arbitrary symmetric $T_{\mu\nu}$ by the expression (3.62).

Now, for definiteness, I will choose the field u_μ to be the unit vector field orthogonal everywhere to the hypersurfaces Σ_η of constant coordinate time η. This choice (called the *normal frame* in some references) means that u_μ is identical to the normal field n_μ discussed in Section 3.2.2. That is, according to Eq. (3.42), in our chart, which is specified by the perturbed metric (3.2), we have

$$u_\mu = (-a(1+\phi), 0). \tag{3.66}$$

Thus in the case of a precisely homogeneous universe, if we choose the Σ_η to coincide with the surfaces of homogeneity, then the field u_μ defines the comoving frame discussed in Section 2.2.1. In this case the functions q_μ and $\pi_{\mu\nu}$ must vanish, and we recover the homogeneous energy-momentum tensor, Eq. (2.60).

To understand the physical meaning of the perturbations q_μ and $\pi_{\mu\nu}$ in the general case, recall Section 2.1.2 where I showed that (in Minkowsky spacetime) the components T^{i0} give the momentum density (or energy flux) of the matter and T^{ij} gives the momentum flux, or stress. Evaluating the mixed form of our perturbed energy-momentum tensor, Eq. (3.62), I find in our chart

$$T^0_{\ 0} = -\rho, \qquad T^0_{\ i} = \frac{q_i}{a}, \qquad T^i_{\ j} = P\delta^i{}_j + \pi^i{}_j, \tag{3.67}$$

to first order in the perturbations. Thus q_i determines the momentum density seen by an observer at constant spatial coordinates. The function P again determines the isotropic stress, or pressure, and the tensor $\pi^i{}_j$ gives the anisotropic stress. The function ρ, of course, still gives the energy density.

(Note that the component $T^i_{\ 0}$ in general differs from $T^0_{\ i}$, but we will only need one of these for a complete set.)

Decomposing the energy-momentum tensor into homogeneous background and perturbation, $T^\mu_{\ \nu}(x^\mu) = {}^0T^\mu_{\ \nu}(\eta) + \delta T^\mu_{\ \nu}(x^\mu)$, we have

$$\delta T^0_{\ 0} = -\delta\rho, \qquad \delta T^0_{\ i} = \frac{q_i}{a}, \qquad \delta T^i_{\ j} = \delta P \delta^i_{\ j} + \pi^i_{\ j}. \tag{3.68}$$

The energy density and pressure perturbations are manifestly scalar quantities. The spatial momentum density and anisotropic stress can be decomposed into scalar-, vector-, and tensor-derived parts using Eqs. (3.4) and (3.6), just as I did for the metric perturbation functions. Thus I can decompose the momentum density as

$$q_i \equiv q_{|i} + v_i, \tag{3.69}$$

where the vector part v_i is transverse, $v^i_{\ i} = 0$. The anisotropic stress can be decomposed according to

$$\pi_{ij} \equiv \Pi_{|ij} - \frac{1}{3}\nabla^2\Pi\gamma_{ij} + \pi_{(i|j)} + \tau_{ij}, \tag{3.70}$$

where $\pi^i_{\ |i} = 0$ and the tensor part is TT, *i.e.* $\tau^i_{\ i} = \tau^i_{\ j|i} = 0$, and symmetric, $\tau_{ij} = \tau_{ji}$. The Laplacian term ensures that π_{ij} is traceless. With these definitions, we can decompose the non-trivial components of the perturbed energy-momentum tensor into the scalar-, vector-, and tensor-derived parts

$$\delta T^{0(\mathrm{s})}_{\ i} = \frac{q_{|i}}{a}, \qquad \delta T^{0(\mathrm{v})}_{\ i} = \frac{v_i}{a} \tag{3.71}$$

and

$$\delta T^{i(\mathrm{s})}_{\ j} = \left(\delta P - \frac{1}{3}\nabla^2\Pi\right)\delta^i_{\ j} + \frac{1}{a^2}\Pi^{|i}_{\ \ j}, \qquad \delta T^{i(\mathrm{v})}_{\ j} = \frac{1}{a^2}\gamma^{il}\pi_{(l|j)}, \qquad \delta T^{i(\mathrm{t})}_{\ j} = \tau^i_{\ j}. \tag{3.72}$$

3.3.2 Conservation

I will now work out the components of the covariant conservation law $T^{\mu\nu}_{\ \ ;\mu} = 0$ for the general perturbed energy-momentum tensor (3.62). Taking the covariant derivative of this expression for $T^{\mu\nu}$, I find

$$\begin{aligned} T^{\mu\nu}_{\ \ ;\mu} &= (\rho+P)_{,\mu}u^\mu u^\nu + (\rho+P)\left(u^\mu_{\ ;\mu}u^\nu + u^\mu u^\nu_{\ ;\mu}\right) + P_{,\mu}g^{\mu\nu} \\ &\quad + 2q^{(\mu}_{\ \ ;\mu}u^{\nu)} + 2q^{(\mu}u^{\nu)}_{\ \ ;\mu} + \pi^{\mu\nu}_{\ \ ;\mu} = 0. \end{aligned} \tag{3.73}$$

The energy conservation law will be determined by the component parallel to u_ν, which is

$$u_\nu T^{\mu\nu}_{\ \ ;\mu} = -\rho_{,\mu}u^\mu - (\rho+P)u^\mu_{\ ;\mu} - q^\mu_{\ ;\mu} - q^\nu u^\mu u_{\nu;\mu} - \pi^{\mu\nu}u_{\nu;\mu} = 0. \tag{3.74}$$

Here I have used the fact that $u_\nu u^\nu = -1$, which implies that $u_\nu u^\nu{}_{;\mu} = 0$, and the defining orthogonality conditions $q^\nu u_\nu = \pi^{\mu\nu} u_\nu = 0$. I will now proceed to evaluate the terms in Eq. (3.74) in terms of the matter and metric variables alone. First, recall from Section 3.2.2 that the expansion of the set of integral curves of the vector field u^μ is, to first order,

$$\theta = u^\mu{}_{;\mu} = 3H(1-\phi) - 3\dot{\psi} + \frac{1}{a^2}\nabla^2\sigma. \tag{3.75}$$

Similarly, the acceleration is

$$a_\nu = u^\mu u_{\nu;\mu} = (0, \phi_{,i}), \tag{3.76}$$

so that

$$q^\nu u^\mu u_{\nu;\mu} = 0 \tag{3.77}$$

to first order in the perturbations. Also, using Eq. (3.24) I can write

$$\begin{aligned} \pi^{\mu\nu} u_{\nu;\mu} &= \pi^{\mu\nu}(K_{\mu\nu} - a_\nu u_\mu) && (3.78)\\ &= \pi^{ij} K_{ij}, && (3.79) \end{aligned}$$

since to first order $\pi^{\mu\nu}$ is purely spatial. Therefore, inserting the explicit expression (3.44) for the extrinsic curvature K_{ij}, and using the tracelessness of $\pi^{\mu\nu}$, I find

$$\pi^{\mu\nu} u_{\nu;\mu} = 0 \tag{3.80}$$

to first order in the perturbations. The term $q^\mu{}_{;\mu}$ in Eq. (3.74) is easy to evaluate explicitly:

$$q^\mu{}_{;\mu} = q^\mu{}_{,\mu} + \Gamma^\mu_{\mu l} q^l = q^i{}_{,i} = \frac{1}{a^2}\nabla^2 q \tag{3.81}$$

to first order, since q^μ is spatial and to lowest order the connection coefficient vanishes according to Eq. (2.73).

Continuing with the evaluation of the terms in Eq. (3.74), I next decompose the exact energy density into a background and a perturbation according to

$$\rho(t, x^i) = {}^0\rho(t) + \delta\rho(t, x^i) \tag{3.82}$$

and similarly for the pressure. Then

$$\begin{aligned} \rho_{,\mu} u^\mu + (\rho + P) u^\mu{}_{;\mu} &= {}^0\rho_{,0} u^0 + ({}^0\rho + {}^0P)\theta + \delta\rho_{,\mu} u^\mu + (\delta\rho + \delta P)\theta && (3.83)\\ &= ({}^0\rho + {}^0P)\left(-3\dot{\psi} + \frac{1}{a^2}\nabla^2\sigma\right) + \delta\dot{\rho} + 3H(\delta\rho + \delta P). \end{aligned}$$

Here I have used Eq. (3.42) for the component u^0 and I have applied the background energy conservation law, Eq. (2.65). Finally, I will now drop the background superscript ^{0}, and combine all of these results to write the first order perturbed energy conservation law as

$$\delta\dot{\rho} + 3H(\delta\rho + \delta P) + (\rho + P)\left(-3\dot{\psi} + \frac{1}{a^2}\nabla^2\sigma\right) + \frac{1}{a^2}\nabla^2 q = 0. \tag{3.84}$$

Note that, as expected, only scalar matter and metric functions enter into this expression.

Next I will evaluate the spatial part of the energy-momentum conservation law by multiplying Eq. (3.73) by the spatial projection tensor $h_{\lambda\nu}$. This will give the momentum conservation law, which of course was trivially satisfied in the homogeneous case in Section 2.2.3. Combining expressions (3.63) and (3.73), we have

$$\begin{aligned} h_{\lambda\nu}T^{\mu\nu}{}_{;\mu} = (\rho + P)_{,\mu}u^\mu u_\lambda + (\rho + P)\left(u^\mu{}_{;\mu}u_\lambda + u^\mu u_{\lambda;\mu}\right) + P_{,\lambda} \\ + q^\mu{}_{;\mu}u_\lambda + q^\mu u_{\lambda;\mu} + q_{\lambda;\mu}u^\mu + q_\lambda u^\mu{}_{;\mu} + \pi^\mu{}_{\lambda;\mu} + u_\lambda u_\nu T^{\mu\nu}{}_{;\mu} = 0. \end{aligned} \tag{3.85}$$

Next, evaluating this expression for the spatial component $\lambda = i$, each term proportional to u_λ drops out because of Eq. (3.66). We are left with the exact expression

$$h_{i\nu}T^{\mu\nu}{}_{;\mu} = (\rho + P)a_i + \delta P_{,i} + q^j K_{ij} + q_{i;\mu}u^\mu + q_i\theta + \pi^\mu{}_{i;\mu} = 0. \tag{3.86}$$

Here I have used Eq. (3.24) and the fact that q^μ is purely spatial. As expected, this equation reduces to the trivial $0 = 0$ when all perturbations vanish. The terms in this last expression can be easily calculated to first order using the explicit expressions for a_i, θ, and K_{ij} from Section 3.2.2, and by evaluating the covariant derivatives. The results are

$$q^j K_{ij} = Hq_i, \tag{3.87}$$

$$q_{i;\mu}u^\mu = \dot{q}_i - Hq_i, \tag{3.88}$$

$$\pi^\mu{}_{i;\mu} = \pi^j{}_{i,j} = \frac{1}{a^2}\nabla^2\left(\frac{2}{3}\Pi_{,i} + \frac{1}{2}\pi_i\right). \tag{3.89}$$

For the last expression, I used the anisotropic stress decomposition (3.70) and the transverseness of the vector and tensor parts of the anisotropic stress, π_i and τ_{ij}. Combining these results, the first order momentum conservation equation becomes

$$(\rho + P)\phi_{,i} + \delta P_{,i} + \dot{q}_i + 3Hq_i + \frac{1}{a^2}\nabla^2\left(\frac{2}{3}\Pi_{,i} + \frac{1}{2}\pi_i\right) = 0. \tag{3.90}$$

Note that in this equation ρ and P can be considered background quantities without changing the result to first order. Finally, this equation can be decomposed into scalar- and vector-derived parts (the tensor part clearly vanishes) using the momentum density decomposition [Eq. (3.69)] $q_i = q_{,i} + v_i$:

$$\dot{q} + 3Hq + (\rho + P)\phi + \delta P + \frac{2}{3a^2}\nabla^2\Pi = 0, \qquad \text{scalar part,} \tag{3.91}$$

$$\dot{v}_i + 3Hv_i + \frac{1}{2a^2}\nabla^2\pi_i = 0, \qquad \text{vector part.} \tag{3.92}$$

In writing the scalar equation, I have integrated an equation involving spatial gradients. The resulting arbitrary spatially constant function $C(t)$ can be absorbed into q, since the physical momentum density is determined by the gradient $q_{,i}$.

3.4 Gauge transformations

3.4.1 General form

Why consider gauge transformations?

The tensor transformation law

$$\widetilde{S}_{\mu\nu}(\tilde{x}^\rho) = \frac{\partial x^\kappa}{\partial \tilde{x}^\mu}\frac{\partial x^\lambda}{\partial \tilde{x}^\nu} S_{\kappa\lambda}(x^\rho) \tag{3.93}$$

gives the new value of an arbitrary tensor $S_{\mu\nu}$ after the coordinate change $x^\mu \to \tilde{x}^\mu$, but with the tensor evaluated at the same physical spacetime event. Thus the law describes how the components of the tensor transform. In the case of a scalar, the transformation law therefore reduces to the trivial

$$\tilde{\phi}(\tilde{x}^\mu) = \phi(x^\mu), \tag{3.94}$$

i.e. the scalar has the same value at the event irrespective of the choice of coordinate basis.

We can, of course, ask about the transformation properties of tensors evaluated at different events. In fact, these transformation properties will be extremely important in actually solving Einstein's equation. Because of its general covariance, we cannot expect to find unique solutions to Einstein's equation: if metric tensor $g_{\mu\nu}(x)$ is a solution for energy-momentum tensor $T_{\mu\nu}(x)$, then so must $\tilde{g}_{\mu\nu}(x)$ be a solution for $\widetilde{T}_{\mu\nu}(x)$ after coordinate change $x^\mu \to \tilde{x}^\mu$. This ability to start with any given solution and generate "new solutions" with the same physical content corresponds to the gauge freedom of general relativity, and the associated coordinate transformations are called gauge transformations. A general gauge transformation consists of four arbitrary functions, namely the four components of $(\tilde{x}^\mu - x^\mu)$. This gauge freedom will mean that only six of the ten degrees of freedom in $g_{\mu\nu}$ are "physical" and the rest correspond to "gauge modes". This is consistent with the fact that the four relations (the contracted Bianchi identity or energy-momentum conservation)

$$G^\mu{}_{\nu;\mu} = 0 \tag{3.95}$$

reduce the number of independent components of Einstein's equation from ten to six. Indeed, in Section 2.1.2 general covariance was used to derive the Bianchi identity.

In the context of linear perturbation theory, it will be very useful to understand how tensors change under infinitesimal gauge transformations. In fact, since any perturbation is only defined to within such a transformation, we are *forced* to consider such transformations. To understand this, recall that we are interested in the behaviour of some perturbation $\delta S(t, x^i)$ defined through

$$S(t, x^i) = {}^0S(t) + \delta S(t, x^i), \tag{3.96}$$

where $S(t, x^i)$ is some exact quantity and ${}^0S(t)$ is a (fictional) homogeneous background quantity. We perform this split with the expectation that the equations of motion for both the background and the perturbation will be simpler than the full equations for the exact quantity (because of symmetries of the background and perhaps an approximation of linearity for the perturbations). In a non-relativistic theory, this procedure poses no conceptual problems whatsoever: We find solutions to the perturbation and background equations and then trivially sum them to recover the dynamics of the exact quantity. However, in general relativity Einstein's equation tells us that any perturbation $\delta S(t, x^i)$ must be accompanied by a perturbation of the metric (if S was not already the metric!). Thus the exact and background spacetimes *are not the same*, and hence there is no unique way of "mapping" or identifying events in one with events in the other. In particular, we cannot say whether the coordinates on the lhs of Eq. (3.96) refer to the same event as the coordinates on the rhs. Therefore the perturbation $\delta S(t, x^i)$ is not uniquely defined. Conversely, if we do solve the background and perturbation dynamics, we cannot unambiguously sum them to recover the evolution of the exact quantity S. The best we can say is that

$$\widetilde{S}(t, x^i) = {}^0S(t) + \delta S(t, x^i), \tag{3.97}$$

where $\widetilde{S}(t, x^i)$ is related to $S(t, x^i)$ by some coordinate transformation; this is just a gauge transformation as discussed in the previous paragraph in the context of general covariance. In the context of linear perturbation theory, we are helped somewhat by the fact that the gauge transformations we allow in Eq. (3.97) must be appropriately "small" in order not to destroy the smallness of the perturbation $\delta S(t, x^i)$. However, we still have four arbitrary (but small) functions with which to gauge transform.

To summarize, we must determine how quantities change under gauge transformations since a perturbation is only defined up to such a transformation. While this may seem to throw cosmological perturbation theory into a state of hopeless ambiguity, in fact we will soon see that the gauge freedom will be very useful in simplifying the form of the dynamical equations if we choose the gauge appropriately. In the remainder of this subsection I will thus derive the tensor transformation law under infinitesimal gauge transformations, with applications to linear perturbation theory in mind.

Change of a tensor

I wish to determine the change of a tensor under an infinitesimal gauge transformation

$$x^\mu \longrightarrow \tilde{x}^\mu = x^\mu - \xi^\mu, \tag{3.98}$$

where the four functions ξ^μ are small in a sense to be determined. First, consider the infinitesimal form of the general transformation law, Eq. (3.93),

$$\widetilde{S}_{\mu\nu}(\tilde{x}) = S_{\mu\nu}(x) + \xi^\kappa{}_{,\mu} S_{\kappa\nu} + \xi^\kappa{}_{,\nu} S_{\mu\kappa} + \mathcal{O}(\xi^2). \tag{3.99}$$

We can Taylor expand $\widetilde{S}_{\mu\nu}(\tilde{x})$ to express it in terms of x,

$$\widetilde{S}_{\mu\nu}(\tilde{x}) = \widetilde{S}_{\mu\nu}(x) - \xi^\kappa S_{\mu\nu,\kappa}(x) + \mathcal{O}(\xi^2). \tag{3.100}$$

Combining these two expressions gives

$$\widetilde{S}_{\mu\nu}(x) = S_{\mu\nu}(x) + \xi^\kappa{}_{;\mu} S_{\kappa\nu} + \xi^\kappa{}_{;\nu} S_{\mu\kappa} + \xi^\kappa S_{\mu\nu;\kappa} + \mathcal{O}(\xi^2). \tag{3.101}$$

In this expression, in replacing the partial derivatives with covariant derivatives the terms involving connection coefficients have cancelled. Defining the Lie derivative of a second rank covariant tensor by

$$\mathcal{L}_\xi S_{\mu\nu} = \xi^\kappa{}_{;\mu} S_{\kappa\nu} + \xi^\kappa{}_{;\nu} S_{\mu\kappa} + \xi^\kappa S_{\mu\nu;\kappa}, \tag{3.102}$$

we can finally write

$$\widetilde{S}_{\mu\nu}(x) = S_{\mu\nu}(x) + \mathcal{L}_\xi S_{\mu\nu}(x) + \mathcal{O}(\xi^2). \tag{3.103}$$

The Lie derivative can be similarly defined for other types of tensors and expressions analogous to Eq. (3.103) will hold. In particular, for a scalar χ the Lie derivative becomes simply

$$\mathcal{L}_\xi \chi = \xi^\mu \chi_{,\mu}, \tag{3.104}$$

so that

$$\tilde{\chi}(x) = \chi(x) + \xi^\mu \chi_{,\mu} + \mathcal{O}(\xi^2) \tag{3.105}$$

as expected.

We can now re-express the statement on gauge freedom from above using the Lie derivative. If metric $g_{\mu\nu}(x)$ and energy-momentum tensor $T_{\mu\nu}(x)$ solve Einstein's equation, then so will

$$\tilde{g}_{\mu\nu}(x) = g_{\mu\nu}(x) + \mathcal{L}_\xi g_{\mu\nu}(x) + \mathcal{O}(\xi^2) \tag{3.106}$$

and

$$\widetilde{T}_{\mu\nu}(x) = T_{\mu\nu}(x) + \mathcal{L}_\xi T_{\mu\nu}(x) + \mathcal{O}(\xi^2), \tag{3.107}$$

where ξ^μ is an infinitesimal gauge transformation. Similarly, in the context of perturbation theory, the statement above that a perturbation can only be defined up to a gauge transformation of the exact quantity can be expressed now as the statement that to any perturbation δS can be added a Lie derivative,

$$\delta S \to \delta\widetilde{S} = \delta S + \mathcal{L}_\xi S + \mathcal{O}(\xi^2), \tag{3.108}$$

without any physical consequences, as long as *all* matter and metric quantities are similarly transformed, *i.e.* all quantities are presented in the same gauge.

To close this subsection, I will explicate the sense in which the functions ξ^μ must be "small" if we are to approximate a gauge transformation with the Lie derivative. In order that we can ignore the $\mathcal{O}(\xi^2)$ terms in Eq. (3.99), the derivatives $\xi^\mu{}_{,\nu}$ must be much smaller than unity. This means that *the size of ξ^μ must be much smaller than the characteristic length scale of variation of $\xi^\mu(x)$*. Similarly, if we are to ignore the $\mathcal{O}(\xi^2)$ terms in the Taylor expansion (3.100), then the second order term proportional to $\xi^2 S_{\mu\nu,\kappa\lambda}$ must be much smaller than the first order term. That is, *the size of ξ^μ must be much smaller than the characteristic length scale of curvature of $S_{\mu\nu}(x)$*.

3.4.2 Gauge transformations of the metric

Returning to the perturbed metric tensor, the gauge ambiguity mentioned above can be removed by arbitrarily adopting a particular coordinate system, *i.e.* by fixing the four functions ξ^μ. This will amount to imposing four coordinate conditions on $g_{\mu\nu}$, which together with the six independent components of Einstein's equation will give an unambiguous solution.

To begin, I will describe the effect of infinitesimal gauge transformations on scalar, vector, and tensor metric perturbations. Using the vector decomposition, Eq. (3.4), an arbitrary coordinate change can be written

$$\xi^\mu = (\xi^0, \xi^{|i} + \xi^i_{\mathrm{tr}}), \tag{3.109}$$

where ξ^0 and ξ are scalar functions and $\xi^i_{\rm tr}$ is a transverse spatial vector field. (Note that I define

$$\xi^{|i} = \gamma^{ij}\xi_{|j}, \tag{3.110}$$

where $\xi_{|j}$ is the covariant derivative with respect to $x^j = x^\mu|_{\mu=j}$.) For the metric tensor, whose covariant derivative vanishes, the Lie derivative Eq. (3.102) reduces to

$$\mathcal{L}_\xi g_{\mu\nu} = 2\xi_{(\mu;\nu)}, \tag{3.111}$$

although the non-covariant form

$$\mathcal{L}_\xi g_{\mu\nu} = 2\xi^\kappa{}_{,(\mu}g_{\nu)\kappa} + \xi^\kappa g_{\mu\nu,\kappa} \tag{3.112}$$

is slightly more convenient for these calculations. Note that the change $\mathcal{L}_\xi g_{\mu\nu}$ in the metric tensor under an infinitesimal gauge transformation is a linear function of covariant derivatives of ξ^μ and hence, by the result stated at the end of Section 3.1, we expect that a scalar gauge transformation

$$\xi^\mu = (\xi^0, \xi^{|i}) \tag{3.113}$$

will change only the scalar-derived terms in the perturbed metric, a vector transformation

$$\xi^\mu = (0, \xi^i_{\rm tr}) \tag{3.114}$$

will change only the vector part of the metric, and no infinitesimal gauge transformation will change the tensor part of the metric. While the metric will change under all types of linear gauge transformations, we will see in Section 3.4.3 that due to the homogeneity of the background, only the temporal scalar transformation ξ^0 will effect the quantities that appear in the dynamical equations. Nevertheless, I will exhibit the effect of all gauge transformations in this section.

Anticipating Section 3.5, I will consider the case where the perturbed metric functions and their derivatives are small in the sense that the products

$$\xi^\mu{}_{,\nu}\phi \qquad \text{and} \qquad \xi^\mu\phi_{,\nu} \tag{3.115}$$

can be ignored, and similarly for the remaining metric perturbation functions. This means that to linear order we can consider the metric in the expression (3.112) to be the homogeneous background FRW metric, Eq. (2.54). I will use the comoving, conformal time background chart specified by that metric, and assume spatial flatness and Cartesian coordinates so that spatial covariant derivatives become partial derivatives.

I will start by calculating the effect of an infinitesimal vector gauge transformation, Eq. (3.114), on the metric. The spatial part of the metric tensor, g_{ij}, changes by

$$\begin{aligned}\mathcal{L}_\xi g_{ij} &= 2\xi^{\kappa}{}_{,(i}g_{j)\kappa} + \xi^{\kappa} g_{ij,\kappa} &(3.116)\\ &= 2\xi^{k}{}_{,(i}g_{j)k} &(3.117)\\ &= 2a^2\xi^{k}{}_{,(i}\gamma_{j)k} &(3.118)\\ &= 2a^2\xi_{(i,j)}. &(3.119)\end{aligned}$$

Here I've used $g_{ij,k} = 0$ for the homogeneous background and $\xi^0 = 0$ for a vector transformation. Now from Section 3.1, the completely general spatial part of the perturbed metric tensor is

$$g_{ij} = a^2\left[(1-2\psi)\gamma_{ij} + 2E_{,ij} + 2F_{(i,j)} + h_{ij}\right], \tag{3.120}$$

for scalars ψ and E, transverse vector F_i, and TT tensor h_{ij}. Thus we see from Eq. (3.119) and from the uniqueness of the decomposition (3.120) that the effect of the vector gauge transformation on the spatial metric consists of a shift in the vector F_i,

$$F_i \to \widetilde{F}_i = F_i + \xi_{i,\mathrm{tr}}, \tag{3.121}$$

while ψ, E, and h_{ij} are unchanged, as expected.

By a very similar calculation, the time-space component of the metric tensor changes under a vector gauge transformation by

$$\mathcal{L}_\xi g_{0i} = a^2\xi_i' \tag{3.122}$$

where $\xi_i' \equiv \xi_{i,0} = \partial\xi_i/\partial\eta$. Thus the only change in the time-space part of the general perturbed metric

$$g_{0i} = a^2(B_{,i} + S_i) \tag{3.123}$$

is in the vector part

$$S_i \to \widetilde{S}_i = S_i + \xi'_{i,\mathrm{tr}}. \tag{3.124}$$

For a vector gauge transformation

$$\mathcal{L}_\xi g_{00} = 0 \tag{3.125}$$

so ϕ is unchanged.

Next I will consider the change in the metric due to a scalar gauge transformation, Eq. (3.113). Calculations similar to those for the vector case give

$$\begin{aligned}\mathcal{L}_\xi g_{00} &= -2a^2(\xi^{0\prime} + \mathcal{H}\xi^0), &(3.126)\\ \mathcal{L}_\xi g_{0i} &= a^2(\xi_{,i}' - \xi^0_{,i}), &(3.127)\\ \mathcal{L}_\xi g_{ij} &= 2a^2(\xi_{,ij} + \mathcal{H}\xi^0\gamma_{ij}). &(3.128)\end{aligned}$$

These expressions imply that a scalar gauge transformation changes only the scalar metric functions as follows:

$$\phi \to \tilde{\phi} = \phi + \xi^{0\prime} + \mathcal{H}\xi^0, \tag{3.129}$$
$$B \to \widetilde{B} = B + \xi' - \xi^0, \tag{3.130}$$
$$\psi \to \tilde{\psi} = \psi - \mathcal{H}\xi^0, \tag{3.131}$$
$$E \to \widetilde{E} = E + \xi. \tag{3.132}$$

3.4.3 Gauge transformations of the energy-momentum tensor

We have now seen that the functions appearing in the general perturbed metric (3.2) all change under the arbitrary gauge transformation

$$\xi^\mu = (\xi^0, \xi^{|i} + \xi^i_{\mathrm{tr}}). \tag{3.133}$$

We expect the perturbed energy-momentum tensor to also change under such a transformation. In fact, in this section I will show that, to first order, the mixed form components $\delta T^0_{\ 0}$, $\delta T^0_{\ i}$, and $\delta T^i_{\ j}$ only change under the temporal transformation ξ^0. Thus the corresponding components of the linearized Einstein equation

$$\delta G^\mu_{\ \nu} = 8\pi G \delta T^\mu_{\ \nu} \tag{3.134}$$

will be gauge invariant under linear spatial transformations.

I will calculate the change in the energy-momentum tensor under a gauge transformation by using the mixed form of Eq. (3.102),

$$\mathcal{L}_\xi T^\mu_{\ \nu} = -\xi^\mu_{\ ;\lambda} T^\lambda_{\ \nu} + \xi^\lambda_{\ ;\nu} T^\mu_{\ \lambda} + \xi^\lambda T^\mu_{\ \nu;\lambda} \tag{3.135}$$
$$= -\xi^\mu_{\ ,\lambda} T^\lambda_{\ \nu} + \xi^\lambda_{\ ,\nu} T^\mu_{\ \lambda} + \xi^\lambda T^\mu_{\ \nu,\lambda}. \tag{3.136}$$

I choose this form because the components in our chart contain fewer troublesome appearances of the scale factor a. Inserting the components $T^\mu_{\ \nu}$ from Eq. (3.67) and the general infinitesimal gauge transformation (3.133) into Eq. (3.136), I find

$$\mathcal{L}_\xi T^0_{\ 0} = -\rho' \xi^0, \tag{3.137}$$
$$\mathcal{L}_\xi T^0_{\ i} = -(\rho + P)\xi^0_{\ ,i}, \tag{3.138}$$
$$\mathcal{L}_\xi T^i_{\ j} = P' \xi^0 \delta^i_{\ j}. \tag{3.139}$$

Indeed, as promised, the spatial transformations ξ or ξ^i_{tr} do not appear in these expressions. Note that the spatial transformations *do* appear in the component $\mathcal{L}_\xi T^i_{\ 0}$ (as well as in the covariant or contravariant forms $\mathcal{L}_\xi T_{\mu\nu}$ and $\mathcal{L}_\xi T^{\mu\nu}$). However, only the components $\delta T^0_{\ 0}$, $\delta T^0_{\ i}$, and $\delta T^i_{\ j}$ are needed to write a complete set of components of Einstein's equation (3.134), namely the energy and momentum constraints and the space-space dynamical equation. The expressions (3.137) to (3.139) imply that that complete set of equations is gauge invariant under linear spatial transformations. In particular, this means that the vector (as well as the tensor) part of the equations of motion will be gauge invariant, and the scalar part will depend only on the single degree of gauge freedom ξ^0. Physically, this is due to the homogeneity of the background—as we move infinitesimally along the constant-time hypersurfaces Σ_η, no physical quantity changes at zeroth order.

The gauge transformations of $T^\mu_{\ \nu}$ which I have just calculated can be readily expressed in terms of the matter variables defined in the decompositions (3.68), (3.71), and (3.72), just as I expressed the gauge transformations of $g_{\mu\nu}$ in terms of the metric functions in the previous subsection. Under the arbitrary gauge transformation (3.133), we have

$$\delta\rho \to \delta\tilde{\rho} = \delta\rho + \rho'\xi^0, \tag{3.140}$$

$$\delta P \to \delta\widetilde{P} = \delta P + P'\xi^0, \tag{3.141}$$

$$q \to \tilde{q} = q - a(\rho + P)\xi^0, \tag{3.142}$$

and each of

$$v_i, \quad \Pi, \quad \pi_i, \quad \text{and} \quad \tau^i_{\ j} \tag{3.143}$$

are gauge invariant.

What these results mean is that when we write the complete set of linearized equations,

$$\delta G^0_{\ 0} = 8\pi G\delta T^0_{\ 0}, \qquad \delta G^0_{\ i} = 8\pi G\delta T^0_{\ i}, \qquad \delta G^i_{\ j} = 8\pi G\delta T^i_{\ j}, \tag{3.144}$$

we are guaranteed that only combinations of the metric functions that are spatially gauge invariant will appear. Thus the homogeneity of the background itself reduces the number of *physical* metric degrees of freedom from four to three for the scalar modes, and from four to two for the vector modes. Of course the remaining scalar gauge mode is still not physical, and fixing ξ^0 will reduce the number of independent scalar degrees of freedom to two. The various choices for fixing ξ^0 will be the subject of the next section.

3.4.4 Choice of gauge

I proved in the previous section that the vector- and tensor-derived parts of the linearly perturbed Einstein equation are gauge invariant, while the scalar-derived part depends only on the single gauge function ξ^0, *i.e.* on the choice of time slicing. Considering the gauge transformations (3.121) and (3.124) for the vector functions F_i and S_i, it follows that the vector equation of motion must only contain the gauge-invariant combination

$$\sigma_i = a(F_i' - S_i). \tag{3.145}$$

In Section 3.2.2 I found that the quantity σ_i determines the shear of the normal curves to constant-time hypersurfaces. Similarly, considering the gauge transformations (3.129) to (3.132) for the scalar metric functions, the scalar equation of motion must only contain

$$\phi, \quad \psi, \quad \text{and} \quad \sigma = a(E' - B), \tag{3.146}$$

each of which are gauge invariant under spatial transformations. Recalling Eq. (3.54), the quantity σ determines the shear of the normal curves in the scalar case.

For reference I will summarize here all of the geometrically or physically important metric and matter functions that depend only on the choice of time slicing ξ^0. The metric function ϕ determines the lapse function through Eq. (3.43). The function ψ determines the Ricci curvature scalar of the constant-time hypersurfaces Σ_η through Eq. (3.21). The function σ determines the shear of the normals to the Σ_η, as I just mentioned. These three functions can be combined according to Eq. (3.53) to form the perturbation in the expansion of the normals, $\delta\theta$. The matter side is described by the perturbed energy density, $\delta\rho$, the perturbed pressure, δP, and the momentum density scalar, q.

I will now summarize the gauge transformation properties of each of these seven scalar perturbation functions. The forms of the expressions are cleaner if we write them in terms of the shift in proper time $T \equiv a\xi^0$, rather than conformal time ξ^0. Collecting expressions (3.129) to (3.132), (3.140) to (3.142), and definitions (3.53) and (3.55), I find

$$\tilde{\phi} = \phi + \dot{T}, \tag{3.147}$$
$$\tilde{\psi} = \psi - HT, \tag{3.148}$$
$$\tilde{\sigma} = \sigma + T, \tag{3.149}$$
$$\delta\tilde{\theta} = \delta\theta + \left(3\dot{H} + \frac{1}{a^2}\nabla^2\right) T, \tag{3.150}$$
$$\delta\tilde{\rho} = \delta\rho + \dot{\rho}T, \tag{3.151}$$
$$\delta\widetilde{P} = \delta P + \dot{P}T, \tag{3.152}$$
$$\tilde{q} = q - (\rho + P)T. \tag{3.153}$$

This list of expressions makes it easy to define several physically meaningful choices of time slicing T. Namely, with a suitable choice of T we can set any one of the seven quantities to zero. Thus each such choice will simplify the equations of motion while completely removing the gauge ambiguity (except for synchronous and static curvature gauges; see below). After defining some notation and giving some general results I will next summarize each of these gauge choices.

Notation and general results

I will follow a simple notation convention. Consider any two matter or metric perturbation variables p and r (*e.g.*, $p = \psi$ and $r = \delta\rho$). When written unsubscripted, such a variable will refer to that quantity in an arbitrary gauge, unless otherwise noted. On the other hand, the variable p in a gauge specified by $r = 0$ is given the symbol p_r. (For example, the uniform density gauge curvature perturbation is written $\psi_{\delta\rho}$, or ψ_ρ for short.) Thus $p_p = 0$ for all p. In a gauge defined by $r = 0$, the hypersurfaces Σ_t of constant coordinate time coincide with the hypersurfaces of vanishing r, which I denote Σ_r. Thus $p_r = p|_{\Sigma_r}$.

When p and r transform like

$$\tilde{p} = p + \alpha(t)T, \tag{3.154}$$

$$\tilde{r} = r + \beta(t)T, \tag{3.155}$$

[as most of the important variables do; recall Eqs. (3.147) to (3.153)], some simple results with clear geometrical interpretations apply. The gauge transformation

$$T_{\to p} = -\frac{p}{\alpha} \tag{3.156}$$

results in

$$\tilde{p} = p_p = 0, \tag{3.157}$$

and similarly r is gauged away with the shift

$$T_{\to r} = -\frac{r}{\beta}. \tag{3.158}$$

The shift $T_{\to p}$ is simply the temporal displacement between an arbitrary gauge hypersurface and a vanishing-p hypersurface, Σ_p, and similarly for r. Thus we have

$$p_r = p - \frac{\alpha}{\beta} r \tag{3.159}$$

and

$$r_p = r - \frac{\beta}{\alpha} p. \tag{3.160}$$

It follows that p_r and r_p are proportional,

$$p_r = -\frac{\alpha}{\beta} r_p. \tag{3.161}$$

This result has a simple geometrical interpretation. Both p_r and r_p are proportional to the temporal displacement $T_{p\to r}$ between hypersurfaces Σ_p and Σ_r of vanishing p and r,

$$T_{p\to r} = T_{\to r} - T_{\to p} = \frac{p}{\alpha} - \frac{r}{\beta}. \tag{3.162}$$

Indeed, performing the shift $T_{p\to r}$ from Σ_p, where $p = 0$, to Σ_r, where $p = p_r$, we see [with Eq. (3.154)] that p_r must be proportional to $T_{p\to r}$, and similarly for r_p.

Synchronous gauge

The choice

$$\dot{T} = -\phi, \tag{3.163}$$

that is

$$T = -\int \phi \, dt + C(x^i), \tag{3.164}$$

where $C(x^i)$ is an arbitrary spatial function, results in

$$\tilde{\phi} = 0. \tag{3.165}$$

This choice, known as *synchronous gauge*, is particularly useful in that it simplifies considerably the form of the equations of motion. This is because $\phi = 0$ implies that the lapse function is unperturbed, so that coordinate time coincides with proper time along constant spatial coordinate worldlines.

However, there is a cost to this simplicity. Since synchronous gauge only fixes $\dot{T}$, there remains the *residual* gauge freedom in T represented by the spatial function $C(x^i)$. This can complicate the treatment of solutions, as the residual gauge mode must be tracked and not mistakenly interpreted as a physical mode. Ultimately this residual freedom corresponds to a (spatially dependent) arbitrariness in the choice of time origin, and can in fact be useful in understanding the behaviour of long-wavelength perturbations, as I will show in Chapter 5.

Static curvature gauge

Setting

$$(\dot{HT}) = \dot{\psi}, \tag{3.166}$$

that is

$$T = \frac{\psi + C(x^i)}{H} \tag{3.167}$$

where $C(x^i)$ is an arbitrary spatial function, results in

$$\dot{\tilde{\psi}} = 0 \tag{3.168}$$

and

$$\tilde{\psi} = -C(x^i). \tag{3.169}$$

Since ψ is static with this slicing [and hence, recalling Eq. (3.21), ${}^{(3)}R$ has only the trivial time dependence ${}^{(3)}R \propto a^{-2}$], I will name this gauge *static curvature gauge,* although I am not aware of this gauge discussed in the literature under any name.

As with synchronous gauge, this gauge only fixes a time derivative, and hence is only defined up to an arbitrary residual gauge function $C(x^i)$. However, it will also prove very useful in discussing long-wavelength dynamics in Chapter 5.

Uniform curvature gauge

If we fix the residual gauge function $C(x^i)$ in static curvature gauge to vanish, *i.e.* if we choose

$$T = \frac{\psi}{H}, \tag{3.170}$$

then we obtain precisely

$$\tilde{\psi} = 0 \tag{3.171}$$

and hence

$${}^{(3)}R = 0. \tag{3.172}$$

Thus the constant-time hypersurfaces Σ_t are spatially flat in this gauge, if the background is flat. (In the general case, the Σ_t are of constant scalar curvature.)

Zero-shear or longitudinal gauge

A time slicing specified by

$$T = -\sigma \tag{3.173}$$

yields

$$\tilde{\sigma} = 0. \tag{3.174}$$

This condition has the geometrical interpretation that the normals to the Σ_t do not experience shear. As described below, with a further spatial transformation we can set

$$\widetilde{B} = \widetilde{E} = 0, \tag{3.175}$$

which defines *longitudinal* gauge.

Uniform expansion gauge

Choosing for the time slicing T the solution to the equation

$$\left(3\dot{H} + \frac{1}{a^2}\nabla^2\right) T = -\delta\theta \tag{3.176}$$

results in

$$\delta\tilde{\theta} = 0. \tag{3.177}$$

That is, in this gauge the expansion rate of the normals to the Σ_t is given by the unperturbed value $3H$.

Uniform density gauge

The choice of time slicing

$$T = -\frac{\delta\rho}{\dot{\rho}} \tag{3.178}$$

means that

$$\delta\tilde{\rho} = 0, \tag{3.179}$$

that is in this gauge the hypersurfaces Σ_t are constant-energy-density surfaces. Thus the uniform density gauge (together with the following two gauge choices) has a particularly striking physical meaning. Note, however, that the gauge transformation T becomes singular if the background density becomes stationary, $\dot{\rho} = 0$. I will discuss this situation in Section 3.4.5.

Uniform pressure gauge

We can similarly set the pressure perturbation to zero on the Σ_t,

$$\delta\tilde{P} = 0, \tag{3.180}$$

by performing a gauge transformation with

$$T = -\frac{\delta P}{\dot{P}}. \tag{3.181}$$

This gauge shares with uniform density gauge the potential for singular behaviour.

Comoving gauge

A particularly intuitive gauge choice results from the slicing

$$T = \frac{q}{\rho + P}. \tag{3.182}$$

Then we have

$$\tilde{q} = 0 \tag{3.183}$$

on the constant-time hypersurfaces, so that the momentum density vanishes in our coordinates. This gauge choice is, in this sense, a natural generalization of the comoving coodinates we used in the homogeneous case. It can also become singular.

Spatial gauge choice

Recall again that spatial gauge transformations $\xi^{|i} + \xi^i_{\rm tr}$ do not affect the final equations of motion, and hence only the choice of time slicing T is "physical" in this sense. However, spatial transformations *do* affect the metric functions E and B according to Eqs. (3.130) and (3.132). Therefore in addition to each of the gauge choices described above we can make a further spatial transformation to eliminate either E or B. The advantage to this procedure is that in simplifying the form of the metric it will make metric-based calculations easier.

In particular, according to Eq. (3.132), performing a spatial gauge transformation with

$$\xi = -E \tag{3.184}$$

results in

$$\widetilde{E} = 0. \tag{3.185}$$

Similarly, according to Eq. (3.130), setting instead

$$\xi = \int (\xi^0 - B)\, d\eta + C(x^i), \tag{3.186}$$

where $C(x^i)$ is an arbitrary spatial function, gives

$$\widetilde{B} = 0. \tag{3.187}$$

As in the case of synchronous gauge described above, the arbitrary function $C(x^i)$ represents a residual spatial gauge freedom.

For example, in uniform curvature gauge, where $\tilde{\psi} = 0$, we can further set $\widetilde{E} = 0$ and the scalar line element (3.8) becomes

$$ds^2_{(\rm s)} = a^2(\eta)\left[-(1+2\tilde{\phi})d\eta^2 + 2\widetilde{B}_{|i}dx^i d\eta + \gamma_{ij}dx^i dx^j\right]. \tag{3.188}$$

That is, the spatial 3-metric γ_{ij} is completely unperturbed.

Likewise, in zero-shear gauge, where $\tilde{\sigma} = 0$, we can also set $\widetilde{E} = 0$. Thus, since $\tilde{\sigma} = a(\widetilde{E}' - \widetilde{B})$, we also have

$$\widetilde{B} = 0. \tag{3.189}$$

In this case, zero-shear gauge is commonly referred to as *longitudinal* gauge. According to Eqs. (3.147), (3.148), and (3.173), the scalar metric functions ϕ and ψ become

$$\tilde{\phi} = \phi_\sigma = \phi - \dot{\sigma} \equiv \Phi, \tag{3.190}$$
$$\tilde{\psi} = \psi_\sigma = \psi + H\sigma \equiv \Psi. \tag{3.191}$$

The symbols Φ and Ψ are conventionally used to designate the longitudinal gauge scalar functions. The scalar line element becomes

$$ds^2_{(\mathrm{s})} = a^2(\eta)\left[-(1+2\Phi)d\eta^2 + (1-2\Psi)\gamma_{ij}dx^i dx^j\right] \tag{3.192}$$

in this gauge. This simple form makes longitudinal gauge particularly suited to doing metric calculations. In the absence of anisotropic stress, the longitudinal gauge metric will take an even simpler form, as we will see.

Vector gauge

I proved above that Einstein's equation for linear vector perturbations is gauge invariant (in a homogeneous background). However, just as with the scalar gauge choices listed above, there is a "trick" we can perform to make intermediate calculations with the vector metric simpler. According to Eqs. (3.121) and (3.124), performing a spatial gauge transformation with

$$\xi^i_{\mathrm{tr}} = -F^i \tag{3.193}$$

produces

$$\widetilde{F}_i = 0 \tag{3.194}$$

and

$$\widetilde{S}_i = S_i - F_i'. \tag{3.195}$$

I will call this choice "vector gauge". The metric in this gauge takes the very simple form

$$ds^2_{(\mathrm{v})} = a^2(\eta)\left(-d\eta^2 + 2\widetilde{S}_i dx^i d\eta + \gamma_{ij}dx^i dx^j\right). \tag{3.196}$$

After completing a metric-based derivation of the equation of motion, we can, recalling definition (3.145) and Eq. (3.195), simply replace each occurence of $-a\widetilde{S}_i$ with the gauge-invariant physical variable σ_i representing the shear.

Gauge fixing *vs.* "gauge-invariant variables"

The approach of several works on cosmological perturbation theory [25, 26, 28] is to attempt to construct metric and matter variables that are independent of the choice of gauge. For example, the metric function Ψ, defined in Eq. (3.191),

$$\Psi \equiv \psi_\sigma = \psi + H\sigma, \tag{3.197}$$

is considered such a "gauge-invariant" variable. Indeed, it is very easy to confirm with Eqs. (3.148) and (3.149) that the combination $\psi + H\sigma$ does not change under arbitrary linear gauge transformations. This is to be expected according to the interpretation surrounding Eq. (3.162), under which ψ_σ is simply proportional to the (manifestly gauge-invariant) temporal displacement between uniform curvature and zero shear hypersurfaces.

However, Eq. (3.197) is a prescription for taking an arbitrary gauge perturbation ψ and translating it into zero-shear gauge. Thus the "gauge-invariance" of Ψ can be seen as just a reflection of the fact that this prescription *must* produce the same result on arbitrary initial gauges. In fact, it is not hard to convince oneself that such "gauge-invariant" variables can be constructed simply by specifying the prescription for taking any metric or matter variable and representing it in *any* completely fixed gauge [29, 30]. That is, working with any such "gauge-invariant" variable is completely equivalent to working in some specific gauge—equations of motion, for example, will be identical in either approach. While the interpretation of a variable such as Ψ as a gauge-invariant displacement between hypersurfaces is attractive, not all "gauge-invariant" variables can be interpreted this way (*e.g.*, $\Phi = \phi_\sigma$ cannot). Instead, I find it physically clearer and safer to simply fix all variables according to some gauge condition, and this is the approach I take in this thesis.

3.4.5 A gauge transformation too far?

In this subsection I will make some cautionary remarks on the applicability of gauge transformations. Situations can arise in which a transformation to some particular gauge is singular or ill-defined. Indeed, there is no gauge transformation (infinitesimal or not) that can gauge away density perturbations after they have collapsed sufficiently in the process of forming galactic structure. That is, uniform density gauge is ill-defined at late enough times.

To see why this is so, consider the gauge transformation of some scalar quantity, χ, decomposed as usual into a homogeneous background and a perturbation:

$$\mathcal{L}_\xi \chi = \xi^\mu \chi_{,\mu} = T\dot{\chi} + \mathcal{O}(\xi^i \delta\chi_{,j}). \tag{3.198}$$

[recall Eq. (3.104)]. As above, $T = a\xi^0$ is the proper time displacement associated with the transformation. Thus, exactly as in the preceeding subsection, if we perform a gauge transformation with

$$T = -\frac{\delta\chi}{\dot{\chi}}, \tag{3.199}$$

the result is

$$\delta\tilde{\chi} = 0, \tag{3.200}$$

that is we appear to have gauged away the χ perturbations.

This procedure should be geometrically clear upon examining Fig. 3.1. There I plot the homogeneous background $\chi(t)$ as well as the perturbation $\delta\chi(x^i, t)$ for some spatial coordinate value x^i. As I explained in Section 3.4.1, any perturbation is only ever defined up to a gauge transformation of the corresponding background quantity. In this case, this means that we are free to shift the two curves relative to each other along the time direction, in a way that varies smoothly with time and space. In Fig. 3.1(a), for the case of a small amplitude perturbation $\delta\chi$, it is clear that we can shift the two curves into coincidence and hence eliminate the perturbation.

As the amplitude of the perturbation oscillation increases, we may need to violate the conditions that the Lie derivative be a good approximation to the gauge transformation, namely that the size of ξ^μ not exceed the variation scale of ξ^μ or the curvature scale of the background (recall the end of Section 3.4.1). This only means that a more accurate representation of the gauge transformation is needed than the Lie derivative. However, if the perturbation amplitude becomes great enough that the exact quantity is no longer monotonically decreasing, then no gauge transformation whatsoever can gauge away the perturbation. This should be clear from Fig. 3.1(b). Essentially the transformation would need to be discontinuous and not one-to-one (so that it is no longer a diffeomorphism). This is essentially the reason that at late enough times density perturbations cannot be gauged away.

A related scenario is illustrated in Fig. 3.1(c). Here the background quantity is no longer monotonic, and near the times when $\dot{\chi} = 0$ the gauge transformation is not defined, even if the oscillation amplitude of the perturbations is small. Notice that the expression (3.199) for T diverges at these times.

Conversely, if near the times when $\dot{\chi} = 0$ we restrict gauge transformations to "small" shifts T, then the gauge transformation (3.198) will have negligible effect on the perturbation $\delta\chi$. At sufficiently late times in the evolution of the universe all background time derivatives become negligible. This is the reason that (small) gauge transformations have no significant effect on late time perturbations.

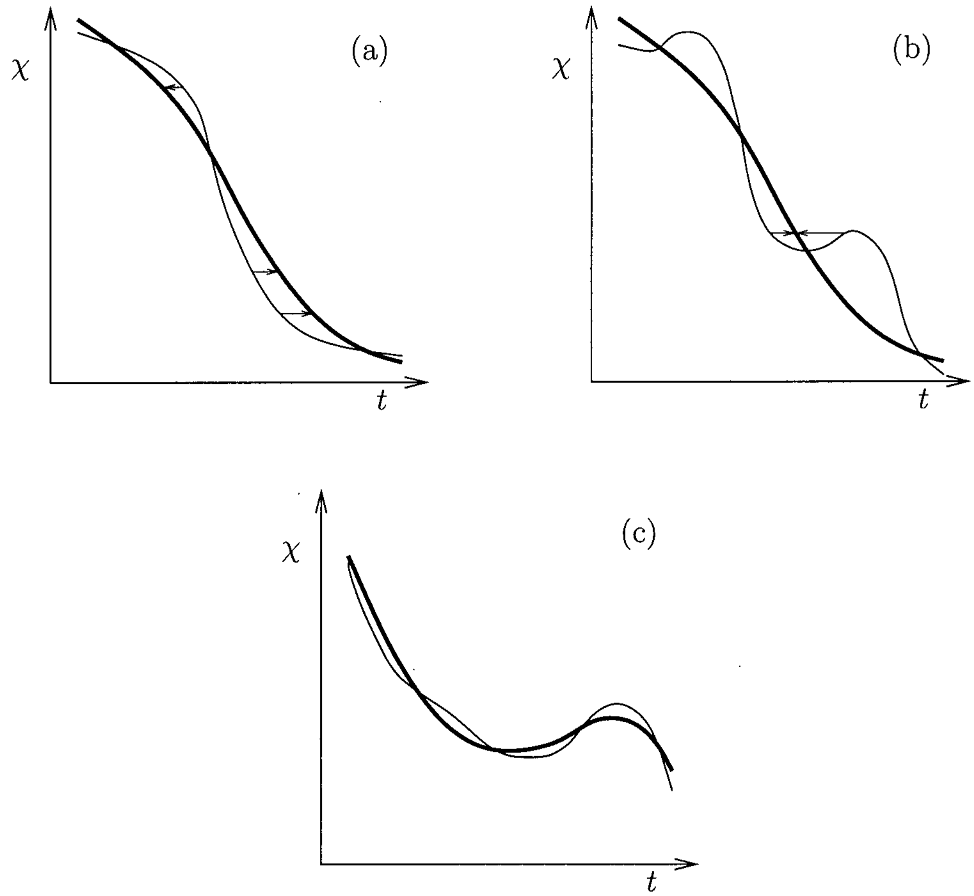

Figure 3.1: A scalar homogeneous background quantity $\chi(t)$ (heavy lines) and its perturbation $\delta\chi(x^i, t)$ (fine lines) along some worldline. In (a) the perturbation can be gauged away with a time-dependent shift in time (small arrows). In (b), this cannot be done smoothly. In (c) it cannot be done because $\dot{\chi}$ changes sign.

Of the several gauge choices discussed in the previous section, uniform density, uniform pressure, and comoving gauges all share the potential for singular behaviour. For uniform density gauge, the gauge transformation T becomes singular if the background density becomes stationary, $\dot{\rho} = 0$. This can in fact happen during the phase of reheating when the inflaton field oscillates about the minimum of its potential. From the conservation equation (2.65) and Eqs. (2.128) and (2.130), a universe dominated by homogeneous scalar fields φ_A satisfies

$$\dot{\rho} = -3H\dot{\varphi} \cdot \dot{\varphi}. \tag{3.201}$$

Thus when the scalar field velocities vanish, uniform density gauge becomes singular. This gauge also becomes singular in the slow-roll limit, Eq. (2.158). Similarly, for $\dot{P} = 0$, uniform pressure gauge is singular, and, since $\rho + P = \dot{\varphi} \cdot \dot{\varphi}$ for homogeneous scalar fields, comoving gauge can become singular as well. Thus these gauge choices should be used with caution in inflationary studies.

3.5 Linear perturbation dynamics

3.5.1 General form of perturbed Einstein tensor

The previous sections in this chapter provide us with two important tools which greatly simplify the treatment of the perturbed metric and energy-momentum tensor. The vector and tensor decompositions allow us to write separate equations of motion for scalar, vector, and tensor modes, and to concentrate on the scalar modes as sources of large-scale structure. An appropriate choice of gauge can furthermore eliminate the ambiguity in the metric and allows a considerable simplification of the form of the metric and hence of the equations of motion. In this section, I will use these two results and introduce a technique that will further simplify the dynamical equations. Namely, I will derive the equations of motion under the approximation of linear perturbation theory, which the cosmological principle and actual observations imply will be accurate.

Recall from earlier in this chapter that the general approach in cosmological perturbation theory is to decompose any arbitrarily perturbed exact quantity into a homogeneous and isotropic background and a perturbation. The exact metric $g_{\mu\nu}$ and energy-momentum tensor $T_{\mu\nu}$ are supposed to represent the real universe, and they are the quantities we are ultimately interested in determining. In doing so we concoct the fictitious homogeneous background quantities ${}^0g_{\mu\nu}$ and ${}^0T_{\mu\nu}$, which "exist" on a fictitious homogeneous and isotropic spacetime manifold. On this background manifold these quantities are decreed to precisely satisfy Einstein's equation, *i.e.*

$$ {}^0G_{\mu\nu} = 8\pi G\, {}^0T_{\mu\nu}. \tag{3.202} $$

To define the perturbations we must specify a correspondence, or mapping, between events on the background and exact manifolds. As I discussed in Section 3.4.1, the ambiguity in this mapping amounts precisely to the ability to perform arbitrary gauge transformations of the exact quantities. Thus, for example, the exact metric tensor can be decomposed as the sum

$$ \tilde{g}_{\mu\nu}(t, x^i) = {}^0g_{\mu\nu}(t) + \delta g_{\mu\nu}(t, x^i). \tag{3.203} $$

Here I have indicated that this decomposition is only defined up to a gauge transformation: $\tilde{g}_{\mu\nu}$ is the result of performing an arbitrary gauge transformation $x^\mu \rightarrow \tilde{x}^\mu$ on the exact metric tensor $g_{\mu\nu}$.

In defining the background and perturbed quantities, I must choose certain conventions. I will raise and lower indices on background quantities with the unperturbed metric ${}^0g_{\mu\nu}$. Thus in particular we have

$$ {}^0g^{\mu\lambda}\,{}^0g_{\lambda\nu} = \delta^\mu{}_\nu \tag{3.204} $$

as an exact statement. Indices on perturbed quantities can be raised or lowered with the background or the exact metric—the results are the same to first order. Also, I will define

$$ \delta g^\mu{}_\nu \equiv g^{\mu\lambda}\delta g_{\lambda\nu} \tag{3.205} $$

so that by Eq. (2.13), or equivalently via a simple calculation using Eq. (3.204),

$$ \delta(g^{\mu\nu}) = -\delta g^{\mu\nu} + \mathcal{O}(\delta g^2). \tag{3.206} $$

However, I define

$$ \delta G^\mu{}_\nu \equiv \delta(g^{\mu\lambda}G_{\lambda\nu}) \tag{3.207} $$

$$ = -\delta g^{\mu\lambda}G_{\lambda\nu} + g^{\mu\lambda}\delta G_{\lambda\nu} \tag{3.208} $$

to first order, so that in general

$$ \delta G^\mu{}_\nu \neq g^{\mu\lambda}\delta G_{\lambda\nu}. \tag{3.209} $$

Now I will proceed with the derivation of the perturbed Einstein equation. Using the definition of the connection coefficients, Eq. (2.5), the perturbations $\delta\Gamma^\mu_{\nu\lambda}$ are

$$ \delta\Gamma^\mu_{\nu\lambda} = g^{\mu\rho}\left(-\delta g_{\rho\sigma}\Gamma^\sigma_{\nu\lambda} + \delta g_{\rho(\nu,\lambda)} - \frac{1}{2}\delta g_{\nu\lambda,\rho}\right). \tag{3.210} $$

In this expression I have dropped the $\mathcal{O}(\delta g^2)$ terms, and I will henceforth only write the first order terms in all perturbation equations. This is the linear approximation, where all products of perturbations are supposed to be negligible. Note that $\delta\Gamma^\mu_{\nu\lambda}$ is in fact a third-rank tensor: writing

$$ \delta g_{\rho\nu,\lambda} = \delta g_{\rho\nu;\lambda} + \Gamma^\sigma_{\rho\lambda}\delta g_{\nu\sigma} + \Gamma^\sigma_{\nu\lambda}\delta g_{\rho\sigma} \tag{3.211} $$

and similarly for the permuted terms in Eq. (3.210), the connection coefficients cancel and we are left with

$$ \delta\Gamma^\mu_{\nu\lambda} = \delta g^\mu{}_{(\nu;\lambda)} - \frac{1}{2}\delta g_{\nu\lambda}{}^{;\mu}. \tag{3.212} $$

Next, linearly perturbing the definition of the Riemann tensor, Eq. (2.4), gives

$$\delta R_{\mu\nu} = 2\delta\Gamma^{\lambda}_{\mu[\nu,\lambda]} + \Gamma\delta\Gamma \text{ terms.} \tag{3.213}$$

Here the $\Gamma\delta\Gamma$ terms are each proportional to the zeroth order connection coefficients. Therefore, if we choose locally geodesic coordinates each of these terms vanishes, and we can immediately write the perturbed Ricci tensor as the covariant expression

$$\delta R_{\mu\nu} = 2\delta\Gamma^{\lambda}_{\mu[\nu;\lambda]} \tag{3.214}$$

$$= \frac{1}{2}\left(\delta g^{\lambda}{}_{\mu;\nu\lambda} + \delta g^{\lambda}{}_{\nu;\mu\lambda} - \delta g_{\mu\nu;\lambda}{}^{\lambda} - \delta g_{;\mu\nu}\right), \tag{3.215}$$

where

$$\delta g \equiv \delta g^{\mu}{}_{\mu} \tag{3.216}$$

is the trace of the metric perturbation. Finally, writing

$$G^{\mu}{}_{\nu} = g^{\mu\lambda}R_{\lambda\nu} - \frac{1}{2}\delta^{\mu}{}_{\nu}R, \tag{3.217}$$

I find the linear perturbation of the Einstein tensor to be

$$\delta G^{\mu}{}_{\nu} = -\delta g^{\mu\lambda}R_{\lambda\nu} + g^{\mu\lambda}\delta R_{\lambda\nu} - \frac{1}{2}\delta^{\mu}{}_{\nu}\delta R \tag{3.218}$$

$$= \frac{1}{2}\left(-\delta g^{;\mu}{}_{\nu} - \delta g^{\mu}{}_{\nu;\lambda}{}^{\lambda} + \delta g^{\lambda}{}_{\nu}{}^{;\mu}{}_{\lambda} + \delta g^{\mu}{}_{\lambda;\nu}{}^{\lambda} - \delta^{\mu}{}_{\nu}\delta R\right) - \delta g^{\mu}{}_{\lambda}R^{\lambda}{}_{\nu}, \tag{3.219}$$

where

$$\delta R \equiv \delta\left(g^{\lambda\rho}R_{\lambda\rho}\right) = -\delta g^{\lambda\rho}R_{\lambda\rho} + g^{\lambda\rho}\delta R_{\lambda\rho} \tag{3.220}$$

$$= -\delta g^{\lambda}{}_{\rho}R^{\rho}{}_{\lambda} - \delta g^{;\lambda}{}_{\lambda} + \delta g^{\lambda}{}_{\rho;\lambda}{}^{\rho} \tag{3.221}$$

is the perturbed Ricci scalar. The first order linearized Einstein equation is simply

$$\delta G^{\mu}{}_{\nu} = 8\pi G\delta T^{\mu}{}_{\nu}, \tag{3.222}$$

where $\delta G^{\mu}{}_{\nu}$ is given in terms of the perturbed metric by expression (3.219), and $\delta T^{\mu}{}_{\nu} \equiv \delta(T^{\mu}{}_{\nu})$ is given in terms of the matter variables by Eqs. (3.68), (3.71), and (3.72). Recalling Eq. (3.203), this linearized Einstein equation specifies the perturbations only up to a gauge transformation of the backgrounds.

3.5.2 Explicit form of perturbed Einstein tensor

Scalar perturbations

I will now evaluate the scalar-derived part of the perturbed Einstein tensor. According to Eq. (3.13), only the scalar parts of the perturbed metric will enter into this part of the linearized Einstein tensor. I will use the spatially flat, comoving, conformal time background chart specified by Eq. (2.54). I will perform the calculations in longitudinal gauge, where

$$\delta g_{00} = -2a^2\Phi, \qquad \delta g_{0i} = 0, \qquad \delta g_{ij} = -2a^2\Psi\gamma_{ij}, \tag{3.223}$$

$$\delta g^0{}_0 = 2\Phi, \qquad \delta g^0{}_i = 0, \qquad \delta g^i{}_j = -2\Psi\delta^i{}_j, \tag{3.224}$$

and afterwards generalize to arbitrary gauge.

First I determine the perturbed Ricci scalar. I obtain after straightforward calculations

$$\delta g^\lambda{}_\rho R^\rho{}_\lambda = \frac{6}{a^2}\left[\mathcal{H}'\Phi - (\mathcal{H}' + 2\mathcal{H}^2)\Psi\right], \tag{3.225}$$

$$\delta g^{;\lambda}{}_\lambda = \frac{1}{a^2}\left(-\delta g'' - 2\mathcal{H}\delta g' + \nabla^2\delta g\right), \tag{3.226}$$

$$\delta g^\lambda{}_{\rho;\lambda}{}^\rho = -\frac{2}{a^2}\left[\Phi'' + 5\mathcal{H}\Phi' + 3\mathcal{H}\Psi' + (3\mathcal{H}' + 6\mathcal{H}^2)(\Phi + \Psi) + \nabla^2\Psi\right], \tag{3.227}$$

where

$$\delta g = 2\Phi - 6\Psi. \tag{3.228}$$

Thus by Eq. (3.221)

$$\delta R = -\frac{2}{a^2}\left[3\Psi'' + 3\mathcal{H}(3\Psi' + \Phi') + 6(\mathcal{H}' + \mathcal{H}^2)\Phi + \nabla^2(\Phi - 2\Psi)\right]. \tag{3.229}$$

Now I must calculate the terms needed for the component $\delta G^0{}_0$. Straightforward but somewhat lengthy calculations give

$$\delta g^{;0}{}_0 = \frac{1}{a^2}\left(-\delta g'' + \mathcal{H}\delta g'\right), \tag{3.230}$$

$$\delta g^0{}_{0;\lambda}{}^\lambda = \frac{2}{a^2}\left[-\Phi'' - 2\mathcal{H}\Phi' + 6\mathcal{H}^2(\Phi + \Psi) + \nabla^2\Phi\right], \tag{3.231}$$

$$\delta g^\lambda{}_0{}^{;0}{}_\lambda = \delta g^0{}_{\lambda;0}{}^\lambda = \frac{2}{a^2}\left[-\Phi'' - 2\mathcal{H}\Phi' - 3\mathcal{H}\Psi' + 3\mathcal{H}^2(\Phi + \Psi)\right], \tag{3.232}$$

$$\delta g^0{}_\lambda R^\lambda{}_0 = \frac{6}{a^2}\mathcal{H}'\Phi. \tag{3.233}$$

Summing the terms in Eq. (3.219) thus gives

$$\delta G^0{}_0 = \frac{2}{a^2}\left[3\mathcal{H}(\Psi' + \mathcal{H}\Phi) - \nabla^2\Psi\right] \tag{3.234}$$

as the time-time component of the linearized Einstein tensor in longitudinal gauge.

Completely analogous calculations give

$$\delta G^{0(\mathrm{s})}_{\ \ i} = -\frac{2}{a^2}\left(\Psi' + \mathcal{H}\Phi\right)_{,i} \tag{3.235}$$

and

$$\begin{aligned}\delta G^{i(\mathrm{s})}_{\ \ j} = \frac{1}{a^2}\Bigg[&2\delta^i_{\ j}\left(\Psi'' + \mathcal{H}(2\Psi' + \Phi') + (\mathcal{H}^2 + 2\mathcal{H}')\Phi + \frac{1}{2}\nabla^2(\Phi - \Psi)\right)\\ &+ (\Psi - \Phi)^{,i}_{\ \ ,j}\Bigg]\end{aligned} \tag{3.236}$$

as the remaining scalar-derived components. (As I explained in Section 3.4.3, the component $\delta G^i_{\ 0}$ will provide no independent information.) Notice that the time-time and time-space components contain no second time derivatives—these expressions lead to constraints on the scalar functions and their first time derivatives. As discussed in Section 2.2.4, only the space-space component of Einstein's equation contains second time derivatives and is "dynamical".

Vector perturbations

Next I will evaluate the vector-derived part of the perturbed Einstein tensor. Recall from Section 3.4.3 that this vector part is gauge invariant in a homogeneous background. Also recall from Section 3.4.4 that I can, nevertheless, always perform a spatial gauge transformation on the metric functions to vector gauge, where $F_i = 0$, and thus simplify the metric calculations. In the end I can simply replace all instances of S_i in the Einstein tensor with $-\sigma_i/a$, where σ_i is the shear vector, to obtain the gauge-invariant equation of motion.

Again, only the vector-derived parts of the perturbed metric will enter the calculation of the Einstein tensor, and in vector gauge they become

$$\delta g_{00} = 0, \qquad \delta g_{0i} = a^2 S_i, \qquad \delta g_{ij} = 0, \tag{3.237}$$

$$\delta g^0_{\ 0} = 0, \qquad \delta g^0_{\ i} = -S_i, \qquad \delta g^i_{\ j} = 0. \tag{3.238}$$

Since we cannot linearly construct a scalar from a transverse vector, we can immediately conclude that

$$\delta g = \delta R = 0. \tag{3.239}$$

Similarly we must have

$$\delta G^{0(\mathrm{v})}_{\ \ 0} = 0 \tag{3.240}$$

so that the time-time Einstein equation is trivial.

I calculate the terms needed for the time-space component to be

$$\delta g^{0}{}_{i;\lambda}{}^{\lambda} = \frac{1}{a^2}\left(S_i'' + 2\mathcal{H}S_i' - 6\mathcal{H}^2 S_i - \nabla^2 S_i\right), \tag{3.241}$$

$$\delta g^{\lambda}{}_{i}{}^{;0}{}_{\lambda} = \frac{1}{a^2}\left(S_i'' + 3\mathcal{H}S_i' - 4\mathcal{H}^2 S_i\right), \tag{3.242}$$

$$\delta g^{0}{}_{\lambda;i}{}^{\lambda} = -\frac{1}{a^2}\left[\mathcal{H}S_i' + (2\mathcal{H}' + 6\mathcal{H}^2)S_i\right], \tag{3.243}$$

$$\delta g^{0}{}_{\lambda}R^{\lambda}{}_{i} = -\frac{1}{a^2}\left(\mathcal{H}' + 2\mathcal{H}^2\right)S_i. \tag{3.244}$$

Combining these terms we obtain the concise result

$$\delta G^{0(\mathrm{v})}_{\ i} = \frac{1}{2a^2}\nabla^2 S_i. \tag{3.245}$$

Similar calculations give

$$\delta G^{i(\mathrm{v})}_{\ j} = -\frac{1}{a^2}\gamma^{il}\left(S'_{(l,j)} + 2\mathcal{H}S_{(l,j)}\right) \tag{3.246}$$

for the space-space component of the vector-perturbed Einstein tensor.

Finally, we can express these components in explicitly gauge-invariant form by substituting S_i with $-\sigma_i/a$. The result is

$$\delta G^{0(\mathrm{v})}_{\ i} = -\frac{1}{2a^3}\nabla^2\sigma_i, \tag{3.247}$$

$$\delta G^{i(\mathrm{v})}_{\ j} = \frac{1}{a^2}\gamma^{il}\left(\dot\sigma_{(l,j)} + H\sigma_{(l,j)}\right). \tag{3.248}$$

Tensor perturbations

The derivation is simplest and most elegant for tensor modes (gravitational waves). Here we immediately conclude that

$$\delta g^{\mu}{}_{\nu;\mu} = \delta g = \delta R = \delta G^{0(\mathrm{t})}_{\ 0} = \delta G^{0(\mathrm{t})}_{\ i} = 0, \tag{3.249}$$

since no scalar or vector can be linearly constructed from h_{ij}. Also, we have no gauge freedom, since $h^{i}{}_{j}$ does not change under infinitesimal gauge transformations, as we found in Section 3.4.2.

The first condition in (3.249) is known as the covariant transverseness condition. This condition allows us to derive a compact covariant expression for the perturbed dynamical equation. We can do this by commuting the derivatives in the third and fourth terms in the perturbed Einstein tensor, Eq. (3.219), and then exploiting the transverseness property. Explicitly, the properties of the Riemann tensor imply that

$$2\delta g^{\lambda}{}_{\mu;[\nu\lambda]} = \delta g^{\lambda}{}_{\rho} R^{\rho}{}_{\mu\nu\lambda} - \delta g^{\rho}{}_{\mu} R^{\lambda}{}_{\rho\nu\lambda} \tag{3.250}$$

$$= \delta g^{\lambda}{}_{\rho} R^{\rho}{}_{\mu\nu\lambda} + \delta g^{\rho}{}_{\mu} R_{\rho\nu}. \tag{3.251}$$

However,

$$2\delta g^{\lambda}{}_{\mu;[\nu\lambda]} = \delta g^{\lambda}{}_{\mu;\nu\lambda} - \delta g^{\lambda}{}_{\mu;\lambda\nu} \tag{3.252}$$

$$= \delta g^{\lambda}{}_{\mu;\nu\lambda} \tag{3.253}$$

by transverseness, so we simply have

$$\delta g^{\lambda}{}_{\mu;\nu\lambda} = \delta g^{\lambda}{}_{\rho} R^{\rho}{}_{\mu\nu\lambda} + \delta g^{\rho}{}_{\mu} R_{\rho\nu}. \tag{3.254}$$

Exchanging μ and ν gives

$$\delta g^{\lambda}{}_{\nu;\mu\lambda} = \delta g^{\lambda}{}_{\rho} R^{\rho}{}_{\nu\mu\lambda} + \delta g^{\rho}{}_{\nu} R_{\rho\mu}. \tag{3.255}$$

When we sum these two expressions we will need

$$\delta g^{\lambda\rho} R_{\rho\mu\nu\lambda} + \delta g^{\lambda\rho} R_{\rho\nu\mu\lambda} = \delta g^{\lambda\rho}(R_{\rho\mu\nu\lambda} + R_{\lambda\nu\mu\rho}) \tag{3.256}$$

$$= \delta g^{\lambda\rho}(R_{\rho\mu\nu\lambda} + R_{\mu\rho\lambda\nu}) \tag{3.257}$$

$$= 2\delta g^{\lambda\rho} R_{\rho\mu\nu\lambda}, \tag{3.258}$$

by the symmetries of the Riemann tensor and $\delta g_{\mu\nu}$. We now have all the non-zero terms in the perturbed Einstein tensor, Eq. (3.219). Performing the sum, I find

$$\delta G^{\mu(\mathrm{t})}_{\nu} = \frac{1}{2}\left(-\Box\delta g^{\mu}{}_{\nu} + 2\delta g^{\lambda}{}_{\rho} R^{\rho\mu}{}_{\nu\lambda} + \delta g^{\rho}{}_{\nu} R^{\mu}{}_{\rho} - \delta g^{\mu}{}_{\rho} R^{\rho}{}_{\nu}\right). \tag{3.259}$$

Finally, note that for an isotropic background, $R^{i}{}_{j} \propto \delta^{i}{}_{j}$, and so the last two terms in Eq. (3.259) cancel. We are left with

$$\delta G^{\mu(\mathrm{t})}_{\nu} = -\frac{1}{2}\Box\delta g^{\mu}{}_{\nu} + \delta g^{\lambda}{}_{\rho} R^{\rho\mu}{}_{\nu\lambda} \tag{3.260}$$

as the covariant tensor mode linearized Einstein tensor.

We can also express this tensor in the standard chart defined by Eq. (3.2). In this case we have

$$\delta g_{00} = \delta g_{0i} = 0, \qquad \delta g_{ij} = a^2 h_{ij}, \tag{3.261}$$

$$\delta g^{0}{}_{0} = \delta g^{0}{}_{i} = 0, \qquad \delta g^{i}{}_{j} = h^{i}{}_{j}. \tag{3.262}$$

To calculate the term

$$\delta g^{\lambda}{}_{\rho} R^{\rho i}{}_{j\lambda} = g^{in} \delta g^{l}{}_{m} R^{m}{}_{njl} \tag{3.263}$$

we can use the definition of the Riemann tensor, Eq. (2.4), together with the FRW connection coefficients, Eqs. (2.73), to obtain

$$R^{m}{}_{njl} = \mathcal{H}^2 (\delta^{m}{}_{j} \gamma_{nl} - \delta^{m}{}_{l} \gamma_{nj}). \tag{3.264}$$

Thus we have

$$\delta g^{\lambda}{}_{\rho} R^{\rho i}{}_{j\lambda} = \frac{1}{a^2} \mathcal{H}^2 \gamma^{in} h^{l}{}_{m} (\delta^{m}{}_{j} \gamma_{nl} - \delta^{m}{}_{l} \gamma_{nj}) \tag{3.265}$$

$$= \frac{1}{a^2} \mathcal{H}^2 h^{i}{}_{j}. \tag{3.266}$$

Next, straightforward calculations give

$$\Box \delta g^{i}{}_{j} = \frac{1}{a^2} \left(-h^{i\prime\prime}_{j} - 2\mathcal{H} h^{i\prime}_{j} + 2\mathcal{H}^2 h^{i}{}_{j} + \nabla^2 h^{i}{}_{j} \right), \tag{3.267}$$

and so we finally obtain

$$\delta G^{i(\mathrm{t})}_{j} = \frac{1}{2a^2} \left(h^{i\prime\prime}_{j} + 2\mathcal{H} h^{i\prime}_{j} - \nabla^2 h^{i}{}_{j} \right) \tag{3.268}$$

for the linearized Einstein tensor.

3.5.3 Perturbed Einstein's equations

In this subsection, I will combine the results of the previous subsection for the perturbed Einstein tensor with the results from Section 3.3.1 for the perturbed energy-momentum tensor and write explitly the various components of the scalar-, vector-, and tensor-derived parts of the linearly perturbed Einstein equation.

Scalar part

I calculated the scalar-derived part of the perturbed Einstein tensor in Section 3.5.2 using longitudinal gauge. In this section I will describe a very simple way to generalize those results and write Einstein's equation in arbitrary gauges. The idea is that we *know* how the rhs of Einstein's equation changes under arbitrary gauge transformations—I calculated the transformation of the energy-momentum tensor in Section 3.4.3. Thus we know how the lhs must transform. Now the lhs calculated in Section 3.5.2 is missing terms *linear* in the shear σ, which was set to zero in the longitudinal gauge derivation. Therefore we simply need to add an appropriate term to the lhs, linear in σ, to produce the correct gauge transformation behaviour. (Recall from Section 3.4.4 that because Einstein's equation is spatially gauge invariant, only the functions ϕ, ψ, and σ can appear in the scalar equation of motion.)

To begin I will write the scalar-derived components of Einstein's equation in longitudinal gauge, using Eqs. (3.234) to (3.236) for the perturbed Einstein tensor, and Eqs. (3.68), (3.71), and (3.72) for the perturbed energy-momentum tensor. I will rewrite the expressions using proper time rather than conformal time, to simplify the forms of the gauge transformations. The results are

$$3H(\dot{\Psi} + H\Phi) - \frac{1}{a^2}\nabla^2\Psi = -4\pi G\delta\rho, \qquad \text{0-0 component,} \tag{3.269}$$

$$(\dot{\Psi} + H\Phi)_{,i} = -4\pi G q_{,i}, \qquad \text{0-}i\text{ component.} \tag{3.270}$$

The perturbed matter variables must also be specified in longitudinal gauge here. The space-space components can be decomposed into an off-diagonal part (essentially the tracefree part), and a trace part, $\delta G^i{}_i = 8\pi G\delta T^i{}_i$:

$$(\Psi - \Phi)^{,i}{}_j = 8\pi G\Pi^{,i}{}_j, \qquad i \neq j, \tag{3.271}$$

$$\ddot{\Psi} + H(3\dot{\Psi} + \dot{\Phi}) + (3H^2 + 2\dot{H})\Phi + \frac{1}{3a^2}\nabla^2(\Phi - \Psi) = 4\pi G\delta P, \qquad \text{trace.} \tag{3.272}$$

Now to generalize these zero-shear gauge expressions to arbitrary gauge, recall that each of these expressions must contain only terms linear in ϕ, ψ, and σ, and their derivatives. Thus the arbitrary-gauge time-time component must have the form

$$3H(\dot{\psi} + H\phi) - \frac{1}{a^2}\nabla^2\psi + f(\sigma) = -4\pi G\delta\rho, \tag{3.273}$$

where $f(\sigma)$ denotes terms linear in σ and its derivatives. The rhs of (3.273) changes under a gauge transformation $t \to \tilde{t} = t - T$ according to Eq. (3.151),

$$\widetilde{\text{rhs}} = \text{rhs} - 4\pi G\dot{\rho}T \tag{3.274}$$

$$= \text{rhs} - 3H\dot{H}T, \tag{3.275}$$

where for the second line I have used the background energy conservation equation (2.65) and the background equation of motion (2.87). Similarly, according to Eqs. (3.147) to (3.149), the lhs of (3.273) changes under a temporal gauge transformation according to

$$\widetilde{\text{lhs}} = \text{lhs} - 3H\dot{H}T + \frac{1}{a^2}\nabla^2 TH + f(\sigma + T) - f(\sigma), \tag{3.276}$$

where I have used the linearity of f. Therefore, equating Eqs. (3.275) and (3.276) I conclude that $f(\sigma)$ *must* transform according to

$$f(\sigma + T) = f(\sigma) - \frac{1}{a^2}\nabla^2 TH \tag{3.277}$$

so that

$$f(\sigma) = -\frac{1}{a^2}\nabla^2 \sigma H. \tag{3.278}$$

Thus, substituting this expression into (3.273), I can now write the energy constraint equation in a completely arbitrary gauge as

$$3H(\dot{\psi} + H\phi) - \frac{1}{a^2}\nabla^2(\psi + H\sigma) = -4\pi G\delta\rho. \tag{3.279}$$

Repeating this procedure for the longitudinal gauge momentum constraint equation, (3.270), I find that both sides already transform identically under a temporal gauge transformation. Thus the shear σ does not appear in this component, and I can immediately write

$$(\dot{\psi} + H\phi)_{,i} = -4\pi G q_{,i} \tag{3.280}$$

as the arbitrary gauge momentum constraint. Note that this equation implies that

$$\dot{\psi} + H\phi = -4\pi G q + C(t), \tag{3.281}$$

for arbitrary spatial constant $C(t)$. However, since the momentum density is determined by $q_{,i}$, the scalar q is only defined up to a spatial constant. Thus we can absorb the constant $C(t)$ into q without altering the physical content and write

$$\dot{\psi} + H\phi = -4\pi G q. \tag{3.282}$$

For the dynamical (space-space) equations, (3.271) and (3.272), I find that shear terms do need to be added. The result for the off-diagonal part is

$$\psi - \phi + \dot{\sigma} + H\sigma = 8\pi G\Pi. \tag{3.283}$$

Here I have used a result paralleling that of Eq. (3.296) in the following section and have absorbed a spatial constant arising from the mixed spatial derivatives into the function Π, just as I did for the momentum constraint equation. The result for the trace part is

$$\ddot{\psi} + H(3\dot{\psi} + \dot{\phi}) + (3H^2 + 2\dot{H})\phi = 4\pi G\left(\delta P + \frac{2}{3a^2}\nabla^2\Pi\right), \tag{3.284}$$

where I have substituted the off-diagonal equation. Each of these equations is valid in any gauge. Note, however, that the off-diagonal equation is actually gauge invariant, since the anisotropic stress Π is gauge invariant.

Taking the time derivative of the energy constraint Eq. (3.279) and using the remaining components of Einstein's equation it is straightforward to verify the perturbed energy conservation equation, (3.84). Similarly, substituting the momentum constraint Eq. (3.282) into the trace equation (3.284) it is very easy to verify the momentum conservation equation, (3.91). The contracted Bianchi identity does indeed hold.

To close this treatment of the scalar dynamics, I will rewrite the two Einstein equations (3.279) and (3.284) in terms of the expansion perturbation $\delta\theta$, Eq. (3.53), and the Ricci curvature scalar ${}^{(3)}R$ of constant-time hypersurfaces, Eq. (3.21). While the equations can be rewritten in many ways, these forms are particularly illuminating. For the energy constraint equation, I find

$$2H\delta H = \frac{8\pi G}{3}\delta\rho - \frac{{}^{(3)}R}{6}, \tag{3.285}$$

where I define $\delta H \equiv \delta\theta/3$. Noting that the Ricci scalar for a homogeneous background is ${}^{(3)}R = 6\mathcal{K}/a^2$ [recall Eq. (2.117)], this equation is *precisely* the result of naively "perturbing" the background Friedmann equation (2.83)! Of course, there is no reason (that I can think of, at least) to expect that perturbing the homogeneous Friedmann equation should yield the correct equation for inhomogeneous perturbations valid in an arbitrary gauge.

Similarly rewriting the trace part of the space-space equation, I find

$$-2\delta\dot{H} + 2\dot{H}\phi - 6H\delta H + \frac{2}{3a^2}\nabla^2\phi = 8\pi G\delta P + \frac{{}^{(3)}R}{6}, \tag{3.286}$$

which looks, again, like a perturbed version of the homogeneous space-space Einstein equation (2.84), taking into account that the coordinate time is perturbed relative to the proper time via the lapse function.

Note that in writing the off-diagonal equation (3.283) I have made an assumption. In particular, I have restricted the universe to be spatially flat *on average.* To see this, notice that according to the *actual* off-diagonal equation,

$$(\psi - \phi + \dot{\sigma} + H\sigma)^{,i}{}_{j} = 8\pi G\Pi^{,i}{}_{j}, \tag{3.287}$$

I can always shift the curvature perturbation according to

$$\psi \to \psi + Cr^2, \tag{3.288}$$

for constant C and comoving radius $r^2 = x_i x^i$. Such a transformation is not allowed in Eq. (3.283), however. (In my derivation of that expression I referred to a result based on a Fourier expansion, which assumes ψ is bounded.) But Eqs. (3.285) and (3.286) show that such a shift in ψ is equivalent to a (homogeneous) perturbation towards a spatially open or closed universe [the momentum constraint is invariant under (3.288)]. Therefore, as claimed, in writing the off-diagonal equation in the form (3.283), I ignore such a possibility and assume spatial flatness on average. But this is a completely reasonable assumption: recall from Section 2.2.7 that any departure from $\Omega_{\mathcal{K}} = 0$ will grow by a tremendous factor between the time near the end of inflation (which I am interested in) and today. (As I explained there it is not the spatial curvature which grows, but its contribution to the Friedmann equation, $\Omega_{\mathcal{K}}$.) Since we know that the universe is currently very close to flat, it is completely valid to ignore such a homogeneous curvature perturbation at early times.

Vector part

The vector-derived Einstein equations are much easier to write down. Combining the expressions (3.247) and (3.248) for the perturbed Einstein tensor with Eqs. (3.71) and (3.72) for the energy-momentum tensor, I find

$$\frac{1}{a^2}\nabla^2\sigma_i = -16\pi G v_i \qquad \text{0-}i\text{ component,} \tag{3.289}$$

$$\dot{\sigma}_{(i,j)} + H\sigma_{(i,j)} = 8\pi G\pi_{(i,j)} \qquad i\text{-}j\text{ component.} \tag{3.290}$$

The function v_i determines the momentum density and π_i determines the anisotropic stress. As I have discussed above, these equations are gauge invariant and describe the evolution of completely general linear vector perturbations. Combining these two equations it is very easy to verify the vector part of the momentum conservation law, Eq. (3.92).

Tensor part

Using the covariant expression (3.260), I can write the tensor part of the perturbed Einstein equation (in an isotropic background) in the covariant form

$$-\frac{1}{2}\Box\delta g^{\mu}{}_{\nu} + \delta g^{\lambda}{}_{\rho}R^{\rho\mu}{}_{\nu\lambda} = 8\pi G\delta T^{\mu(\mathrm{t})}_{\nu}. \tag{3.291}$$

Using instead the component form (3.268) of the Einstein tensor and Eq. (3.72) for the perturbed energy-momentum tensor we have

$$\frac{1}{2a^2}\left(h^{i\prime\prime}_{j} + 2\mathcal{H}h^{i\prime}_{j} - \nabla^2 h^{i}{}_{j}\right) = 8\pi G\tau^{i}{}_{j}. \tag{3.292}$$

The source function $\tau^{i}{}_{j}$ is the tensor part of the anisotropic stress. This equation is gauge invariant, and describes the evolution of linear gravitational waves. Notice also that, without the source term, this equation looks exactly like the equation of motion for a pair of massless, free scalar fields, if metric perturbations are ignored [see Eq. (3.346)].

3.5.4 Behaviour of perturbations in Minkowsky vacuum

Having now derived the first-order dynamical and constraint equations, it will be instructive to consider what they tell us about linear metric perturbations in the special case of a vacuum Minkowsky background spacetime. In this case, we have $\mathcal{H} = \delta T^{\mu}{}_{\nu} = 0$.

First, for the scalar modes, the off-diagonal part of the longitudinal gauge dynamical scalar equation (3.271) gives

$$(\Psi - \Phi)^{,i}{}_{j} = 0, \qquad i \neq j. \tag{3.293}$$

If we Fourier-decompose the perturbation functions into momentum-space,

$$\Psi(\eta, x^i) = \int \frac{d^3k}{(2\pi)^3} e^{ik_j x^j} \Psi_{\mathbf{k}}(\eta) \tag{3.294}$$

and similarly for Φ, then the off-diagonal equation becomes

$$k^i k_j (\Psi - \Phi)_{\mathbf{k}} = 0 \tag{3.295}$$

for each mode $\mathbf{k}$. This implies that, for each $\mathbf{k}$, $k^i = 0$ for two distinct i. However, the orientation of the spatial coordinates is arbitrary. Thus we must have $k^i = 0$ for all i, and hence $\Psi - \Phi$ is a spatial constant,

$$\Psi - \Phi = (\Psi - \Phi)(\eta). \tag{3.296}$$

The longitudinal gauge energy constraint equation (3.269) becomes in Minkowsky vacuum

$$\nabla^2 \Psi = 0. \tag{3.297}$$

The only bounded solution is the spatial constant

$$\Psi = \Psi(\eta). \tag{3.298}$$

This is clear in k-space, where $k^2 \Psi_{\mathbf{k}} = 0$ implies $k = 0$. Combining Eqs. (3.296) and (3.298) gives

$$\Phi = \Phi(\eta). \tag{3.299}$$

Such spatially constant perturbations can always be absorbed into the background or transformed away. The longitudinal gauge scalar-derived metric is

$$ds^2_{(s)} = \tilde{a}^2(\eta)\left\{-[1+2\Phi(\eta)]d\eta^2 + [1-2\Psi(\eta)]\gamma_{ij}dx^i dx^j\right\}. \tag{3.300}$$

Defining

$$a^2(\eta) = \tilde{a}^2(\eta)[1-2\Psi(\eta)] \tag{3.301}$$

the metric becomes

$$ds^2_{(s)} = a^2(\eta)\left\{-[1+2\Phi(\eta)+2\Psi(\eta)]d\eta^2 + \gamma_{ij}dx^i dx^j\right\}. \tag{3.302}$$

Now if I perform a gauge transformation with $\xi^{0\prime} = -\Phi(\eta) - \Psi(\eta)$ and $\xi^i = 0$, then according to Eqs. (3.129) to (3.132), Φ is transformed away while B gains a spatially constant part. Since only the gradient $B_{,i}$ appears in the metric, this shift in B is irrelevant. The metric now becomes

$$ds^2_{(s)} = a^2(\eta)\left(-d\eta^2 + \gamma_{ij}dx^i dx^j\right). \tag{3.303}$$

The background equations of motion of course imply that $a =$ const for the Minkowsky vacuum. At any rate, the important result here is that there are no physical scalar metric perturbations in the Minkowsky vacuum at linear order.

Next, for the vector modes, when the momentum density v_i vanishes, the momentum constraint Eq. (3.289) becomes

$$\nabla^2\sigma_i = 0. \tag{3.304}$$

Again, this implies

$$\sigma_i = \sigma_i(\eta). \tag{3.305}$$

Recalling from Section 3.2.2 that for vector perturbations the geometry of the spacetime is completely determined by the shear tensor

$$\sigma_{ij} = \sigma_{(i,j)} = 0, \tag{3.306}$$

we see that there are no physical vector perturbations when the momentum density vanishes. Equivalently, considering the gauge transformations (3.121) and (3.124), when $\sigma_i = \sigma_i(\eta)$, we can gauge away S_i, while F_i becomes an irrelevant spatial constant. This conclusion holds in particular, of course, for the Minkowsky vacuum. When $v_i \neq 0$, we can still constrain S_i tightly in the case that the anisotropic stress $\pi_{(i,j)}$ vanishes. Then the dynamical Einstein equation (3.290) becomes

$$\dot{\sigma}_{ij} + H\sigma_{ij} = 0, \tag{3.307}$$

which can be readily integrated to give

$$\sigma_{ij} \propto a^{-1}. \tag{3.308}$$

Thus in an expanding universe, the shear tensor σ_{ij} decays towards zero. To summarize, when the anisotropic stress vanishes but the momentum density does not, linear vector mode perturbations decay in an expanding universe, while when the momentum density vanishes there are no physical linear vector modes whatsoever.

Finally, for tensor modes, the single Einstein equation (3.292) becomes in Minkowsky vacuum

$$h^{i\,\prime\prime}_{\;j} - \nabla^2 h^i{}_j = 0. \tag{3.309}$$

The tensor perturbation $h^i{}_j$ cannot be gauged away, and Eq. (3.309) describes the evolution of (physical!) gravitational waves. The TT tensor perturbation $h^i{}_j$ contains two degrees of freedom, and these correspond to the two transverse polarization states of the wave.

The covariant transverse and traceless conditions

$$\delta g^\mu{}_{\nu;\mu} = \delta g = 0 \tag{3.310}$$

are often used to define *transverse traceless* gauge. It is now clear that the metric can always be brought into this gauge in Minkowsky vacuum, where the physical scalar and vector modes vanish leaving only the tensor modes which trivially satisfy Eq. (3.310).

3.6 Scalar fields

Having developed the completely general formalism for treating linear perturbations in a cosmological background, it is now time to apply the formalism to a concrete and important example, that of the system of N minimally coupled scalar fields φ_A introduced in Section 2.3.1. Here I will decompose the scalar fields into homogeneous background parts and "small" perturbations,

$$\varphi_A(t, x^i) = {}^0\varphi_A(t) + \delta\varphi_A(t, x^i). \tag{3.311}$$

First I will calculate the perturbed energy-momentum tensor $\delta T^\mu{}_\nu$ for a set of scalar fields, and then I will calculate the equations of motion for the metric functions as well as for the field perturbations $\delta\varphi_A$.

3.6.1 Perturbed energy-momentum tensor

Recall that in Section 2.3.1 I derived the general form of the scalar field energy-momentum tensor directly from the Lagrangian. The result was

$$T^{\mu}{}_{\nu} = \varphi^{,\mu} \cdot \varphi_{,\nu} - \frac{1}{2}\delta^{\mu}{}_{\nu} \left(\varphi^{,\lambda} \cdot \varphi_{,\lambda} + 2V(\varphi_A) \right), \tag{3.312}$$

where I defined

$$\varphi \cdot \varphi \equiv \varphi_A \varphi^A. \tag{3.313}$$

Now I will perturb the scalar fields according to Eq. (3.311). However, for consistency with Einstein's equation, I must also perturb the metric, according to

$$g_{\mu\nu}(t, x^i) = {}^0g_{\mu\nu}(t) + \delta g_{\mu\nu}(t, x^i). \tag{3.314}$$

Thus the perturbation of Eq. (3.312) reads

$$\delta T^{\mu}{}_{\nu} \equiv \delta\left(T^{\mu}{}_{\nu}\right) = \delta\left(g^{\mu\lambda}\varphi_{,\lambda} \cdot \varphi_{,\nu}\right) - \frac{1}{2}\delta^{\mu}{}_{\nu}\left[\delta\left(g^{\lambda\rho}\varphi_{,\lambda} \cdot \varphi_{,\rho}\right) + 2\delta V(\varphi_A)\right], \tag{3.315}$$

where for example

$$\delta\left(g^{\mu\lambda}\varphi_{,\lambda} \cdot \varphi_{,\nu}\right) = -\left(\delta g^{\mu\lambda}\right) {}^0\varphi_{,\lambda} \cdot {}^0\varphi_{,\nu} + {}^0g^{\mu\lambda}\delta\varphi_{,\lambda} \cdot {}^0\varphi_{,\nu} + {}^0g^{\mu\lambda}\delta\varphi_{,\nu} \cdot {}^0\varphi_{,\lambda} \tag{3.316}$$

and here I have used Eq. (3.206). A typical term in this expression is

$$\delta\left(g^{0\lambda}\varphi_{,\lambda} \cdot \varphi_{,0}\right) = -\left(\delta g^{00}\right) {}^0\varphi_{,0} \cdot {}^0\varphi_{,0} + 2\left({}^0g^{00}\right) \delta\varphi_{,0} \cdot {}^0\varphi_{,0} \tag{3.317}$$

$$= \frac{2}{a^2}\left({}^0\varphi' \cdot {}^0\varphi'\phi - \delta\varphi' \cdot {}^0\varphi'\right), \tag{3.318}$$

where I have used

$$\delta g^{00} = -\frac{2}{a^2}\phi, \tag{3.319}$$

which follows from the general perturbed metric (3.2). Also we have

$$\delta V(\varphi_A) = V_{,\varphi} \cdot \delta\varphi. \tag{3.320}$$

Combining all such terms, and dropping henceforth the superscript ^{0} on the background quantities, I find to linear order

$$\delta T^0_{\ 0} = -\dot{\varphi} \cdot \delta\dot{\varphi} + \dot{\varphi} \cdot \dot{\varphi}\phi - V_{,\varphi} \cdot \delta\varphi, \tag{3.321}$$

$$\delta T^0_{\ i} = -\frac{1}{a}\dot{\varphi} \cdot \delta\varphi_{,i}, \tag{3.322}$$

$$\delta T^i_{\ j} = \left(\dot{\varphi} \cdot \delta\dot{\varphi} - \dot{\varphi} \cdot \dot{\varphi}\phi - V_{,\varphi} \cdot \delta\varphi\right) \delta^i_{\ j}. \tag{3.323}$$

Thus the defining relations (3.68), (3.71), and Eq. (3.72) give

$$\delta\rho = \dot{\varphi}\cdot\delta\dot{\varphi} - \dot{\varphi}\cdot\dot{\varphi}\phi + V_{,\varphi}\cdot\delta\varphi, \tag{3.324}$$

$$q = -\dot{\varphi}\cdot\delta\varphi, \tag{3.325}$$

$$\delta P = \dot{\varphi}\cdot\delta\dot{\varphi} - \dot{\varphi}\cdot\dot{\varphi}\phi - V_{,\varphi}\cdot\delta\varphi, \tag{3.326}$$

$$\Pi = 0, \tag{3.327}$$

and of course all of the vector or tensor matter variables vanish for the scalar field. It is important to point out here that the anisotropic stress Π vanishes for any system of minimally-coupled scalar fields, to linear order. Also, the energy density and pressure perturbations look like perturbations of the homogeneous background relations, (2.128) and (2.130), considering that the coordinate time is perturbed relative to the proper time.

It is now very straightforward to write down the perturbed energy and momentum conservation equations, (3.84) and (3.91), for the scalar field matter variables listed above. The perturbed energy conservation law gives the equation of motion for $\delta\varphi_A$ (recall from Section 2.3.1 that energy-momentum conservation implies the Klein-Gordon equation), though I derive this equation more directly below. The momentum conservation law turns out to be automatically satisfied for the scalar field.

The value of the scalar field perturbation $\delta\varphi_A$ will of course depend on the gauge. The change in $\delta\varphi_A$ under a gauge transformation $t \to t - T$ is given by the transformation for a scalar quantity, Eq. (3.104). In the present case this becomes the rather obvious

$$\delta\tilde{\varphi}_A = \delta\varphi_A + \dot{\varphi}_A T. \tag{3.328}$$

It is easy to check that this transformation law is consistent with the previously calculated transformations of the matter variables in Eqs. (3.324) to (3.326).

3.6.2 Perturbed equations of motion

The four distinct components of Einstein's equation are easy to write down now, using the general expressions (3.279) and (3.282) to (3.284) with the scalar field values for the matter variables derived in the previous section. The result is

$$3H(\dot{\psi} + H\phi) + \dot{H}\phi - \frac{1}{a^2}\nabla^2(\psi + H\sigma) = 4\pi G(-\dot{\varphi}\cdot\delta\dot{\varphi} - V_{,\varphi}\cdot\delta\varphi), \tag{3.329}$$

$$\dot{\psi} + H\phi = 4\pi G\dot{\varphi}\cdot\delta\varphi, \tag{3.330}$$

$$\psi - \phi + \dot{\sigma} + H\sigma = 0, \tag{3.331}$$

$$\ddot{\psi} + H(3\dot{\psi} + \dot{\phi}) + (3H^2 + \dot{H})\phi = 4\pi G(\dot{\varphi}\cdot\delta\dot{\varphi} - V_{,\varphi}\cdot\delta\varphi), \tag{3.332}$$

for, respectively, the energy constraint, momentum constraint, off-diagonal space-space, and trace part of space-space equations. These equations are valid in any gauge. I have made a slight simplification by using the background relations $\dot{\varphi} \cdot \dot{\varphi} = \rho + P$ and (2.87). The energy constraint and trace equations can be put into a more compact form by subtracting from them $3H$ times the momentum constraint equation and using the background equation of motion. The results are

$$\dot{H}\phi - \frac{1}{a^2}\nabla^2(\psi + H\sigma) = 4\pi G(-\dot{\varphi} \cdot \delta\dot{\varphi} + \ddot{\varphi} \cdot \delta\varphi), \tag{3.333}$$

$$\ddot{\psi} + H\dot{\phi} + \dot{H}\phi = 4\pi G(\dot{\varphi} \cdot \delta\dot{\varphi} + \ddot{\varphi} \cdot \delta\varphi). \tag{3.334}$$

The second equation here is just the time derivative of the momentum constraint, which demonstrates that the components of Einstein's equation are not independent, but are of course related by the contracted Bianchi identity (2.34).

The vanishing of the anisotropic stress in the off-diagonal equation resulted in a significant simplification of the trace equation. In fact, the off-diagonal equation reduces the number of physical metric functions from two (ϕ, ψ, and σ with one gauge degree of freedom) to one. This is most apparent if we write the equations in zero-shear or longitudinal gauge. Setting $\sigma = 0$, the off-diagonal equation becomes

$$\Psi = \Phi. \tag{3.335}$$

(Recall that upper-case symbols designate values in longitudinal gauge.) Thus the equations of motion in longitudinal gauge take the very simple form

$$3H(\dot{\Phi} + H\Phi) + \dot{H}\Phi - \frac{1}{a^2}\nabla^2\Phi = 4\pi G(-\dot{\varphi} \cdot \delta\dot{\varphi} - V_{,\varphi} \cdot \delta\varphi), \tag{3.336}$$

$$\dot{\Phi} + H\Phi = 4\pi G\dot{\varphi} \cdot \delta\varphi, \tag{3.337}$$

$$\ddot{\Phi} + 4H\dot{\Phi} + (3H^2 + \dot{H})\Phi = 4\pi G(\dot{\varphi} \cdot \delta\dot{\varphi} - V_{,\varphi} \cdot \delta\varphi). \tag{3.338}$$

The scalar field perturbation $\delta\varphi_A$ must also be expressed in longitudinal gauge here.

An equation of motion for the field perturbation $\delta\varphi_A$ can be derived from the components of Einstein's equation, Eqs. (3.329) to (3.332) [or, as mentioned already, from the energy conservation equation, (3.84)]. However, this approach involves unilluminating algebra, and so instead I will simply perturb the exact Klein-Gordon equation, $\Box\varphi_A = V_{,\varphi_A}$. With the identity (2.30) we can write

$$\Box\varphi^A = \frac{1}{\sqrt{g}}\left(\sqrt{g}g^{\mu\nu}\varphi^A_{,\nu}\right)_{,\mu}. \tag{3.339}$$

In perturbing this expression, there will be four terms. Using Eq. (2.14) we have

$$\delta\sqrt{g} = \frac{1}{2}\sqrt{g}g^{\mu\nu}\delta g_{\mu\nu} = \frac{1}{2}\sqrt{g}\delta g \tag{3.340}$$

and hence

$$\delta\left(\frac{1}{\sqrt{g}}\right) = -\frac{1}{2\sqrt{g}}\delta g. \tag{3.341}$$

We will also need

$$\sqrt{g} = a^4 \tag{3.342}$$

for Cartesian coordinates in a flat background. Since the scalar field equation of motion cannot contain vector- or tensor-derived parts, the only metric perturbations we need are

$$\delta g^{00} = -\frac{2\phi}{a^2}, \qquad \delta g^{0i} = -\frac{B^{,i}}{a^2}, \qquad \delta g^{ij} = -\frac{2}{a^2}\left(\psi\gamma^{ij} - E^{,ij}\right) \tag{3.343}$$

which imply that the trace is

$$\delta g = 2\phi - 6\psi + 2\nabla^2 E. \tag{3.344}$$

Finally, we will need the variation

$$\delta V_{,\varphi_A} = V_{,\varphi_A\varphi_B}\delta\varphi_B \equiv V_{,\varphi_A\varphi}\cdot\delta\varphi \tag{3.345}$$

Combining these results, I find

$$\Box\delta\varphi_A + \left(\dot{\phi} + 3\dot{\psi} - \frac{1}{a^2}\nabla^2\sigma\right)\dot{\varphi}_A - 2\phi V_{,\varphi_A} = V_{,\varphi_A\varphi}\cdot\delta\varphi \tag{3.346}$$

for the perturbed Klein-Gordon equation, valid in any gauge, where

$$\Box\delta\varphi_A = -\delta\ddot{\varphi}_A - 3H\delta\dot{\varphi}_A + \frac{1}{a^2}\nabla^2\delta\varphi_A \tag{3.347}$$

to first order. In longitudinal gauge this equation takes the more compact form

$$\Box\delta\varphi_A + 4\dot{\Phi}\dot{\varphi}_A - 2\Phi V_{,\varphi_A} = V_{,\varphi_A\varphi}\cdot\delta\varphi. \tag{3.348}$$

Again, $\delta\varphi_A$ must also be in longitudinal gauge in this final expression.

Chapter 4

Parametric Resonance and Backreaction

4.1 Introduction

It is now time to apply the results I have carefully derived in Chapters 2 and 3 to a specific scalar field inflationary model. The model will provide a period of ordinary chaotic inflation, as described in Section 2.3.2. However, its importance will lie in the description it provides of the post-inflationary period of reheating, when the inflaton field begins oscillating. It has long been realized that reheating is a crucial part of the inflationary scenario. During reheating the large energy density contained within the coherently oscillating inflaton field is converted into particle excitations of whatever fields are coupled to the inflaton, vastly increasing the temperature and entropy density and setting the stage for the standard big bang phase. The term "reheating" refers to the appearance of this thermal state after the "cold" emptiness of inflation, which was possibly preceeded by a very early thermal state. If inflation is ever to be a useful picture for describing the early universe, then it is essential to understand the details of how the vacuum energy is transformed into familiar particles.

The original (or "elementary") theory of reheating described the process of the slow perturbative decay of the zero-mode inflaton particles into the other types of particles the inflaton is coupled to [21]. In recent years it has been realized that reheating can occur much more efficiently than the old theory described, through the process of parametric resonance [1, 31, 32]. Field modes within certain resonance bands in momentum space can grow exponentially with time, defining the "preheating" era. This name originates from the fact that the spectrum produced by parametric resonance is highly non-thermal, and thus the fields after preheating must still thermalize.

The possibility of resonant growth of linear scalar *metric* perturbations was first studied in [2]. Recently it has been argued that the resonance of scalar metric perturbations can extend to $k \ll aH$ (where k is a comoving momentum), *i.e.* that super-Hubble perturbations can be amplified [33, 34]. This opens up the possibility of new observational consequences, since the scales relevant to the cosmic microwave background and large-scale structure are much larger than the Hubble radius during preheating. In [35] it was found that simple single-field chaotic inflation models do not exhibit super-Hubble growth beyond what is expected in the absence of parametric resonance. The absence of parametric amplification of super-Hubble modes in these single-field models was shown to hold in a full nonlinear treatment [3], and a general no-go theorem in these models was suggested in [36]. Preheating has also been studied in non-minimally coupled single-field [4] and multi-field [5] models.

For the first model which was claimed to exhibit growth of super-Hubble metric perturbations beyond that of the usual theory of reheating [34] (see also [37]), it was soon realized that the growth was unimportant since it followed a period of exponential *damping* during inflation [38–40]. This damping of super-Hubble modes arises because the field perturbations which are amplified during preheating have an effective mass greater than the Hubble parameter during inflation. This results in a very "blue" power spectrum at the end of inflation, with a severe deficit at the largest scales [40]. The relatively plentiful small-scale modes can also grow resonantly during preheating. Thus the end of parametric resonance occurs when the backreaction of the dominant small-scale modes becomes important, and the cosmological-scale modes are still negligible. An obvious class of models to study, then, consists of those with small masses during inflation and strong super-Hubble resonance [41, 42]. A simple example was provided by Bassett and Viniegra [43], namely that of a massless self-coupled inflaton φ coupled to another scalar field χ, *i.e.* a model with potential

$$V(\varphi_A) = \frac{\lambda}{4}\varphi^4 + \frac{g^2}{2}\varphi^2\chi^2. \tag{4.1}$$

This model has been studied in detail, but in the absence of metric perturbations, by Greene *et al.* [44], who found that the model contains a strong resonance band for χ fluctuations which extends to $k = 0$ for the choice $g^2 = 2\lambda$. Bassett and Viniegra [43] found that super-Hubble metric perturbations are resonantly amplified as well in this model (see also [42, 45]).

To date, however, a thorough analysis of the parametric amplification of super-Hubble-scale metric fluctuations in the model (4.1), including the effects of backreaction on the evolution of the fluctuations, has not been performed. The backreaction of the growing modes on the background fields is expected to shut the growth down at some point, but exactly when? Backreaction is also the only hope to make models which exhibit parametric amplification of super-Hubble cosmological perturbations compatible with the Cosmic Background Explorer (COBE) normalization [46].

In this chapter I will investigate the effects of backreaction on the growth of matter and metric fluctuations using the Bassett and Viniegra model (4.1) as my toy model. I will begin in Section 4.2 with a detailed description of the model, including an analytical and numerical description of the parametric resonance it exhibits. I will then continue in Section 4.3 by studying the growth of scalar field and scalar metric perturbations, including the effect of backreaction in the Hartree approximation. I carefully treat the evolution during inflation, which can be very important for super-Hubble scales. I will compare the large-scale normalization predicted for this model with the COBE value, and find that, although backreaction is crucial in limiting the growth of the fluctuations, the final amplitude is larger than allowed by the COBE normalization (for supersymmetry-motivated coupling constant values). Note that the final amplitude of fluctuations in my model is independent of the scalar field coupling constant (unlike what happens without parametric resonance effects). In Section 4.4 I also extend the model to study the effect of χ-field self-coupling, which can be important in limiting the growth of fluctuations. The importance of this work is that it indicates that the dynamics of super-Hubble scales during preheating must be carefully analyzed in inflationary models in order to decide whether or not this dynamics invalidates the standard inflationary prediction of metric perturbations.

4.2 Model and linearized dynamics

4.2.1 Equations of motion

The inflationary model I will consider in this chapter consists of a set of scalar fields minimally coupled to standard Einstein gravity, so that the system is described by a Lagrangian density of the general form

$$\mathcal{L} = \mathcal{L}_{\mathrm{g}} + \mathcal{L}_{\mathrm{m}} = \frac{\sqrt{g}}{2}\left(\frac{m_{\mathrm{P}}^2}{8\pi}R - \varphi^{;\mu}\cdot\varphi_{;\mu} - 2V(\varphi_A)\right) \tag{4.2}$$

[recall Eqs. (2.6) and (2.118)]. I continue to use the shorthand notation

$$\varphi\cdot\varphi \equiv \varphi_A\varphi^A, \tag{4.3}$$

where $A = 1, \ldots, N$ labels the scalar field. In this chapter I will only consider the case of $N = 2$ real scalar fields, $\varphi_1 \equiv \varphi$ and $\varphi_2 \equiv \chi$. The potential that I will consider through most of this chapter corresponds to a massless, self-interacting φ field coupled quartically to the χ field,

$$V(\varphi_A) = \frac{\lambda}{4}\varphi^4 + \frac{g^2}{2}\varphi^2\chi^2. \tag{4.4}$$

(In Section 4.4 I will consider the effect of a self-interacting χ term.) In these two-field models the field φ will drive inflation and hence is referred to as the inflaton field.

Note that the behaviour of this system is expected to be robust under the addition of a small mass term $m_\varphi^2\varphi^2$ with $m_\varphi \ll \sqrt{\lambda}m_{\mathrm{P}}$ and for the ratio of coupling constants satisfying $g/\sqrt{\lambda} < \sqrt{\lambda}m_{\mathrm{P}}/m_\varphi$ [44]. In particular, this will be the case for supersymmetric models, which motivate the choice $g^2 = 2\lambda$ [47]. In addition, I will show that large values of $g/\sqrt{\lambda}$ are in fact inconsistent with the significant amplification of super-Hubble modes. On the other hand, for $g/\sqrt{\lambda} > \sqrt{\lambda}m_{\mathrm{P}}/m_\varphi$, the theory of "stochastic resonance" for a massive inflaton may need to be applied [31].

In describing the dynamics of our inflationary model, we can use the techniques developed in Chapter 3. Namely, to satisfy the cosmological principle, we can separate the metric and matter variables into homogeneous background parts and small perturbations that should satisfy the linearized evolution equations. That is, we can write

$$\tilde{g}_{\mu\nu}(t,x^i) = {}^0g_{\mu\nu}(t) + \delta g_{\mu\nu}(t,x^i) \tag{4.5}$$

and

$$\tilde{\varphi}_A(t,x^i) = {}^0\varphi_A(t) + \delta\varphi_A(t,x^i). \tag{4.6}$$

Here, as discussed in Section 3.4.1, the tildes indicate that these decompositions are only defined up to a gauge transformation. Thus, in writing the perturbation equations of motion we have the freedom to specify a gauge choice that simplifies those equations.

Homogeneous dynamics

The equations of motion for the homogeneous parts of the inflaton and χ fields are determined by the general homogeneous Klein-Gordon equation, Eq. (2.126), which arises from the matter-sector variational principle $\delta S_{\mathrm{m}} = 0$. For the potential (4.4), the evolution equations become (henceforth I drop the superscript ^{0} on background quantities)

$$\ddot{\varphi} + 3H\dot{\varphi} + \lambda\varphi^3 + g^2\chi^2\varphi = 0 \tag{4.7}$$

and

$$\ddot{\chi} + 3H\dot{\chi} + g^2\varphi^2\chi = 0. \tag{4.8}$$

Recall from Section 2.3.1 that for scalar fields, the covariant energy-momentum conservation law is equivalent to the Klein-Gordon equation, so it provides no additional information here.

To complete the homogeneous background dynamics we must specify the evolution of the background spacetime metric. In a homogeneous cosmology, the spacetime is described by the FRW metric, Eq. (2.51),

$$ds^2 = -dt^2 + a^2(t)\gamma_{ij}(x^k)\, dx^i dx^j. \tag{4.9}$$

The evolution of the metric is provided by Einstein's equation, which I derived from the gravitational variational principle $\delta S_{\rm g} = 0$ in Section 2.1.1. I found the time-time component of Einstein's equation to be the Friedmann equation, Eq. (2.83), which, with the scalar field energy density Eq. (2.128), becomes in the two-field case

$$H^2 = \frac{8\pi}{3m_{\rm P}^2}\left[\frac{1}{2}\dot{\varphi}^2 + \frac{1}{2}\dot{\chi}^2 + V(\varphi,\chi)\right]. \tag{4.10}$$

Here I have assumed spatially flat homogeneous constant-time hypersurfaces, *i.e.* $\mathcal{K} = 0$, which is well justified soon after inflation starts, as I discussed in Section 2.3.2. Also, I have ignored any explicit cosmological constant. Since the Einstein field equation implies the covariant conservation of the energy-momentum tensor, no new information is provided by the space-space dynamical equation.

Perturbation dynamics

The completely general metric perturbation $\delta g_{\mu\nu}$ can be decomposed into scalar-, vector-, and tensor-derived parts as described in Section 3.1. However, recall from Section 3.6.1 that the perturbed scalar field energy-momentum tensor contains only scalar-derived parts (as should be obvious!). Thus so must the first-order perturbed Einstein tensor, when the matter sector is exclusively scalar-field. [Importantly this will no longer be true for the second-order perturbed Einstein tensor, as I explained at the end of Section 3.1. Also, the vanishing of of $\delta G^{\mu(\rm t)}_{\ \nu}$ does not imply the vanishing of $\delta g^{\mu(\rm t)}_{\ \nu}$ (recall Eq. (3.268)). However, gravitational waves do not couple to scalars at linear order and therefore can be treated separately.] Therefore we can write the general-gauge perturbed metric as the scalar form [recall Eq. (3.8)],

$$ds^2 = a^2(\eta)\left\{-(1+2\phi)d\eta^2 + 2B_{|i}dx^i d\eta + \left[(1-2\psi)\gamma_{ij} + 2E_{|ij}\right]dx^i dx^j\right\}. \tag{4.11}$$

To simplify the appearance of this metric, I will choose for this chapter to work in longitudinal (zero-shear) gauge, which I described in Section 3.4.4. With this choice, $B = E = 0$ and the metric becomes Eq. (3.192),

$$ds^2 = a^2(\eta)\left[-(1+2\Phi)d\eta^2 + (1-2\Psi)dx^i dx_i\right] \tag{4.12}$$

(recall from Section 3.4.4 that the upper case symbols Φ and Ψ designate zero shear gauge quantities). As discussed in Section 3.2, the metric function Φ determines the lapse function, while Ψ determines the three-curvature of the zero-shear hypersurfaces. We can make a final simplification to the line element by recalling Section 3.6.2, where I showed that the vanishing of the first-order anisotropic stress Π in the off-diagonal space-space Einstein equation led directly to the equality of the remaining metric functions,

$$\Psi = \Phi. \tag{4.13}$$

Thus the metric can finally be written, using proper time,

$$ds^2 = -(1 + 2\Phi)dt^2 + a^2(t)(1 - 2\Phi)dx^i dx_i. \tag{4.14}$$

I will Fourier-decompose all perturbations $X(t, x^i)$ into comoving wavenumber k, according to

$$X_k(t) = \frac{1}{(2\pi)^{3/2}} \int X(t, x^i) e^{ik_i x^i} \, d^3x. \tag{4.15}$$

Then the linear partial differential evolution equations decouple into ordinary differential equations for each mode k. I will often drop the subscript k on these modes to reduce clutter. The evolution of the perturbations is described, to first order, by the components of the linearized Einstein equation derived in Section 3.5. The energy constraint equation, applied to scalar fields and written in longitudinal gauge, is Eq. (3.336),

$$3H\dot{\Phi} + \left(\frac{k^2}{a^2} + 3H^2\right)\Phi = -\frac{4\pi}{m_{\mathrm{P}}^2}\left(\dot{\varphi}\cdot\delta\dot{\varphi} - \Phi\dot{\varphi}\cdot\dot{\varphi} + V_{,\varphi}\cdot\delta\varphi\right). \tag{4.16}$$

Similarly, I found the momentum constraint equation to be Eq. (3.337),

$$\dot{\Phi} + H\Phi = \frac{4\pi}{m_{\mathrm{P}}^2}\dot{\varphi}\cdot\delta\varphi. \tag{4.17}$$

To complete the description of the perturbation dynamics, I will use the perturbed Klein-Gordon equation, Eq. (3.348),

$$\delta\ddot{\varphi}_A + 3H\delta\dot{\varphi}_A + \frac{k^2}{a^2}\delta\varphi_A + V_{,\varphi_A\varphi}\cdot\delta\varphi = 4\dot{\Phi}\dot{\varphi}_A - 2V_{,\varphi_A}\Phi, \tag{4.18}$$

rather than the space-space Einstein equation. Note that in each of these equations the field perturbations $\delta\varphi_A$ must be specified in longitudinal gauge. Equations (4.16) and (4.17) can be combined to give

$$\Phi = \frac{\dot{\varphi}\cdot\delta\dot{\varphi} + (3H\dot{\varphi} + V_{,\varphi})\cdot\delta\varphi}{-(m_{\mathrm{P}}^2/4\pi)(k/a)^2 + \dot{\varphi}\cdot\dot{\varphi}}, \tag{4.19}$$

which fixes Φ once the matter fields are known.

In Section 5.1.2 I show that the quantity

$$\zeta_{\text{MFB}} = \Phi - \frac{H}{\dot{H}}\left(\dot{\Phi} + H\Phi\right) \tag{4.20}$$

is conserved for adiabatic perturbations on large scales, *i.e.* $k/a \ll H$, and when the anisotropic stress Π vanishes:

$$(\rho + P)\dot{\zeta}_{\text{MFB}} = 0. \tag{4.21}$$

Geometrically, as I show in Section 5.1.2, ζ_{MFB} is simply the comoving curvature perturbation ψ_q (when $\Pi = 0$), and approaches the uniform density curvature perturbation, ψ_ρ, on large scales. Since ζ is conserved for long-wavelength adiabatic modes (I drop the subscript MFB for the remainder of this chapter), the behaviour of ζ will be crucial in determining whether non-adiabatic metric perturbations are amplified by resonance in the current model, Eq. (4.4).

4.2.2 Analytical theory of parametric resonance

Homogeneous backgrounds

The two-field model I am considering in this chapter, Eq. (4.4), should produce inflation at early enough times if it is to be a viable model. Indeed, for initial χ sufficiently small, the Klein-Gordon and Friedmann equations, (4.7) and (4.10), approach those of the inflationary single-field case described in Section 2.3.2. Since the effective masses of χ and φ fields are comparable (for $g^2 \sim \lambda$), an initially small χ remains small until parametric resonance begins, so the inflationary dynamics is essentially that of $(\lambda/4)\varphi^4$ chaotic inflation. That is, for $\varphi \gtrsim m_{\text{P}}$ the universe undergoes slow-roll inflation, with φ undergoing overdamped decay, and with H decreasing slowly ($|\dot{H}| \ll H^2$) and the scale factor a increasing approximately exponentially with time. The dynamics during the inflationary phase is illustrated in Fig. 4.1. This figure shows a numerical simulation of the homogeneous background equations of motion, Eqs. (4.7) and (4.10), for the case $\chi = 0$. The integration began at $\varphi = 4m_{\text{P}}$, and we see that approximately 50 e-folds of inflation occur before the oscillatory preheating phase begins. In addition to the decaying φ and H curves, the figure also shows the equation of state, which can be seen to be close to the value $w = -1$ during most of inflation.

As slow-roll ends, the damping term $3H\dot{\varphi}$ becomes less important in Eq. (4.7) and the homogeneous field begins underdamped oscillations about $\varphi = 0$. This marks the start of the preheating period. Averaged over several oscillations, the equation of state (in the absence of backreaction) is very nearly

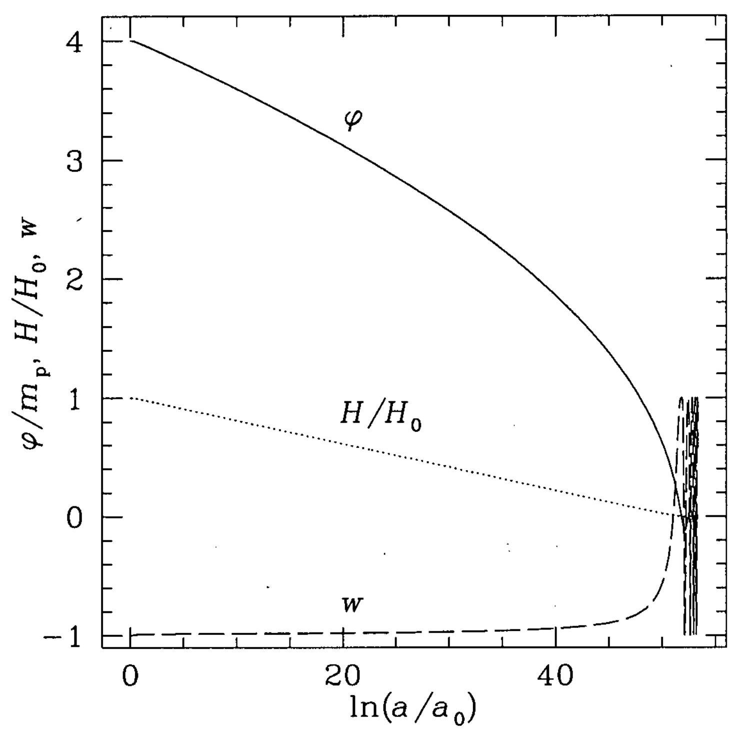

Figure 4.1: Numerical simulation of homogeneous background scalar field $\varphi/m_{\rm P}$, equation of state w, and H/H_0, where H_0 is the Hubble parameter at the start of the simulation, for potential $\lambda\varphi^4/4$. We see about 50 e-folds of inflation before the underdamped oscillations begin.

that of a radiation-dominated universe [32], and the amplitude of the inflaton's oscillations decays as a^{-1}. This is a consequence of the near conformal invariance of this massless model, which considerably simplifies the treatment of parametric resonance as compared with the massive case [31, 44].

To see this explicitly, consider the conformally scaled fields

$$\tilde{\varphi}_A \equiv a\varphi_A. \tag{4.22}$$

Rewriting the homogeneous Klein-Gordon equation (4.7) in terms of conformal time and the conformal fields, I find (setting the homogeneous χ to zero)

$$\tilde{\varphi}'' + \left(-\frac{a''}{a} + \lambda\tilde{\varphi}^2\right)\tilde{\varphi} = 0. \tag{4.23}$$

The term proportional to $a''/a = a^2R/6$ would have been precisely canceled had I chosen a conformally coupled Lagrangian with $\xi = 1/6$, as I discussed at the beginning of Section 2.3.1. However, during the oscillatory preheating phase this term is negligible: Using the Einstein equations, (2.83) and (2.84), and the scalar field expressions for energy density and pressure, Eqs. (2.128) and (2.130), I find

$$\frac{a''}{a} = \frac{4\pi}{3}\left(\lambda\tilde{\varphi}^2\frac{\varphi^2}{m_{\rm P}^2} - \frac{\varphi'^2}{m_{\rm P}^2}\right). \tag{4.24}$$

Thus the a''/a term in Eq. (4.23) is of order $\varphi^2/m_{\rm P}^2$ smaller than the $\lambda\tilde{\varphi}^2$ term. But $\varphi^2/m_{\rm P}^2 \simeq 10^{-2}$ at the end of inflation and only decreases thereafter. Thus to enable an analytical description of the preheating dynamics, I will ignore the a''/a term in this subsection. The validity of this approximation will be confirmed when I perform numerical integration of the *exact* equations of motion later in this chapter.

The homogeneous Klein-Gordon equation now takes the very simple form

$$\tilde{\varphi}'' + \lambda\tilde{\varphi}^3 = 0. \tag{4.25}$$

This equation describes undamped oscillations of the conformal field $\tilde{\varphi}$ in a quartic well, and hence we can immediately conclude that, as claimed above, the amplitude of oscillation of the physical field φ decays like a^{-1}. The period of oscillation is constant in conformal time. To obtain an analytical solution to the homogeneous inflaton equation (4.25), we can begin with the first integral of this equation,

$$\frac{1}{2}\tilde{\varphi}'^2 + \frac{\lambda}{4}\tilde{\varphi}^4 = \text{const} \equiv \frac{\lambda}{4}\tilde{\varphi}_0^4, \tag{4.26}$$

where $\tilde{\varphi}_0$ is the amplitude of oscillation of $\tilde{\varphi}$. Now defining a scaled conformal time x by

$$x \equiv \sqrt{\lambda}\tilde{\varphi}_0\eta \tag{4.27}$$

and a scaled field by

$$f(x) \equiv \frac{\tilde{\varphi}}{\tilde{\varphi}_0}, \tag{4.28}$$

Eq. (4.26) can be easily rewritten as

$$\sqrt{2}\int_f^1 \frac{df^*}{\sqrt{1-f^{*4}}} = x - x_0, \tag{4.29}$$

where the constant x_0 represents the arbitrariness of the time origin. The integral in this last equation is an elliptic integral of the first kind. It can be evaluated in terms of the Jacobian elliptic functions [48]; the result is

$$\sqrt{2}\int_f^1 \frac{df^*}{\sqrt{1-f^{*4}}} = \text{cn}^{-1}\left(f, \frac{1}{\sqrt{2}}\right), \tag{4.30}$$

where the inverse elliptic function $\mathrm{cn}^{-1}(f, 1/\sqrt{2})$ is known as an (inverse) *cosine amplitude*. The second argument, $1/\sqrt{2}$, is called the *modulus*. We can now finally write the conformal field solution as

$$\tilde{\varphi}(x) = \tilde{\varphi}_0 \,\mathrm{cn}\left(x - x_0, \frac{1}{\sqrt{2}}\right). \tag{4.31}$$

This elliptic cosine function is very similar to the ordinary trigonometric cosine: its Fourier series expansion is dominated by the fundamental harmonic. (We have [48]

$$\mathrm{cn}\left(x, \frac{1}{\sqrt{2}}\right) \simeq 0.9550 \cos\left(\frac{2\pi x}{T}\right) + 0.04305 \cos\left(\frac{6\pi x}{T}\right) + \dots \tag{4.32}$$

where

$$T = \sqrt{2\pi}\,\frac{\Gamma(1/4)}{\Gamma(3/4)} \simeq 7.4163 \tag{4.33}$$

is the period of the elliptic cosine.)

In Fig. 4.2 I illustrate the evolution of the homogeneous background quantities in the case $\chi = 0$. The plot covers the time period from just before the end of inflation to several oscillations of φ into the preheating period. The exact equations of motion Eqs. (4.7) and (4.10) were integrated numerically. The conformal time parameter x is used in the plot, and it is apparent that the oscillations have a constant period with a value in agreement with Eq. (4.33).

Perturbations

In the absence of metric perturbations, the linearized dynamics in the model described above is known to exhibit parametric resonance of the scalar field perturbations during preheating [44]. To demonstrate that the resonance persists in the presence of the metric perturbations, I will write the linearized equation of motion (4.18) in terms of the conformal fields and scaled conformal time, x. This is similar to the approach taken in Ref. [43]. I will assume the homogeneous χ background is negligible and ignore the a''/a terms, as I did for the homogeneous inflaton equation of motion. For the χ field perturbations, I obtain

$$\frac{\partial^2 \delta\tilde{\chi}_k}{\partial x^2} + \left(\kappa^2 + \frac{g^2}{\lambda}\frac{\tilde{\varphi}^2}{\tilde{\varphi}_0^2}\right)\delta\tilde{\chi}_k = 0, \tag{4.34}$$

where $\kappa^2 \equiv k^2/(\lambda\tilde{\varphi}_0^2)$ is a dimensionless momentum parameter. Similarly, for the inflaton perturbations I find

$$\frac{\partial^2 \delta\tilde{\varphi}_k}{\partial x^2} + \left(\kappa^2 + 3\frac{\tilde{\varphi}^2}{\tilde{\varphi}_0^2}\right)\delta\tilde{\varphi}_k = \frac{1}{\lambda\tilde{\varphi}_0^2}\left(4\frac{\partial\Phi_k}{\partial x}\frac{\partial\tilde{\varphi}}{\partial x} - 2\lambda\Phi_k\tilde{\varphi}^3\right). \tag{4.35}$$

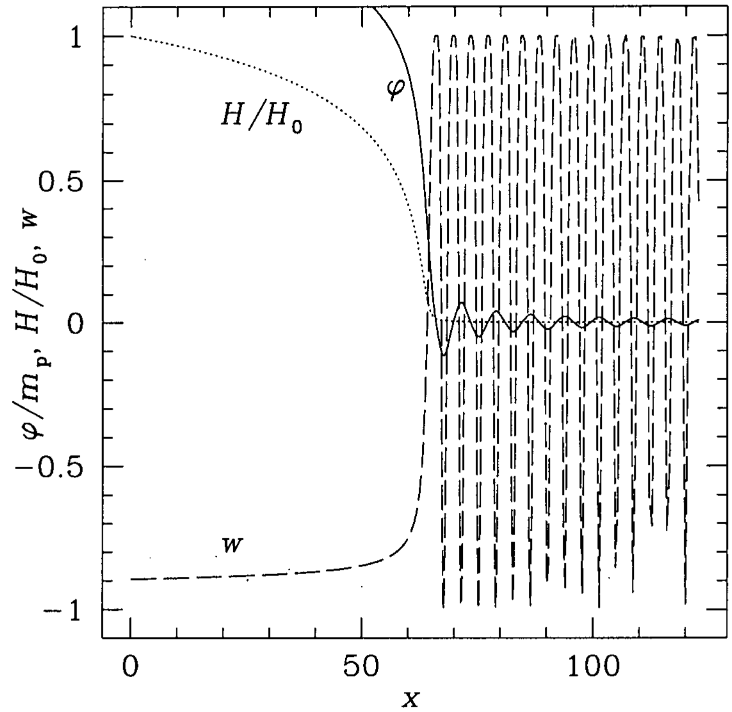

Figure 4.2: Numerical simulation of scalar field φ/m_P, Hubble parameter H, and equation of state w near the end of inflation and into the preheating stage. H_0 refers to the value at the start of the integration, when $\varphi \simeq 1.38m_\mathrm{P}$.

Notice that for the $\delta\tilde{\chi}$ equation there is no coupling to the metric perturbations at first order, when the homogeneous χ vanishes. Thus the results of [44] also apply in this treatment. Namely, for Eq. (4.34), which is known as a Lamé equation, there are bands of parameter space in which the perturbations $\delta\tilde{\chi}$ grow exponentially with time x. However, for the $\delta\tilde{\varphi}$ equation, there are forcing terms which depend on the metric function Φ. Nevertheless, the resonant behaviour of the homogeneous part of Eq. (4.35) can still be studied—indeed the homogeneous part of Eq. (4.35) corresponds to Eq. (4.34) for the special case $g^2/\lambda = 3$.

Floquet form

Parametric resonant behaviour is well known in mechanical systems [49]. It differs from the even more familiar forced resonance in that no non-homogeneous forcing terms are required in the equations of motion. Instead it is a periodicity in a system parameter, in particular the frequency, that drives the resonance. To see how such instability is expected generically, first note that Eq. (4.34) is of the general form of a linear dynamical system with periodic coefficients,

$$\dot{\boldsymbol{\eta}} = \mathsf{A}(t)\boldsymbol{\eta}, \tag{4.36}$$

where $\boldsymbol{\eta}(t)$ is an N-dimensional real state vector and $\mathsf{A}(t)$ is a time-periodic matrix, $\mathsf{A}(t+T) = \mathsf{A}(t)$. [To write Eq. (4.34) in this form, let

$$\boldsymbol{\eta} = \begin{pmatrix} \delta\tilde{\chi}_k \\ \delta\dot{\tilde{\chi}}_k \end{pmatrix} \tag{4.37}$$

and

$$\mathsf{A} = \begin{pmatrix} 0 & 1 \\ -\omega^2 & 0 \end{pmatrix}, \tag{4.38}$$

where

$$\omega^2 \equiv \kappa^2 + \frac{g^2}{\lambda}\frac{\tilde{\varphi}^2}{\tilde{\varphi}_0^2}, \tag{4.39}$$

and rename t to x.]

Consider a set of N independent solutions to Eq. (4.36), $\boldsymbol{\eta}_n(t)$, where $n = 1, \ldots, N$. Because the dynamical equation is invariant under $t \to t+T$, the functions $\boldsymbol{\eta}_n(t+T)$ must also be solutions to Eq. (4.36), and hence can be expanded in the set $\boldsymbol{\eta}_n(t)$, *i.e.*

$$\boldsymbol{\eta}_n(t+T) = \alpha_{nm}\boldsymbol{\eta}_m(t), \tag{4.40}$$

where α_{nm} is a constant matrix and summation is implied. Now consider some arbitrary solution $\widetilde{\boldsymbol{\eta}}(t)$ to Eq. (4.36) with expansion

$$\widetilde{\boldsymbol{\eta}}(t) = \beta_n\boldsymbol{\eta}_n(t). \tag{4.41}$$

Combining the last two equations, we have

$$\widetilde{\boldsymbol{\eta}}(t+T) = \beta_n\alpha_{nm}\boldsymbol{\eta}_m(t). \tag{4.42}$$

The vector β_n will be an eigenvector of α_{nm} with eigenvalue $e^{\mu T}$, *i.e.*

$$\beta_n\alpha_{nm} = \beta_m e^{\mu T}, \tag{4.43}$$

if the secular equation is satisfied,

$$\det(\alpha_{mn} - e^{\mu T}\mathrm{I}_{mn}) = 0, \tag{4.44}$$

for $N \times N$ identity I_{mn}. Ignoring the special case when some roots coincide, we will have N real or complex-conjugate pair eigenvalues $e^{\mu T}$, since the solutions $\boldsymbol{\eta}(t)$ are real. Choosing β_n to be such an eigenvector, Eq. (4.42) becomes

$$\widetilde{\boldsymbol{\eta}}(t+T) = e^{\mu T}\beta_m \boldsymbol{\eta}_m(t), \tag{4.45}$$

and using the definition Eq. (4.41), this becomes

$$\widetilde{\boldsymbol{\eta}}(t+T) = e^{\mu T}\widetilde{\boldsymbol{\eta}}(t). \tag{4.46}$$

The most general function with this property is

$$\widetilde{\boldsymbol{\eta}}(t) = f(t)e^{\mu t}, \tag{4.47}$$

where f is periodic, $f(t+T) = f(t)$. To summarize, we generally expect either exponentially growing or decaying solutions to the original system (4.36) for $\mathrm{Re}(\mu) \neq 0$, or oscillatory solutions if $\mathrm{Im}(\mu) \neq 0$, in both cases modulated by a period-T oscillation. The form (4.47) for the solutions is called the *Floquet form*, and the values μ are called the *Floquet indices.* The demonstration above has a close parallel with Bloch's theorem regarding quantum mechanical states in spatially periodic potentials.

One property of the Floquet indices can be easily derived for the case of the two-dimensional non-dissipative system

$$\dot{\boldsymbol{\eta}} = \frac{d}{dt}\begin{pmatrix} \eta \\ \dot{\eta} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\omega^2(t) & 0 \end{pmatrix} \boldsymbol{\eta}, \tag{4.48}$$

of which our $\delta\tilde{\chi}$ equation (4.34) is an example. For the two solutions $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ satisfying Eq. (4.46) we have

$$\ddot{\eta}_1 + \omega^2\eta_1 = 0, \tag{4.49}$$
$$\ddot{\eta}_2 + \omega^2\eta_2 = 0. \tag{4.50}$$

Multiplying these two equations by η_2 and η_1, respectively, and subtracting, we find

$$\frac{d}{dt}(\dot{\eta}_1\eta_2 - \dot{\eta}_2\eta_1) = 0. \tag{4.51}$$

However, according to Eqs. (4.46) and (4.36),

$$(\dot{\eta}_1\eta_2 - \dot{\eta}_2\eta_1)\big|_{t=t_0+T} = e^{(\mu_1+\mu_2)T}\,(\dot{\eta}_1\eta_2 - \dot{\eta}_2\eta_1)\big|_{t=t_0} \tag{4.52}$$

where μ_1 and μ_2 are the Floquet indices corresponding to $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$, respectively. The last two equations are only consistent if

$$\mu_1 = -\mu_2. \tag{4.53}$$

Thus a pair of real eigenvalues corresponds to two eigenvectors growing and contracting like $e^{\mu t}$ and $e^{-\mu t}$, and a complex-conjugate pair must be purely imaginary, corresponding to rotating eigenvectors. This circumstance is a special case of the general case of symplectic systems, typified by Hamiltonian dynamics, which are characterized by pairs of eigenvalues λ, $1/\lambda$, a result which is related to the conservation of phase space volumes in such systems.

It is important to stress again that exponential instability is expected quite generally, as long as $\mathrm{Re}(\mu_\mathrm{u}) \neq 0$ for some eigenvalue μ_u. For then generic initial conditions will contain some component of the eigenvector corresponding to $|\mathrm{Re}(\mu_\mathrm{u})|$, and hence this component will grow exponentially according to Eq. (4.47). This closely parallels the situation in the linearized dynamics of an autonomous system about a fixed point. In fact, for the non-autonomous system (4.36), we can define an autonomous "time-T map" by sampling the dynamics at time intervals T. The state $\boldsymbol{\eta} = 0$ will be a fixed point of this map. For this linear map, there will be a spectrum of eigenvalues corresponding to the Floquet indices of the the full system (4.36) and describing the usual possible behaviours near the fixed point, namely that of a source, sink, spiral point, pure oscillation, etc.

While this should be sufficient mathematical motivation to look for instabilities in our scalar field perturbation equation, (4.34), a more physical motivation is provided by the observation that Eq. (4.34) is in the form of the equation of motion for a linear pendulum with periodically modulated frequency, and by recalling the familiar experience of modulating a children's swing's frequency by bending the knees and hence pumping the oscillations.

Relation to Mathieu's equation

While the Floquet form (4.47) tells us what the general solutions to the $\delta\tilde{\chi}$ equation of motion must look like, there still remains the problem of determining the indices μ as a function of the parameters κ^2 and g^2/λ. A closely related problem, but with purely sinusoidally varying frequency, is presented by *Mathieu's equation*,

$$\ddot{\eta} + \omega_0^2[1 + b\cos(\omega t)]\eta = 0. \tag{4.54}$$

This equation describes perturbation dynamics in a massive inflaton model [31],

$$V(\varphi_A) = \frac{1}{2}m^2\varphi^2 + \frac{1}{2}g^2\varphi^2\chi^2. \tag{4.55}$$

[Recall Eq. (2.157)]. The Mathieu equation can be analyzed analytically in the case $b \ll 1$ (see, *e.g.*, Ref. [49]). In this case Eq. (4.54) has resonant bands, with one positive Floquet exponent, near

$$\omega = 2\omega_0/n, \qquad n = 1, 2, \ldots. \tag{4.56}$$

The bands have vanishing width at $b = 0$ and widen with increasing b. In addition, the problem can be analysed even with the inclusion of a damping term in the equation of motion using the method of multiple time scales.

In our case, Eq. (4.34), the frequency varies periodically like the square of an elliptic cosine. Recall from Eq. (4.32) that our elliptic cosine is dominated by its first harmonic. Thus we might expect that our Lamé equation, (4.34), posseses resonant band structure similar to the Mathieu case described above. Approximating the elliptic cosine by its first harmonic, we can write

$$\kappa^2 + \frac{g^2}{\lambda}\mathrm{cn}^2\left(x, \frac{1}{\sqrt{2}}\right) \simeq \kappa^2 + \frac{g^2}{2\lambda} + \frac{g^2}{2\lambda}\cos\left(\frac{4\pi x}{T}\right), \tag{4.57}$$

where here T is the period of the elliptic cosine, Eq. (4.33). Thus, comparing this expression with the standard form of the Mathieu equation, Eq. (4.54), we should expect that, according to Eq. (4.56), resonance will occur in Eq. (4.34) near

$$\kappa^2 + \frac{g^2}{2\lambda} \simeq \left(\frac{2\pi n}{T}\right)^2 \simeq 0.7178n^2, \qquad n = 1, 2, \ldots. \tag{4.58}$$

For $g^2/\lambda \ll 1$, we expect narrow resonace bands near

$$\kappa^2 \simeq 0.7178n^2, \qquad n = 1, 2, \ldots, \tag{4.59}$$

and these bands will extend along roughly the directions

$$\kappa^2 + \frac{g^2}{2\lambda} = \text{const}, \tag{4.60}$$

getting wider for increasing g^2/λ. The bands will reach the axis $\kappa^2 = 0$ at very roughly the points

$$\frac{g^2}{\lambda} \simeq 1.4n^2, \qquad n = 1, 2, \ldots. \tag{4.61}$$

Further analytical progress is difficult for the Lamé equation, although some results are known [44]. To accurately illustrate the complete resonance structure in the κ^2-g^2/λ plane, the Lamé equation must be numerically integrated. In Fig. 4.3 I present the results of a numerical simulation of the $\delta\tilde{\chi}$ equation of motion, Eq. (4.34). For a grid of κ and g^2/λ parameter values I integrated this equation with some initial value $\delta\tilde{\chi}(0)$, and calculated the Floquet index from

$$\mu = \frac{1}{x}\ln\left(\frac{\delta\tilde{\chi}(x)}{\delta\tilde{\chi}(0)}\right). \tag{4.62}$$

For sufficiently long integration times x, the exponential growth dominates the oscillatory behaviour of $\delta\tilde{\chi}$, and so this expression converges to the true Floquet index.

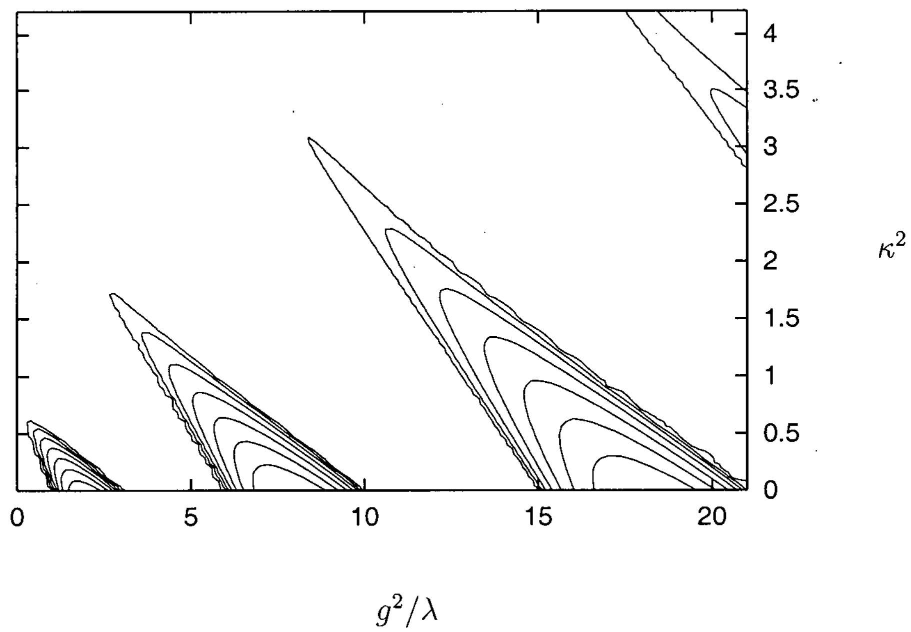

Figure 4.3: Floquet index μ for perturbation $\delta\tilde{\chi}$ calculated by numerically evolving Eq. (4.34). Parameter κ is a rescaled wavenumber. The seven contour levels are equally spaced from $\mu = 0.03$ to 0.21.

As expected, we see resonance band structure, in rough agreement with the predictions based on Eq. (4.58). The strongest resonance is for $k = 0$, where we find a sequence of $\delta\tilde{\chi}_k$ resonance bands, centred at $g^2/\lambda = 2n^2$ with width $2n$, for positive integral n, in agreement with [44]. In particular, $\delta\tilde{\chi}_k$ exhibits strong resonance at $k = 0$ for the supersymmetric point $g^2/\lambda = 2$. Resonance bands are weak at small scales. Thus $\delta\tilde{\varphi}_k$, which, according to Eq. (4.35) should evolve like $\delta\tilde{\chi}_k$ for $g^2/\lambda = 3$, exhibits only weak small-scale resonance. The Floquet index reaches a maximum value of $\mu_{\max} \simeq 0.238$ at the centre of each $k = 0$ band.

4.2.3 Numerical results

For my numerical calculations, I was primarily interested in the behaviour of cosmological-scale matter and metric modes. Thus I evolved a scale which left the Hubble radius (at time t_0) at about $N = 50$ e-folds before the end of inflation. For $(\lambda/4)\varphi^4$ models, the number of e-folds during slow-roll inflation after initial time t_0 is [21]

$$N \simeq \pi \left(\frac{\varphi(t_0)}{m_{\rm P}} \right)^2 ; \tag{4.63}$$

thus I used the homogeneous inflaton initial value of $\varphi(t_0) = 4m_{\rm P}$. I began the calculations with the modes still somewhat inside the Hubble radius, so the initial conditions for the matter field fluctuations were simply given by the adiabatic vacuum state

$$\delta\varphi_{Ak}(t_0) = \frac{1}{a^{3/2}(t_0)} \left(\frac{1}{2\omega_A(t_0)} \right)^{1/2}, \tag{4.64}$$

$$\delta\dot{\varphi}_{Ak}(t_0) = -i\omega_A(t_0)\delta\varphi_{Ak}(t_0), \tag{4.65}$$

with $\omega_\varphi^2(t) = (k/a)^2 + 3\lambda\varphi^2 + g^2\chi^2$ and $\omega_\chi^2(t) = (k/a)^2 + g^2\varphi^2$. Physically, the $a^{-3/2}$ dependence arises because particle number densities $n_k \propto |\delta\varphi_{Ak}|^2$ must decay like a^{-3} in the massive, adiabatic regime. The initial metric perturbations were then determined by Eq. (4.19).

To illustrate the dynamics in the absence of backreaction, I numerically integrated the coupled set of background equations (4.7) and (4.10) and perturbation equations (4.17) and (4.18) using the initial conditions described above, and for $g^2/\lambda = 2$, $\lambda = 10^{-14}$, and a zero χ background. (According to Fig. 4.3 these parameters are expected to give strong resonance for $\delta\chi_k$ as $k \to 0$.) The set of coupled ordinary differential equations was integrated with the NIST Core Math Library routine ddriv1 called from Fortran. I used the constraint Eq. (4.19) as well as the conservation equation (4.21) to check the accuracy of the calculations. In Fig. 4.4 I display the evolution of my cosmological modes, together with the comoving curvature perturbation ζ_k, during inflation and preheating. For each of the perturbations $X_k = \delta\chi_k$, $\delta\varphi_k$, Φ_k, and ζ_k I plot the power spectrum [51]

$$\mathcal{P}_X(k) = \frac{k^3}{2\pi^2}|X_k|^2, \tag{4.66}$$

rather than the mode amplitudes, to facilitate comparison with the COBE measured normalization which gives $\mathcal{P}_\Phi \sim 10^{-10}$ [46].

Figure 4.4 shows how the modes begin early in inflation as sub-Hubble oscillations, and become "frozen in" after they exit the Hubble radius. Note that the $\delta\chi_k$ fluctuation experiences some damping late in inflation, when its effective mass squared $g^2\varphi^2$ becomes somewhat greater than H^2, which

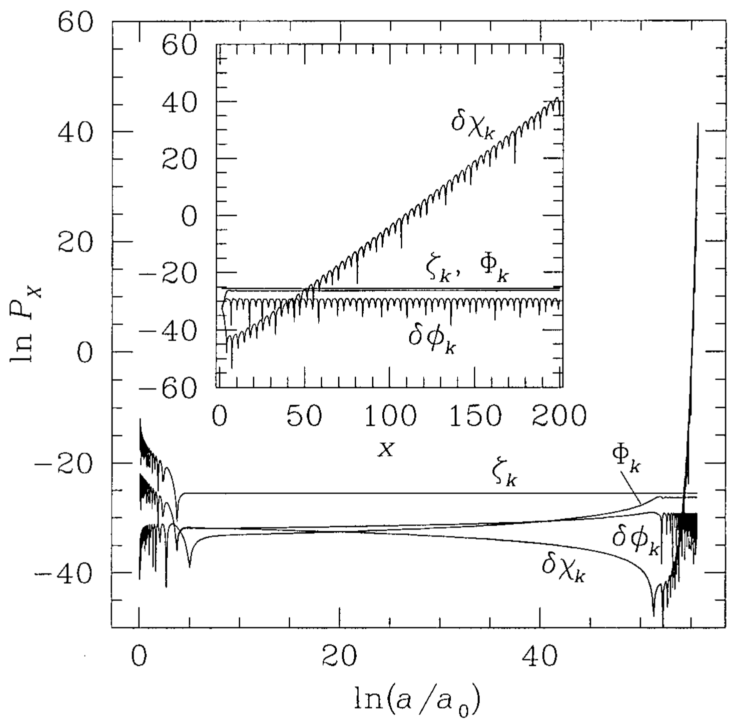

Figure 4.4: Numerical simulation of linear cosmological-scale perturbations in the two field model described in the text, in the absence of backreaction. Plotted are the logarithms of the power spectra $\mathcal{P}_X(k) = (k^3/2\pi^2)|X_k|^2$, for $X_k = \delta\chi_k$, $\delta\varphi_k$, Φ_k, and ζ_k, and using $m_\mathrm{P} = 1$, $g^2/\lambda = 2$, $\lambda = 10^{-14}$, and zero χ background. The main figure shows the evolution during inflation, while the inset details the behaviour during preheating, using the rescaled conformal time x. This particular comoving scale leaves the Hubble radius approximately 5 e-folds after the start of the simulation, which corresponds to $k/(aH) \sim 10^{-19}$ at the start of preheating.

decreases like φ^4 in the slow-roll approximation. The inflaton perturbation $\delta\varphi_k$, however, stays roughly constant during inflation even though its effective mass is comparable to that of the $\delta\chi_k$ mode. This is because of the coupling between $\delta\varphi_k$ and Φ_k in the linearized perturbation equations. We can also observe a growth of Φ_k between the time the mode exits the Hubble radius and the beginning of preheating, by a factor of approximately 20, in good agreement with the growth predicted from the "conservation law" Eq. (4.21). Also note that after this small growth stage the cosmological-scale metric power spectrum ends up close to the $10^{-10}(\sim e^{-23})$ level, as the standard theory predicts for $\lambda \simeq 10^{-14}$ in the absence of parametric resonance [25]. During preheating we observe exponential growth of $\delta\chi_k$ while the super-Hubble $\delta\varphi_k$ mode does not grow, as expected from the analytical theory. Φ_k and ζ_k remain constant, since according to Eq. (4.17) the metric perturbations couple only to $\delta\varphi_k$ in the absence of a χ background, at linear level [42].

To observe the effect of including a non-zero homogeneous χ background, I repeated the above calculation with an initial value of $\chi(t_0) = 10^{-10}m_{\rm P}$ (this value illustrates well the various stages of evolution). Figure 4.5 indicates that $\delta\chi_k$ grows as before, but $\delta\varphi_k$ and Φ_k now grow initially with twice the Floquet index of $\delta\chi_k$. This is the result of the driving term $2g^2\varphi\chi\delta\chi_k$ in the equation of motion for $\delta\varphi_k$, Eq. (4.18), which contains two factors growing like $e^{\mu_{\rm max}x}$ (the background χ satisfies the $\delta\chi_k$ equation of motion for $k \to 0$). Once the background χ field becomes comparable to the inflaton background, all the perturbations synchronize and grow at the same rate. (Note that although it is not plotted, the comoving curvature perturbation ζ_k behaves in essentially the same way as Φ_k in this and all subsequent figures.) It is important to emphasize that, ignoring the resonant growth during preheating, my model reproduces the standard predictions for $(\lambda/4)\varphi^4$ inflation [this statement holds even for initial homogeneous values as large as $\chi(t_0) \sim m_{\rm P}$]. Thus parametric resonance is an important new effect, which must be examined to determine if the model does in fact make realistic predictions. I will discuss the significance of the homogeneous χ field in relation to the nonlinear evolution of the fields in the next section.

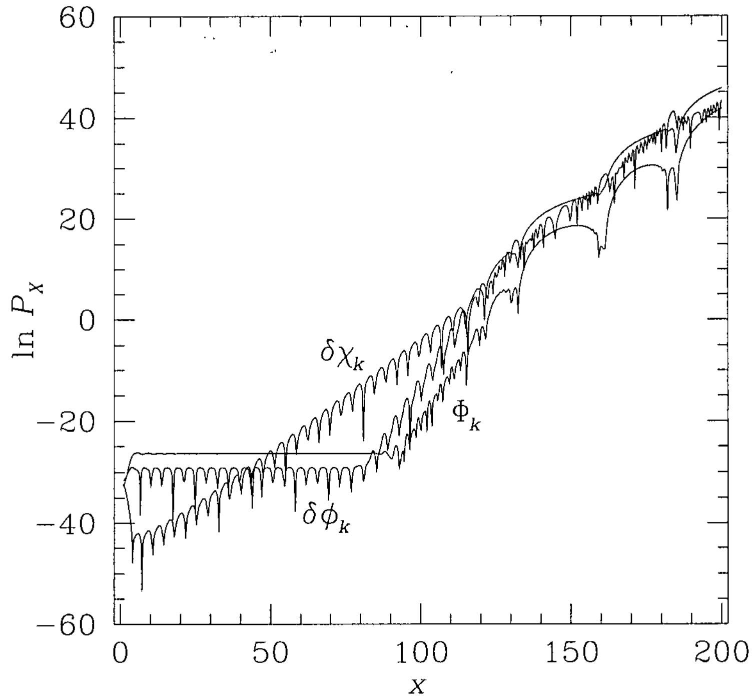

Figure 4.5: Simulation of linear cosmological-scale perturbations in the absence of backreaction as in Fig. 4.4, but with a non-zero initial background $\chi(t_0) = 10^{-10} m_{\rm P}$. Plotted are the logarithms of the power spectra $\mathcal{P}_X(k) = (k^3/2\pi^2)|X_k|^2$, for $X_k = \delta\chi_k$, $\delta\varphi_k$, and Φ_k, and using $m_{\rm P} = 1$, $g^2/\lambda = 2$, $\lambda = 10^{-14}$, and $k/(aH) \sim 10^{-19}$ at the start of preheating. Coupling through the χ background causes $\delta\varphi_k$ and Φ_k to grow initially at twice the rate of $\delta\chi_k$.

4.3 Backreaction[1]

4.3.1 Equations of motion

The linearized equations in the previous section describe unbounded growth of perturbations during resonance. In reality this growth must of course stop at some point, namely when the perturbed field values are on the order of the background values. A full nonlinear simulation will include this effect automatically, but approximation methods can alleviate the computational costs significantly. A common approach to approximate this backreaction on the background and perturbation evolution is to include Hartree terms in the equations of motion [6, 31]. This entails making the replacements $\varphi_A \rightarrow \varphi_A + \delta\varphi_A$, $\delta\varphi_A^2 \rightarrow \langle\delta\varphi_A^2\rangle$, and $\delta\varphi_A^3 \rightarrow 3\langle\delta\varphi_A^2\rangle\delta\varphi_A$. In this approximation, the background equations (4.7) – (4.10) become

$$\begin{aligned}H^2 = \frac{8\pi}{3m_{\mathrm{P}}^2}\Bigg[&V(\varphi,\chi) + \frac{3}{2}\lambda\varphi^2\langle\delta\varphi^2\rangle + \frac{g^2}{2}\varphi^2\langle\delta\chi^2\rangle + \frac{g^2}{2}\chi^2\langle\delta\varphi^2\rangle \\ &+ \frac{1}{2}\sum_A\left(\dot{\varphi}_A^2 + \langle\delta\dot{\varphi}_A^2\rangle + \frac{1}{a^2}\langle\nabla\delta\varphi_A^2\rangle\right)\Bigg],\end{aligned} \tag{4.67}$$

$$\ddot{\varphi} + 3H\dot{\varphi} + V_{,\varphi} + 3\lambda\langle\delta\varphi^2\rangle\varphi + g^2\langle\delta\chi^2\rangle\varphi = 0, \tag{4.68}$$

$$\ddot{\chi} + 3H\dot{\chi} + V_{,\chi} + g^2\langle\delta\varphi^2\rangle\chi = 0. \tag{4.69}$$

Similarly, the momentum-space linearized field perturbation equations (4.18) become

$$\begin{aligned}\delta\ddot{\varphi}_k + 3H\delta\dot{\varphi}_k + \left(\frac{k^2}{a^2} + 3\lambda\langle\delta\varphi^2\rangle + g^2\langle\delta\chi^2\rangle\right)\delta\varphi_k \\ + V_{,\varphi\varphi}\cdot\delta\varphi_k = 4\dot{\varphi}\dot{\Phi}_k - 2V_{,\varphi}\Phi_k,\end{aligned} \tag{4.70}$$

$$\begin{aligned}\delta\ddot{\chi}_k + 3H\delta\dot{\chi}_k + \left(\frac{k^2}{a^2} + g^2\langle\delta\varphi^2\rangle\right)\delta\chi_k \\ + V_{,\chi\varphi}\cdot\delta\varphi_k = 4\dot{\chi}\dot{\Phi}_k - 2V_{,\chi}\Phi_k.\end{aligned} \tag{4.71}$$

In this approach, the field fluctuations are calculated self-consistently from the relations

$$\langle\delta\varphi_A^2\rangle = \frac{1}{(2\pi)^3}\int d^3k|\delta\varphi_{Ak}|^2. \tag{4.72}$$

[1] A version of Sections 4.3, 4.4, and 4.5 has been published. Zibin, J. P., Brandenberger, R., and Scott, D., (2001) Backreaction and the Parametric Resonance of Cosmological Fluctuations, Phys. Rev. D **63**: 043511.

In practice, the strongest resonance band will provide a natural ultraviolet cutoff.

The Hartree terms approximate the full nonlinear dynamics of the fields. To illustrate what this approximation entails, we may consider the exact dynamics of the fields, treating the Klein-Gordon equation as a classical field equation. As an example, consider the exact evolution equation for $\delta\varphi$ in position space, obtained by perturbing the Klein-Gordon equation, setting the background χ to zero, and ignoring metric perturbations,

$$\begin{aligned}\delta\ddot{\varphi} + 3H\delta\dot{\varphi} + \frac{1}{a^2}\nabla^2\delta\varphi + 3\lambda\varphi^2\delta\varphi + 3\lambda\varphi\delta\varphi^2 + \lambda\delta\varphi^3 \\ + g^2\delta\chi^2\varphi + g^2\delta\chi^2\delta\varphi = 0.\end{aligned} \tag{4.73}$$

The terms in this equation describing the interaction between the φ and χ fields become in momentum space

$$\frac{g^2\varphi}{(2\pi)^{3/2}}\int d^3k'\delta\chi_{\mathbf{k}'}\delta\chi_{\mathbf{k}-\mathbf{k}'} + \frac{g^2}{(2\pi)^3}\int d^3k'd^3k''\delta\chi_{\mathbf{k}'}\delta\chi_{\mathbf{k}''}\delta\varphi_{\mathbf{k}-\mathbf{k}'-\mathbf{k}''}. \tag{4.74}$$

Thus the Hartree term $g^2\langle\delta\chi^2\rangle\delta\varphi_k$ in Eq. (4.70) corresponds to the second term in expression (4.74), restricted to $\mathbf{k}'' = -\mathbf{k}'$. Physically, this means that only scattering events which do not change the $\delta\varphi_k$ momentum are included in the Hartree approximation, and "rescattering" events are ignored.

It is important to notice that the first term in (4.74), which scatters particles from the homogeneous inflaton background into mode $\delta\varphi_k$ and rescatters $\delta\chi$ particles, could be larger than the Hartree term since initially $|\varphi| > |\delta\varphi|$, unless the first term vanishes upon averaging (integrating) over the entire phase space of contributing terms (which is what is assumed in the Hartree approximation). If it does not vanish, the first term in (4.74) will act as a driving term for the $\delta\varphi$ modes in (4.73). Since some $\delta\chi$ modes experience parametric amplification with Floquet exponent μ, this term will lead to an important second-order effect, namely the growth of $\delta\varphi$ as $e^{2\mu x}$. This effect is left out in the Hartree approximation. Because the metric perturbations are coupled to $\delta\varphi$ through Eq. (4.17), we also expect that, with the homogeneous χ set to zero, the Hartree approximation will miss the corresponding growth of Φ. Note, however, that by including the χ background term $2g^2\chi\varphi\delta\chi$ in (4.73), and setting $\chi^2 \sim \langle\delta\chi^2\rangle$, we can approximate the effect of the important first term in (4.74), as we saw in Fig. 4.5.

Note that the calculations to follow, based on the Hartree equations of motion listed above, are *not* intended to accurately describe the evolution of the fields after the nonlinear terms become important. Instead, I use the Hartree approximation of the nonlinear terms as an indication of when those terms become important, and hence the growth of fluctuations should stop.

4.3.2 Analytical estimates

Evolution of perturbations during inflation

Perturbations will grow during parametric resonance until backreaction becomes important. We can analytically estimate the amount of growth by estimating the time at which the Hartree term $g^2\langle\delta\chi^2\rangle$ is of the order of the background $\lambda\varphi^2$ [*cf.* Eq. (4.70)]. Note that in the absence of metric fluctuations, such an estimate should be accurate, at least for $g^2/\lambda \sim 1$, as nonlinear lattice simulations indicate [8, 52]. In order to estimate the variance $\langle\delta\chi^2\rangle$, we will need to calculate the evolution of $\delta\chi_k$ modes, starting from the adiabatic vacuum inside the Hubble radius, continuing through inflation, and finally through preheating. The evolution during inflation is quite complicated, and will have a crucial effect on the final variances, so I will describe the inflationary stage in some detail. I consider general values of g^2/λ, rather than just the supersymmetric point.

I will only need to consider the contribution to $\langle\delta\chi^2\rangle$ from modes which are super-Hubble at the start of preheating. To see this, first note that for $g^2/\lambda = 2$, the small-scale boundary of the strongest (and largest-scale) resonance band is at $k_{\max}/a \simeq \sqrt{\lambda}\varphi_0/2$, where $\varphi_0(t)$ is the amplitude of inflaton oscillations during preheating [44]. Next, we can use the Friedmann equation (4.10) to write the Hubble parameter in terms of φ_0, giving

$$H^2 = \frac{2\pi}{3m_{\mathrm{P}}^2}\lambda\varphi_0^4. \tag{4.75}$$

(Note that this equation also applies approximately during slow-roll.) Using the value $\varphi_0 = 0.2m_{\mathrm{P}}$, I calculate the ratio $aH/k_{\max} \simeq 0.6$ at the start of preheating. Thus the Hubble radius corresponds closely to the smallest resonant scale. This result is not very sensitive to g^2/λ as long as we are near the centre of a band, *i.e.* $g^2/\lambda = 2n^2$, since $k_{\max}$ increases only slowly with g^2/λ in this case [44]. Also, we can ignore the resonance bands at higher k values, since they correspond to narrow resonance.

To estimate the evolution of $\delta\chi_k$ on super-Hubble scales during inflation, we can ignore terms containing the background χ as well as the spatial gradient term in Eq. (4.18), resulting in a damped harmonic oscillator equation with time-dependent coefficients,

$$\ddot{\delta\chi}_k + 3H\dot{\delta\chi}_k + g^2\varphi^2\delta\chi_k = 0. \tag{4.76}$$

During slow-roll, we can use the adiabatic approximation to find solutions to this equation, since $|\dot{H}| \ll H^2$. Thus for $g^2\varphi^2 > (3H/2)^2$ we have underdamped oscillations with damping envelope

$$\delta\chi_k \propto \exp\left[-\int (3H/2)dt\right] = a^{-3/2}. \tag{4.77}$$

For $g^2\varphi^2 < (3H/2)^2$, we have the overdamped case with two decaying modes. Ignoring the more rapidly decaying mode, I obtain

$$\delta\chi_k \propto \exp\left[-\int \left(3H/2 - \sqrt{9H^2/4 - g^2\varphi^2}\right) dt\right]. \tag{4.78}$$

In this case, the fluctuations are very slowly decaying in the massless limit $g^2\varphi^2 \ll (3H/2)^2$, while they approach the $a^{-3/2}$ decay as $g^2\varphi^2 \to (3H/2)^2$.

During slow-roll we have $H^2 \propto \varphi^4$ [see Eq. (4.75)], so that H^2 decreases more rapidly than $g^2\varphi^2$, and there is a transition between the over- and underdamped stages. The two types of behaviour are separated by the critically damped case, $g^2\varphi^2 = (3H/2)^2$. Using Eqs. (4.75) and (4.63), we can write this critical damping condition in terms of the number of e-folds after critical damping, $N_{\rm crit}$, as

$$N_{\rm crit} = \ln\left(\frac{a_f}{a_{\rm crit}}\right) = \frac{2}{3}\frac{g^2}{\lambda} \simeq \ln\left(\frac{k_f}{k_{\rm crit}}\right), \tag{4.79}$$

where subscript "f" refers to the end of inflation and "crit" to the time of critical damping. Wavevectors $k_{\rm crit}$ and k_f leave the Hubble radius at $t_{\rm crit}$ and t_f, respectively. We see that as g^2/λ increases, cosmological scales are damped like $a^{-3/2}$ during a greater and greater part of inflation. We thus expect that for large enough g^2/λ, the backreaction of the smaller-scale modes will terminate parametric resonance when cosmological-scale $\delta\chi_k$ modes are still greatly suppressed. In other words, there will be a maximum value of g^2/λ for which there is significant amplification of super-Hubble $\delta\chi_k$ perturbations, as anticipated in [43].

I first consider the evolution of the modes which leave the Hubble radius after $t_{\rm crit}$, *i.e.* $k > k_{\rm crit}$ (but which are still super-Hubble at the end of inflation, $k < k_f$). These modes are effectively massive during inflation, and hence we can simply use the adiabatic vacuum state, Eq. (4.64), which for $k \ll aH$ gives

$$|\delta\chi_k(t_f)|^2 = \frac{1}{2a_f^3 g\varphi_f}. \tag{4.80}$$

Note that if we define the spectral index n through $\mathcal{P}_\chi(k) \propto k^{n-1}$ [51], then for this part of the spectrum we have $n = 4$, an extreme blue tilt.

Next, I will calculate the evolution of modes which leave the Hubble radius before $t_{\rm crit}$, *i.e.* modes with $k < k_{\rm crit}$. In this case, the modes are approximately massless when they exit the Hubble radius [$g^2\varphi^2 < (3H/2)^2$ for $t < t_{\rm crit}$], so we can use the standard result for a massless inflaton [20],

$$|\delta\chi_k(t_k)|^2 = \frac{H^2(t_k)}{2k^3}, \tag{4.81}$$

where t_k is the time that mode $\delta\chi_k$ exits the Hubble radius. We now must use Eq. (4.78) to evolve the modes during the overdamped period, $t_k < t < t_{\rm crit}$. Writing $dt = d\varphi/\dot\varphi$, and using the slow-roll approximation $\dot\varphi \simeq -V_{,\varphi}/3H$, we can perform the integral to obtain

$$|\delta\chi_k(t_{\rm crit})|^2 = \frac{H^2(t_k)}{2k^3} e^{-3F(N_k)}, \tag{4.82}$$

where N_k is the number of e-folds after time t_k and

$$\begin{aligned} F(N_k) \equiv N_k - N_{\rm crit} - \sqrt{N_k}\sqrt{N_k - N_{\rm crit}} \\ +N_{\rm crit} \ln\left(\frac{\sqrt{N_k} + \sqrt{N_k - N_{\rm crit}}}{\sqrt{N_{\rm crit}}}\right). \end{aligned} \tag{4.83}$$

Next we can readily propagate the modes through the underdamped period, $t_{\rm crit} < t < t_f$, using Eqs. (4.77) and (4.79), giving

$$|\delta\chi_k(t_f)|^2 = \frac{H^2(t_k)}{2k^3} e^{-3F(N_k) - 2g^2/\lambda}. \tag{4.84}$$

Since the damping term $F(N_k)$ is positive, we see as expected that large-scale modes are strongly damped for large g^2/λ.

Finally, we can approximate the conformal time dependence of all super-Hubble modes during parametric resonance as

$$\delta\chi_k \propto e^{\mu_{\rm max} x}, \tag{4.85}$$

if we are near the centre of a resonance band. This is valid since, in this case, the Floquet index μ_k varies only slightly for scales larger than a few times the Hubble radius (*i.e.* the smallest resonant scale) [44].

Variances and total resonant growth

Now I can proceed to calculate the field variance, $\langle\delta\chi^2\rangle$. I will use Eq. (4.72), restricting the integral to the resonantly growing modes. I begin with the case $g^2/\lambda = 2$. Equation (4.79) tells us that in this case $N_{\rm crit} \simeq 4/3$, so that essentially all of the evolution during inflation is in the overdamped regime, and we only need to consider modes with $k < k_{\rm crit}$. The variance integral will be dominated by modes with $N_k \gg N_{\rm crit}$, so we may approximate the damping term in Eq. (4.83) as

$$e^{-3F(N_k)} \simeq \left(\frac{N_{\rm crit}}{4N_k}\right)^{g^2/\lambda}. \tag{4.86}$$

For the current case, $g^2/\lambda = 2$, we can now combine the expression (4.84) with Eqs. (4.63), (4.75), (4.85), and (4.86) to obtain for the power spectrum on resonant scales at the end of preheating

$$\mathcal{P}_\chi(k, t_e) = \frac{\lambda m_{\rm P}^2}{54\pi^3} e^{2\mu_{\rm max} x_e}. \tag{4.87}$$

Here t_e is the time that the resonance shuts down, and x_e is the corresponding scaled conformal time. As we will see, the important thing about this result is that the power spectrum is essentially Harrison-Zel'dovich (independent of k), with spectral index $n = 1$.

We can next rewrite the variance integral, Eq. (4.72), in terms of the power spectrum as

$$\langle \delta\chi^2(t_e) \rangle = \int_0^{N_0} dN_k \mathcal{P}_\chi(k, t_e) = N_0 \mathcal{P}_\chi(t_e), \tag{4.88}$$

where $N_0 \simeq 50$ is the total number of e-folds during inflation. Finally, the criterion $g^2 \langle \delta\chi^2(t_e) \rangle \sim \lambda\varphi^2(t_e)$ gives, using the value $\varphi(t_e) \sim 10^{-2} m_{\rm P}$,

$$\mathcal{P}_\chi(t_e) \sim 10^{-6} m_{\rm P}^2 \tag{4.89}$$

for the $\delta\chi_k$ power spectrum on cosmological scales at the end of preheating. Note that this result used only the k-independence of the power spectrum (which is a result of the special choice $g^2/\lambda = 2$), and the values of N_0 and $\varphi(t_e)$. In particular, the result is independent of λ, unless, contrary to my implicit assumption, λ is so large that $g^2\langle\delta\chi^2\rangle > \lambda\varphi^2$ already at the start of preheating. In this case, Eq. (4.89) will be an underestimate.

According to the results from Section 4.2.3, we expect synchronization of the other fields to $\delta\chi_k$, so that in particular we expect $\mathcal{P}_\Phi \sim \mathcal{P}_\chi / m_{\rm P}^2$. Therefore I conclude that, for $g^2/\lambda = 2$, the metric perturbation amplitude will indeed be considerably larger than the COBE measured value, even including the effect of backreaction.

Next I will repeat the preceding analysis for the second super-Hubble resonance band, at $g^2/\lambda = 8$. In this case we have $N_{\rm crit} \simeq 5$, so we must consider modes that exit the Hubble radius both before and after $t_{\rm crit}$. For the large-scale modes, $k < k_{\rm crit}$, it will be sufficient to place an upper limit on the variance. Using Eq. (4.84), but ignoring the damping factor e^{-3F}, I obtain

$$\mathcal{P}_\chi(k, t_f) < \frac{H^2(t_{\rm crit})}{(2\pi)^2} e^{-2g^2/\lambda} \simeq 2 \times 10^{-8} \lambda m_{\rm P}^2 \tag{4.90}$$

on scales $k < k_{\rm crit}$ at the end of inflation. Thus, using Eq. (4.88), the contribution to the variance from modes with $k < k_{\rm crit}$ satisfies the (probably very conservative) bound

$$\langle \delta\chi^2(t_f) \rangle_{k<k_{\rm crit}} < 9 \times 10^{-7} \lambda m_{\rm P}^2. \tag{4.91}$$

Next we can use Eq. (4.80) to calculate the contribution to $\langle \delta\chi^2 \rangle$ from smaller-scale modes with $k_{\rm crit} < k < k_f$,

$$\langle \delta\chi^2(t_f) \rangle_{k_{\rm crit}<k<k_f} = \frac{1}{16\pi^3 a_f^3 g \varphi_f} \int_{k_{\rm crit}}^{k_f} d^3k \tag{4.92}$$

$$= \frac{1}{18}\sqrt{\frac{2}{3\pi}\frac{\lambda}{g^2}\frac{\lambda\varphi_f^5}{m_{\rm P}^3}} \tag{4.93}$$

$$\simeq 3 \times 10^{-6} \lambda m_{\rm P}^2. \tag{4.94}$$

Here I have used $k_{\rm crit}^3 \ll k_f^3$ (which follows from Eq. (4.79) for $g^2/\lambda = 8$), the relation $k_f/a_f = H(t_f)$, Eq. (4.75), and the value $\varphi_f = 0.2 m_{\rm P}$. This value of the small-scale variance exceeds our upper limit on the large-scale variance in Eq. (4.91), so we can ignore the contribution from the large-scale modes, $\langle \delta\chi^2(t_f) \rangle_{k<k_{\rm crit}}$.

Now we can again apply the condition $g^2 \langle \delta\chi^2(t_e) \rangle \sim \lambda\varphi^2(t_e)$, which in this case gives

$$e^{2\mu_{\rm max} x_e} \sim 4\lambda^{-1}. \tag{4.95}$$

Finally, I can use Eq. (4.84) without approximation to calculate the cosmological scale power spectrum at the end of preheating, for the case $g^2/\lambda = 8$,

$$\mathcal{P}_\chi(t_e) = \frac{H^2(t_0)}{(2\pi)^2} \exp\left[-3F(N_0) - 2\frac{g^2}{\lambda} + 2\mu_{\rm max} x_e\right] \tag{4.96}$$

$$\sim 10^{-14} m_{\rm P}^2. \tag{4.97}$$

In this case the growth stops before the cosmological perturbations exceed the COBE value, and thus parametric resonance does not change the standard predictions [25] for the size of the fluctuations. Therefore, since the damping of super-Hubble $\delta\chi_k$ modes increases as g^2/λ increases, the standard predictions are not modified for all resonance bands beyond the first, *i.e.* for $g^2/\lambda \geq 8$.

4.3.3 Numerical results

It is straightforward to check my analytical estimates from the previous section by numerically integrating the coupled set of Hartree approximation evolution equations (4.67) – (4.72) and metric perturbation equation (4.17). I now must evolve a set of modes that fill the relevant resonance band. For example, for $g^2/\lambda = 2$, the first resonance band extends from $\kappa = 0$ to $\kappa = 0.5$ [44]. Again I begin each mode's evolution inside the Hubble radius during inflation, using the initial vacuum state, Eqs. (4.64) and (4.65). Each mode is incorporated into the calculation shortly before it leaves the Hubble radius, so that the spatial gradient terms are never too large. The variances are calculated by performing the discretized integrals, Eqs. (4.72), only over the resonance band; thus they are convergent. Note that the variances are calculated *simultaneously* with the field backgrounds and perturbations.

In Fig. 4.6 I present the evolution of the $\delta\chi_k$, $\delta\varphi_k$, and Φ_k power spectra on the same cosmological scale as was studied in Section 4.2.3. All parameters are the same as for Fig. 4.5, except here I use for the initial background value $\chi(t_0) = 10^{-6} m_{\mathrm{P}}$, which means that during preheating $\chi^2 \simeq \langle \delta\chi^2 \rangle$. The evolution is initially similar to that of Fig. 4.5, only here the growth saturates at $\mathcal{P}_\chi \sim 3 \times 10^{-7} m_{\mathrm{P}}^2$, in good agreement with my prediction based on Eq. (4.89), and also in good agreement with the results of Tsujikawa *et al.* [45]. In addition, the other fields closely follow $\mathcal{P}_\chi$, as expected. Whereas in the linear calculations the Einstein constraint equation (4.19) was satisfied to extremely good accuracy, with the inclusion of backreaction $\mathcal{P}_\Phi$ saturates at a factor of roughly 10^3 higher using Eq. (4.19) than the illustrated result, which used Eq. (4.17). Note that a similar discrepancy was found in [39]. I suspect that this is a fundamental problem related to my attempt to capture some of the nonlinear dynamics with the Hartree approximation. Regardless of which value is used, the cosmological metric perturbations considerably exceed the COBE normalisation.

As discussed above, larger values of g^2/λ result in increased damping of $\delta\chi_k$ on large scales during inflation, and at large enough g^2/λ we expect insignificant amplification of super-Hubble modes. This is illustrated in Fig. 4.7. Here I examine the second resonance band at $g^2/\lambda = 8$, but use otherwise identical parameters to Fig. 4.6. Resonance stops at $\mathcal{P}_\chi \sim 10^{-14} m_{\mathrm{P}}^2$, consistent with my analytical estimate from Eq. (4.97), and not exceeding the standard predictions for $\lambda \simeq 10^{-14}$ [25]. Note that the small rise in $\mathcal{P}_\Phi$ at late times should not be trusted, as my Hartree approximation scheme will not capture the full nonlinear behaviour. For resonance bands at even higher g^2/λ, I find extremely suppressed cosmological $\delta\chi_k$ amplitudes, in quantitative agreement with the calculations of the previous section.

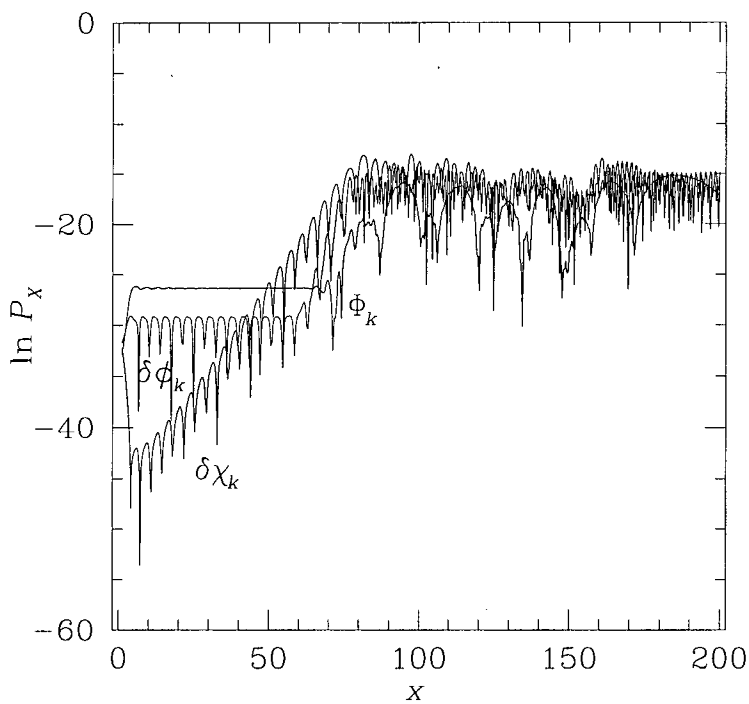

Figure 4.6: Numerical simulation of cosmological-scale perturbations with Hartree backreaction terms included and with a non-zero initial background, $\chi(t_0) = 10^{-6} m_{\rm P}$. Plotted are the logarithms of the power spectra $\mathcal{P}_X(k) = (k^3/2\pi^2)|X_k|^2$, for $X_k = \delta\chi_k$, $\delta\varphi_k$, and Φ_k, using $m_{\rm P} = 1$, $g^2/\lambda = 2$, $\lambda = 10^{-14}$, and $k/(aH) \sim 10^{-19}$ at the start of preheating. Backreaction terminates the growth of each field perturbation.

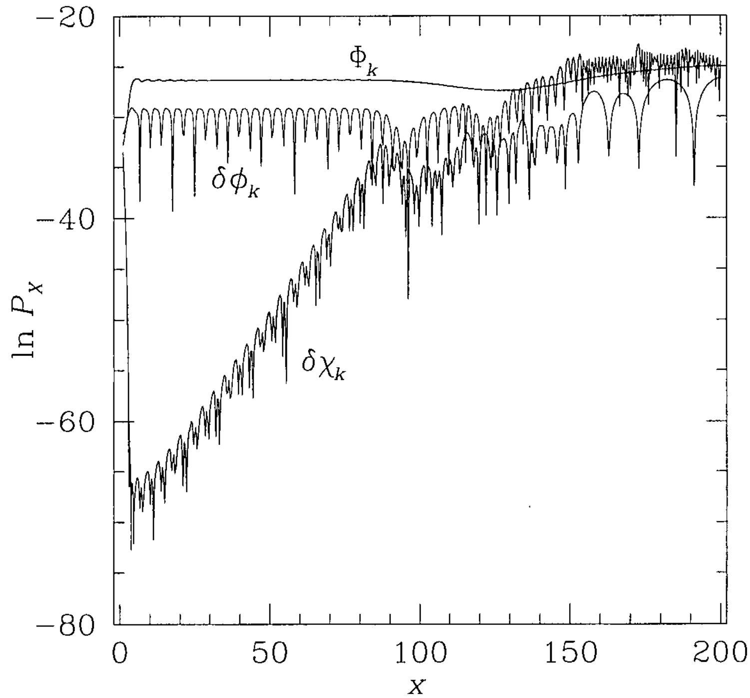

Figure 4.7: Same as Fig. 4.6, except for parameters lying in the second resonance band, *i.e.* $g^2/\lambda = 8$, and with $\chi(t_0) = 10^{-6} m_\mathrm{P}$ and $\lambda = 10^{-14}$. Here backreaction of the small-scale modes terminates the growth before the COBE value is exceeded.

4.4 Self-interacting χ models

4.4.1 Positive coupling

I now consider the addition of a quartic self-interaction term for the χ field, so that our potential becomes

$$V(\varphi, \chi) = \frac{\lambda}{4}\varphi^4 + \frac{g^2}{2}\varphi^2\chi^2 + \frac{\lambda_\chi}{4}\chi^4, \tag{4.98}$$

with $g^2 > 0$. The significance of such a term for parametric resonance was studied in lattice simulations [52] and analytically [54], but in the absence of metric perturbations. Bassett and Viniegra [43] included metric perturbations, but ignored backreaction. Essentially, for $\lambda_\chi \gtrsim \lambda$ we expect the χ self-interaction to limit the growth of perturbations as compared with the $\lambda_\chi = 0$ case studied above, due to the presence of the "potential wall" $(\lambda_\chi/4)\chi^4$.

More precisely, the linearized equation of motion for the χ field perturbation becomes, with χ self-interaction but ignoring metric perturbations,

$$\delta\ddot{\chi}_k + 3H\delta\dot{\chi}_k + \left(\frac{k^2}{a^2} + 3\lambda_\chi\chi^2 + g^2\varphi^2\right)\delta\chi_k + 2g^2\varphi\chi\delta\varphi_k = 0. \tag{4.99}$$

Thus for small enough initial χ background, the initial behaviour of the modes will be essentially unchanged from the $\lambda_\chi = 0$ case. However, when the χ background grows to the point that $\chi^2/\varphi^2 \gtrsim g^2/\lambda_\chi$, the analytical parametric resonance theory of Section 4.2.1 no longer applies, and we may expect the perturbations to stop growing. Since, as discussed above, for the significant production of super-Hubble modes we require $g^2 \simeq \lambda$, we expect that χ self-interaction will shut down the resonance when $\chi^2/\varphi^2 \sim \lambda/\lambda_\chi$, as long as $\lambda_\chi \gtrsim \lambda$. If $\lambda_\chi < \lambda$, then the χ^4 interaction term will not lead to a shutdown of the resonance since (based on my numerical simulations) the homogeneous χ field never substantially exceeds the value of the inflaton background.

I have confirmed this expectation numerically, and I give an example of my results in Fig. 4.8. Here I have included Hartree backreaction and metric perturbations, and used coupling constant values $\lambda = 10^{-14}$, $g^2/\lambda = 2$, and $\lambda_\chi = 10^{-10}$, and initial backgrounds $\varphi(t_0) = 4m_{\mathrm{P}}$ and $\chi(t_0) = 10^{-6}m_{\mathrm{P}}$. We indeed observe the termination of the super-Hubble modes' growth at approximately the time when $\chi^2/\varphi^2 = \lambda/\lambda_\chi$.

Note that, *in the absence of backreaction,* Bassett and Viniegra observed a continued slow growth of super-Hubble perturbations after the initial termination of the resonance when $\chi^2/\varphi^2 \sim \lambda/\lambda_\chi$ [43]. I confirmed this result; however, note that when we include the backreaction term $3\lambda_\chi\langle\delta\chi^2\rangle$ in the evolution equations, we expect backreaction to become important also at the time

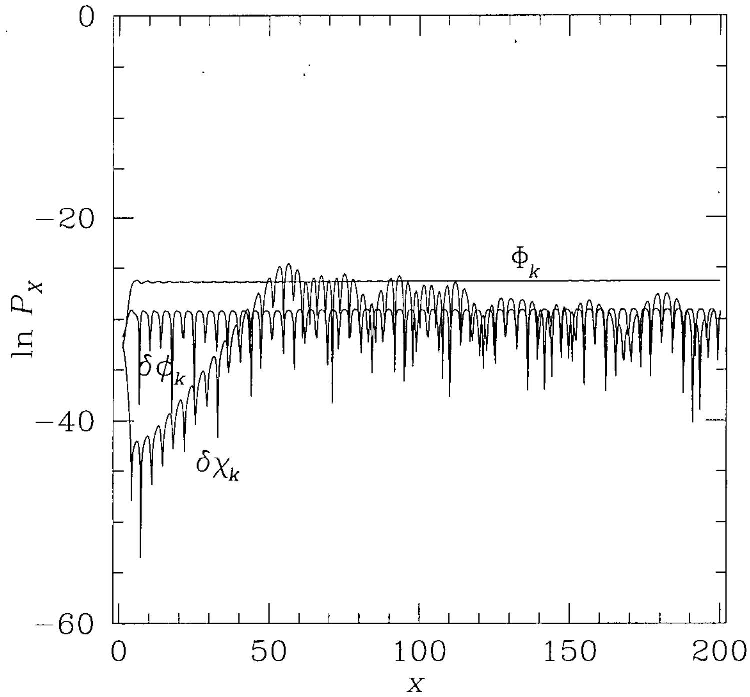

Figure 4.8: Same as for Fig. 4.6, except with χ field self-coupling $\lambda_\chi = 10^{-10}$. The χ self-coupling causes the early termination of growth.

that $\chi^2/\varphi^2 \sim \lambda/\lambda_\chi$, with my choice $\chi^2 \simeq \langle\delta\chi^2\rangle$. Hence, as seen in Fig. 4.8, the slow growth is completely suppressed.

4.4.2 Negative coupling

The presence of χ self-coupling means that we no longer require $g^2 \geq 0$ for global stability. In fact, for the case $g^2 < 0$, the potential will be bounded from below for $\lambda\lambda_\chi/g^4 > 1$ [55]. This *negative coupling* case was studied in the absence of metric perturbations using lattice simulations in [55], and without backreaction in [42]. The behaviour of the fields is qualitatively different in the negative and positive coupling cases. For $g^2 < 0$, potential minima exist with non-zero homogeneous part of the χ field. Thus, assuming the fields fall into these minima, the problem of choice of χ background discussed in previous sections for the positive coupling case is alleviated.

For initial homogeneous χ fields large enough [$\chi(t_0) \gtrsim m_{\mathrm{P}}$], I find numerically that the fields fall into the potential minimum by the end of inflation, and the two fields subsequently evolve in step during preheating. This effectively reduces the system to a single-field system, and hence no resonance is possible on super-Hubble scales.

To see this explicitly, consider the case $\lambda_\chi = \lambda$, for which the symmetry of the potential requires the potential minima to lie along $\chi^2 = \varphi^2$. If we choose the same initial signs for χ and φ, then during preheating the backgrounds lie along the attractor $\varphi = \chi$. Similarly, since the behaviour of super-Hubble modes is essentially the same as that of the backgrounds, we have $\delta\chi = \delta\varphi$ during preheating. Then the perturbation equation (4.99) becomes

$$\delta\ddot{\chi}_k + 3H\delta\dot{\chi}_k + \left[\frac{k^2}{a^2} + 3(\lambda + g^2)\varphi^2\right]\delta\chi_k = 0. \tag{4.100}$$

Thus the effective mass of the $\delta\chi$ oscillations is precisely three times the effective mass of the background inflaton oscillations [*cf.* Eq. (4.7)], so that just as with the case of the inflaton perturbations in Eq. (4.35), there will be no resonance on super-Hubble scales for all allowed values of g^2. I have confirmed this numerically; indeed more generally, as long as initially $\chi(t_0) \sim m_{\rm P}$ but for any $\lambda_\chi \geq \lambda$, the two fields will be proportional during preheating and no super-Hubble resonance will result.

This result assumes that during preheating only the "field" $\delta\chi + \delta\varphi$ is excited. If orthogonal field excitations $\delta\chi - \delta\varphi$ are present, they can grow resonantly. The effective squared mass of $\delta\chi - \delta\varphi$ excitations is $(3\lambda - g^2)\varphi^2$, so that according to the analytical parametric resonance theory of Section 4.2.1, super-Hubble resonance will occur near $3\lambda - g^2 = 2n^2(\lambda + g^2)$, for integral n (we require $n \geq 2$ for negative g^2). That is, super-Hubble $\delta\chi - \delta\varphi$ modes will grow for $g^2 \simeq \lambda(3 - 2n^2)/(1 + 2n^2)$. However, numerically I observe only extremely small components $\delta\chi - \delta\varphi$ by the end of inflation, so their growth is substantially delayed.

On the other hand, for small initial homogeneous part $\chi(t_0) \ll m_{\rm P}$, I find that the potential minima are not reached by the end of inflation, and the two fields evolve in a very complicated manner during preheating. The analytical theory of parametric resonance cannot be applied, but numerically I do find roughly exponential growth of super-Hubble modes in this case, as found in [42]. The growth rate increases as g^2 decreases towards the value at which global instability sets in, $g^2 = -\sqrt{\lambda\lambda_\chi}$.

I have illustrated this case in Fig. 4.9, using the parameter values $\lambda = 10^{-14}$, $g^2 = -0.5\lambda$, $\lambda_\chi = \lambda$, $\varphi(t_0) = 4m_{\rm P}$, and $\chi(t_0) = 10^{-6}m_{\rm P}$. Here the growth rates and final power spectra values are comparable to the $\lambda_\chi = 0$ case of Fig. 4.6, though the $\delta\chi$ field is not damped during inflation for negative coupling. For $\lambda_\chi > \lambda$, the growth is terminated early, just as in the positive coupling case.

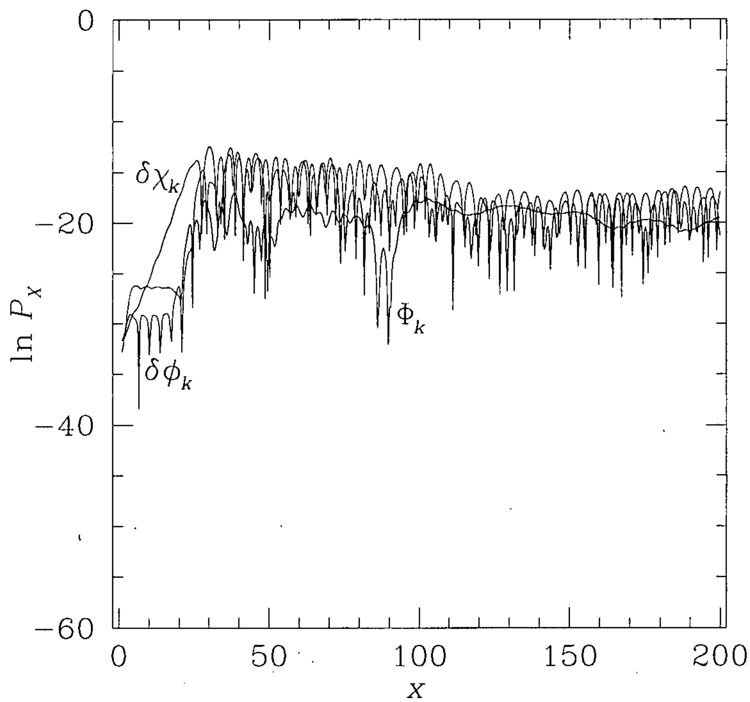

Figure 4.9: Same as for Fig. 4.8, except for the negative coupling case, with parameters $\lambda = 10^{-14}$, $g^2 = -0.5\lambda$, and $\lambda_\chi = \lambda$. Here the growth is comparable to the $\lambda_\chi = 0$ case of Fig. 4.6, even though analytical parametric resonance theory cannot be applied.

4.5 Summary and discussion

In this chapter I have studied backreaction effects on the growth of super-Hubble cosmological fluctuations in a specific class of two field models with a massless inflaton φ coupled to a scalar field χ. My study was based on the Hartree approximation. Ignoring the resonant growth during preheating, my model can reproduce the standard predictions of $(\lambda/4)\varphi^4$ inflation, but preheating changes the picture dramatically.

For the non-self-coupled χ field case, I found that backreaction has a crucial effect in determining the final amplitude of fluctuations after preheating. For values of the coupling constants satisfying $g^2/\lambda = 2$ (the ratio predicted in supersymmetric models), the predicted amplitude of the super-Hubble metric perturbations at the end of preheating is too large to be consistent with the COBE normalization, thus apparently ruling out such models. In addition, the final amplitude of the fluctuation spectrum is independent of the coupling constant λ. Note that the growth of inflaton fluctuations $\delta\varphi_k$ (and hence metric perturbations Φ_k) occurs in these models either through coupling to $\delta\chi_k$ via a homogeneous background χ field or through nonlinear evolution effects.

The situation for $g^2/\lambda \gg 1$ is very similar to the previously studied case of a massive inflaton in the broad resonance regime [38–40]. Cosmological-scale $\delta\chi_k$ modes are significantly damped during inflation, and the end of resonant growth is determined by the growing small-scale modes. Already for the second resonance band (centred at $g^2/\lambda = 8$) cosmological metric perturbations are not amplified above the COBE normalization value. This implies that preheating does not alter the standard predictions for the Φ_k normalization in $(\lambda/4)\varphi^4$ inflation for the second and all higher resonance bands. The important difference between the model I have studied and the massive inflaton case is that, in the massive model, weak super-Hubble suppression at small g^2 is accompanied by weak resonant growth during preheating [39], so that no significant super-Hubble amplification is possible.

The inclusion of χ field self-interaction alters the evolution in a predictable way: the resonant growth stops when $\chi^2/\varphi^2 \sim \lambda/\lambda_\chi$, as long as $\lambda_\chi \gtrsim \lambda$. This means that we are unable to rule out models (on the basis of a too large production of metric perturbations) with $\lambda_\chi/\lambda \gtrsim 10^4$. In the negative coupling case, there are two possibilities. For large initial χ backgrounds, $\chi(t_0) \sim m_{\rm P}$, the system becomes essentially single-field, and no resonance occurs (at least until late times). For small initial χ, exponential growth occurs for large enough allowed $|g^2|$.

The Hartree approximation provides a useful approach for the inclusion of the effects of backreaction. However, as mentioned in Section 4.3.1, this approximation misses terms which could contribute to the evolution of fluctuations in an important way. I believe, nevertheless, that my results are sufficiently accurate to predict, for the models studied, whether or not the metric perturbation amplitude after preheating is consistent with the COBE measurement. Nonlinear effects, or rescattering, will primarily affect the detailed evolution of matter fields after backreaction is important. Still, it is of great interest to extend my analysis to a full nonlinear treatment, as was done in the absence of gravitational fluctuations in [8, 52, 56], and including metric fluctuations in [3] for single-field models.

Chapter 5

Long Wavelength Perturbations

I have now developed the formalism of cosmological perturbation theory in Chapter 3, and applied it in the study of parametric resonance in a particular inflationary model in Chapter 4. I showed that for that model, matter and metric perturbations can grow during preheating exponentially fast for long wavelength ($k \to 0$) modes, and that the growth may even exceed the limits on the primordial amplitude of perturbations determined from CMB measurements.

The model studied in Chapter 4, however, was very special in that the homogeneous inflaton trajectory could be expressed analytically in terms of a periodic elliptic cosine function. In more general multi-field models, the background fields may behave very erratically. The questions I will address in this chapter, then, are what can we determine in general from the behaviour of the background fields about the evolution of long wavelength cosmological perturbations? For what kinds of systems might we expect these modes to grow and hence violate the adiabatic conservation law?

To begin, in Section 5.1, I carefully define the notion of adiabaticity (as it is used in cosmological perturbation theory), and then proceed to provide proofs of the constancy of the adiabatic curvature perturbation on uniform density hypersurfaces, ψ_ρ. I extend previous efforts and attempt to elucidate the geometrical significance of the conservation law, so as to clarify the meaning of its violation in the non-adiabatic case.

Next, in Section 5.2, I systematically address the general evolution of long wavelength modes. I begin by considering exactly homogeneous perturbations. I show that by perturbing the background equations considering the time variable as rigid, we obtain the correct perturbation equations in a gauge in which that time variable can at most be shifted by a constant. This provides a key link in the rigourous connection between the behaviour of a background N-component scalar field dynamical system and the behaviour of long wavelength modes. I show that we expect one adiabatic time-translation homogeneous mode and $2N-1$ physical modes which cannot be gauged away. Next I show how realistic inhomogeneous modes can be generated from such homogeneous solutions. The homogeneous adiabatic mode becomes physical, but locally appears to be a time-translation of the background. An additional constraint reduces the physical homogeneous modes to $2N-2$ inhomogeneous modes.

This work is based upon, but an extension of, previously published work.

Finally, in Section 5.3, I revisit parametric resonance in light of the results of this chapter. I then point out a previously undiscussed general route for the amplification of super-Hubble scalar field and metric curvature perturbations in multi-field models, namely through dynamical chaos in the background field evolution. (Note that "dynamical chaos" here refers to the *evolution* of a low-degree-of-freedom dynamical system of homogeneous fields, and should not be confused with "chaotic inflation", which refers to the insensitivity of certain inflationary models to the inflaton *initial conditions.*) This involves instability in the background fields as is the case with parametric resonance, but applies in much more general cases. Since dynamical chaos is common in nonlinear systems with two or more degrees of freedom (*e.g.* in quartically coupled oscillators), I stress that the possible chaotic overproduction of super-Hubble modes must be considered in such inflationary models.

5.1 Conserved curvature perturbations

In studying the evolution of cosmological perturbations it is clearly important to identify any conserved quantities. Such quantities could help propagate perturbations through the vast stretches of time at very early stages when the detailed physics is not known. For example, the production of scalar field and metric perturbations during inflation is reasonably well understood [25]. In an appropriate gauge, the curvature perturbation ψ [recall our scalar metric, Eq. (3.8)] has long been known to be conserved on sufficiently large scales and for a class of perturbation called *adiabatic* (see, *e.g.*, [65]). Thus for a particular inflationary model, we may take the standard inflationary prediction for the curvature perturbation just after the modes leave the Hubble radius during inflation, and immediately deduce the value of the curvature just before the modes re-enter the Hubble radius at late times. The subsequent behaviour of the modes as they seed structure formation is also well understood, and hence the predictions of the inflationary model can be compared with observations of CMB and large-scale structure.

The crucial assumption in the preceeding discussion is that of *adiabaticity.* As I will show, perturbations in single-scalar-field inflationary models are essentially guaranteed to be adiabatic. However, many inflationary models involve multiple scalar fields, which in general allow non-adiabatic perturbations. Hence the above method for connecting inflationary predictions to observable quantities may be suspect.

5.1.1 Adiabaticity

Adiabatic perturbations arise when there is a well-defined equation of state,

$$P = P(\rho). \tag{5.1}$$

This implies that at any spacetime event, the pressure perturbation δP is proportional to the energy density perturbation, $\delta\rho$,

$$\delta P = \frac{dP}{d\rho}\delta\rho = \frac{\dot{P}}{\dot{\rho}}\delta\rho. \tag{5.2}$$

This will be true, *e.g.*, for a pure radiation or matter equation of state, but not for a mix of both. It is useful to decompose a general pressure perturbation into adiabatic and non-adiabatic parts:

$$\delta P = \frac{\dot{P}}{\dot{\rho}}\delta\rho + \delta P - \frac{\dot{P}}{\dot{\rho}}\delta\rho \tag{5.3}$$

$$\equiv \frac{\dot{P}}{\dot{\rho}}\delta\rho + \dot{P}\Gamma, \tag{5.4}$$

where

$$\Gamma = \frac{\delta P}{\dot{P}} - \frac{\delta\rho}{\dot{\rho}} \tag{5.5}$$

is called the *entropy perturbation.* Then we can call

$$\delta P_{\rm ad} = \frac{\dot{P}}{\dot{\rho}}\delta\rho \tag{5.6}$$

and

$$\delta P_{\rm nad} = \dot{P}\Gamma \tag{5.7}$$

the adiabatic and non-adiabatic pressure perturbations, respectively, and write simply

$$\delta P = \delta P_{\rm ad} + \delta P_{\rm nad}. \tag{5.8}$$

Modes for which $\delta P_{\rm nad} \neq 0$ are variously called non-adiabatic, entropy, or isocurvature perturbations. Note that while this terminology is often technically inaccurate or misleading, it is standard.

The entropy perturbation Γ has a very simple physical interpretation. Recall from Section 3.4.4 [in particular Eqs. (3.178) and (3.181)] that

$$T = -\frac{\delta\rho}{\dot{\rho}} \tag{5.9}$$

is the temporal gauge transformation required to move from an arbitrary gauge to uniform density gauge, and that an analogous expression holds for the pressure. Thus Γ is equal to the temporal displacement between uniform density and uniform pressure hypersurfaces. Hence it is a gauge-invariant quantity. Considering the discussion surrounding Eq. (3.162), Γ must be proportional to both δP_ρ and $\delta\rho_P$. [Recall my notation: p_r represents an arbitrary matter or metric perturbation variable p in a gauge specified by $r = 0$ (or $\delta r = 0$), where r (or δr) is any other matter or metric perturbation.] Indeed, we have

$$\delta P_\rho = \dot{P}\Gamma = \delta P_{\rm nad} \tag{5.10}$$

and

$$\delta\rho_P = -\dot{\rho}\Gamma. \tag{5.11}$$

When $\delta P_{\rm nad} = 0$, this interpretation makes it clear that we can simultaneously gauge away both density and pressure perturbations. For a scalar field system, I showed in Section 3.6.1 that the anisotropic stress Π vanishes to linear order. Indeed the anisotropic stress generally vanishes for a perfect fluid. Thus the adiabaticity condition is very strong: In these cases, an adiabatic perturbation satisfies

$$\delta\rho = \delta P = \Pi = 0 \tag{5.12}$$

(in the appropriate gauge), and the only non-zero matter perturbation is the momentum density q. For an exactly homogeneous perturbation, such an adiabatic mode can be completely gauged away!

5.1.2 Conservation laws: Algebraic derivation

It is extremely easy to derive an adiabatic conservation law for the curvature perturbation on uniform density slices [60], using only the perturbed energy conservation law,

$$\delta\dot{\rho} + 3H(\delta\rho + \delta P) + (\rho + P)\left(-3\dot{\psi} + \frac{1}{a^2}\nabla^2\sigma\right) + \frac{1}{a^2}\nabla^2 q = 0, \tag{5.13}$$

which I derived in Section 3.3.2. Evaluating this energy conservation law in uniform density gauge, we have

$$(\rho + P)\dot{\psi}_\rho = H\delta P_{\rm nad} + \frac{1}{3a^2}\nabla^2[q_\rho + (\rho + P)\sigma_\rho], \tag{5.14}$$

where I have used Eq. (5.10). However, using the gauge transformations for σ and q, Eqs. (3.149) and (3.153), we can easily write

$$q_\sigma = q_\rho + (\rho + P)\sigma_\rho, \tag{5.15}$$

so that

$$(\rho + P)\dot{\psi}_\rho = H\delta P_{\rm nad} + \frac{1}{3a^2}\nabla^2 q_\sigma. \tag{5.16}$$

Similarly, we can equivalently write

$$(\rho + P)\dot{\psi}_\rho = H\delta P_{\rm nad} + \frac{\rho + P}{3a^2}\nabla^2 \sigma_q. \tag{5.17}$$

This is the final result: For adiabatic perturbations, for which $\delta P_{\rm nad} = 0$, *and* when the Laplacian of the shear on comoving hypersurfaces, σ_q, can be ignored (equivalently, when the Laplacian of the momentum density in zero-shear gauge, q_σ can be ignored), then the uniform density gauge curvature perturbation ψ_ρ is conserved. This quantity (ψ_ρ) is precisely equal to the parameter ζ introduced by Bardeen, Steinhardt, and Turner [65], who derived its conservation law by a different route. Note that the conservation law does not necessarily apply when $\rho + P = 0$. However, as I discussed in Section 3.4.5, in this case the uniform density gauge becomes singular, so ψ_ρ is not defined.

A closely related quantity is the curvature perturbation on comoving hypersurfaces, ψ_q. Combining the general transformation law for the curvature perturbation, Eq. (3.148), with the transformation required to move from an arbitrary gauge to comoving gauge, Eq. (3.182), we have, for arbitrary gauge perturbations ψ and q,

$$\psi_q = \psi - \frac{Hq}{\rho + P} \tag{5.18}$$

$$= \psi - \frac{H}{\dot{H}}(\dot{\psi} + H\phi). \tag{5.19}$$

The second line here was obtained using the background equation (2.87) and the perturbed momentum constraint (3.282). This expression can also be written using the background energy constraint (2.83) as

$$\psi_q = \psi + \frac{2}{3}\frac{(H^{-1}\dot{\psi} + \phi)}{1 + w}, \tag{5.20}$$

where $w = P/\rho$ is the equation of state. This final form for the comoving curvature perturbation ψ_q, when evaluated in longitudinal gauge *with the assumption that the anisotropic stress* Π *vanishes,* so that $\Psi = \Phi$, becomes precisely the expression given by Mukhanov, Feldman, and Brandenberger [25] for a quantity they showed is conserved on large scales and for $\Pi = 0$,

$$\zeta_{\rm MFB} = \Phi + \frac{2}{3}\frac{(H^{-1}\dot{\Phi} + \Phi)}{1 + w}. \tag{5.21}$$

To see why the comoving curvature perturbation ψ_q is conserved on large scales, notice that the gauge transformations for ψ and ρ imply that

$$\psi_q = \psi_\rho - \frac{H\delta\rho_q}{\dot{\rho}}. \tag{5.22}$$

However, combining the perturbed energy and momentum constraint equations, (3.279) and (3.282), we have

$$\delta\rho_q = \frac{1}{4\pi G a^2}\nabla^2(\psi + H\sigma). \tag{5.23}$$

That is, ψ_q and ψ_ρ differ by a term that vanishes on sufficiently large scales. Therefore, when the uniform density curvature perturbation is conserved, which I showed above requires $\delta P_{\rm nad} = 0$ and scales large enough that the Laplacian of the comoving shear σ_q can be ignored, then so must the comoving curvature perturbation ψ_q be conserved. Under these conditions, and when the anisotropic stress vanishes, the quantity $\zeta_{\rm MFB}$ must therefore also be conserved.

5.1.3 Conservation laws: Geometrical derivation

While the previous derivation of the adiabatic conservation law was very easy, it offered little insight into the physical origin of this law. To elucidate this origin, I will now describe a method to derive the conservation law in a more general form, based on the work of Lyth and Wands [66].

Exactly homogeneous case

I will begin this discussion of the adiabatic conservation law by considering exactly homogeneous perturbations about a homogeneous FRW background. In this special case it will be trivial to demonstrate an adiabatic "conservation law". However, the usefulness of this approach lies in the insight it provides into the physical origin of the long-wavelength conservation law. In addition, generalization to the realistic long-wavelength case will be relatively straightforward.

To begin, consider the homogeneous energy density conservation law,

$$\dot{\rho} + 3H(\rho + P) = 0, \tag{5.24}$$

which I derived in Section 2.2.3. I can write this law in a more revealing form by defining the "number of e-folds of expansion", N, through

$$dN = H\,dt. \tag{5.25}$$

The integrated expansion

$$N = \int H\, dt \tag{5.26}$$

indeed does give the number of e-folds or Hubble times of linear expansion between any two values of time, since

$$\frac{a}{a_0} = e^N \tag{5.27}$$

holds exactly. With this change of time variable the conservation law becomes

$$\frac{d\rho}{dN} + 3(\rho + P) = 0. \tag{5.28}$$

Now the crucial observations. When there is a well-defined equation of state,

$$P = P(\rho), \tag{5.29}$$

which I explained in Section 5.1.1 implies that perturbations are adiabatic, then there is a *unique* solution $\rho(N)$ to Eq. (5.28). In the ρ-$d\rho/dN$ phase plane, the first order autonomous differential equation (5.28) describes a unique curve in this adiabatic case. The autonomous character or N-*translational symmetry* of this first order equation implies that any solution must be of the form $\rho(N + \delta N)$, where δN is a *constant.* That is, any solution must be a trivial translation of the unique phase plane solution along itself. This means that we can exactly determine the perturbation dynamics: If ${}^0\rho(N)$ is the background evolution, then the perturbed dynamics must be of the form

$${}^0\rho(N) + \delta\rho(N) = {}^0\rho(N + \delta N) \tag{5.30}$$

for some constant δN. Thus the density perturbation is given by

$$\delta\rho(N) = {}^0\rho(N + \delta N) - {}^0\rho(N), \tag{5.31}$$

i.e. $\delta\rho(N)$ is simply the (time-dependent) change in ρ produced by a (constant) shift δN (see Fig. 5.1). Note that this relation is *exact.* It can, however, be readily evaluated to any order in δN. To linear order, we have

$$\delta\rho(N) = \frac{d\rho}{dN}\delta N = \frac{\dot{\rho}}{H}\delta N. \tag{5.32}$$

Therefore the quantity

$$\delta N = H\frac{\delta\rho}{\dot{\rho}} \tag{5.33}$$

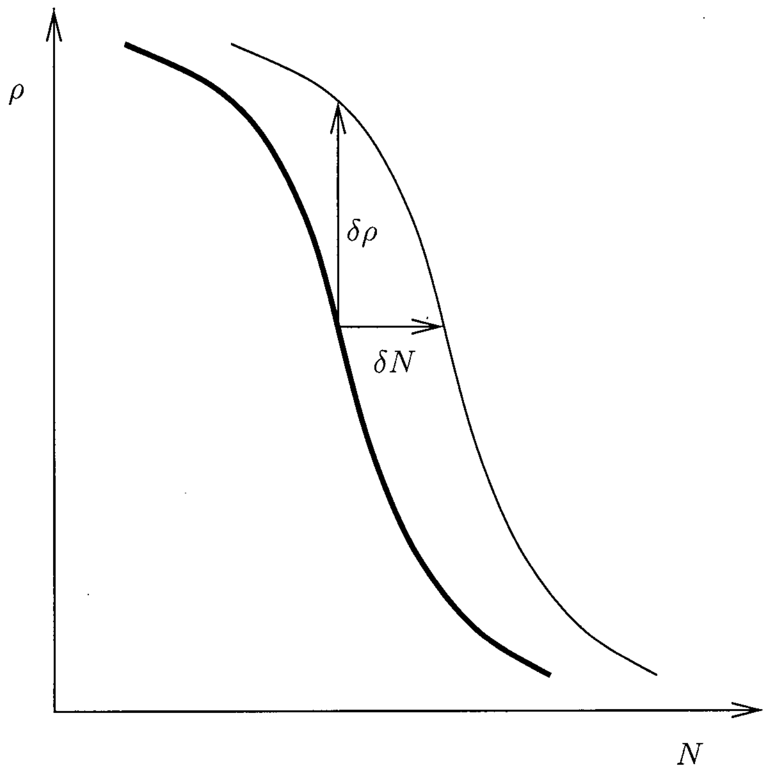

Figure 5.1: For any background evolution ${}^0\rho(N)$ (heavy line) a homogeneous adiabatically perturbed evolution ${}^0\rho(N) + \delta\rho(N)$ (fine line) must be obtained by a constant shift δN. The evolution of $\delta\rho$ is thus trivially determined.

must be conserved to linear order. To second order, we have

$$\delta\rho(N) = \frac{d\rho}{dN}\delta N + \frac{1}{2}\frac{d^2\rho}{dN^2}\delta N^2. \tag{5.34}$$

Inverting this expression and making use of the background equations, I find that the quantity

$$\delta N = H\frac{\delta\rho}{\dot{\rho}} - \frac{\ddot{H}H}{2\dot{H}}\frac{\delta\rho^2}{\dot{\rho}^2} \tag{5.35}$$

must be conserved to second order.

The preceeding "derivation" of adiabatic conserved perturbation quantities should have raised serious alarm bells. First of all, the background equation was perturbed naively, with reckless disregard for general covariance. Secondly, any perturbed quantity, as I explained in great detail in Section 3.4.1, is only ever defined up to a gauge transformation. This certainly applies to the quantity $\delta\rho$ in the above expressions. Given a quantity $H\delta\rho/\dot{\rho}$ which is claimed to be conserved, it is easy to gauge transform it so that it is no longer conserved. So what gauge is $\delta\rho$ to be specified in in order that the expressions (5.33) and (5.35) are conserved? On top of these problems, is the perturbation $\delta\rho$ described above not simply a pure gauge artifact, with no physical significance whatsoever?

To resolve these issues, note first that writing the density perturbation in the form of Eq. (5.31), *i.e.* assuming that it corresponds to a constant shift δN, *amounts to a gauge choice* for the perturbation. In fact, allowing an arbitrary constant (first order) shift $\delta N = H\delta t$ is equivalent to specifying static curvature gauge, where $\dot{\psi} = 0$. Recall from Section 3.4.4 that this gauge choice is only defined up to the residual gauge freedom

$$HT = \delta N = C(x^i), \tag{5.36}$$

for arbitrary spatial function $C(x^i)$. For the exactly homogeneous case I am considering here, it follows that static curvature gauge is defined only up to a *constant* shift δN.

It should be clear now that I *could* have perturbed the background energy conservation equation (5.28) in a more general way, using a time- (or N-) dependent shift δN. This would encompass the full generally covariant freedom available in an exactly spatially homogeneous system. However, I have actually already done this, in deriving (the homogeneous special case of) the perturbed energy conservation equation, Eq. (5.13), which here reads

$$\delta\dot{\rho} + 3H(\delta\rho + \delta P) - 3(\rho + P)\dot{\psi} = 0. \tag{5.37}$$

This equation, evaluated in static curvature gauge (so $\dot{\psi} = 0$), reproduces the "naive" perturbation of the background conservation law above, where I implicitly assumed that the time parameter N was rigid.

Now it is easy to address the second problem raised above. In implicitly assuming that N was unperturbed, or rigid, I restricted the perturbation $\delta\rho$ to be in static curvature gauge. However, if

$$\delta N = H\frac{\delta\rho}{\dot{\rho}} \tag{5.38}$$

is constant for any gauge where $\dot{\psi} = 0$, it must in particular be constant when $\psi = 0$. Therefore

$$\delta N = H\frac{\delta\rho_\psi}{\dot{\rho}} \tag{5.39}$$

must be constant. But according to Eq. (3.161),

$$H\frac{\delta\rho_\psi}{\dot{\rho}} = \psi_\rho, \tag{5.40}$$

and we have recovered the adiabatic long-wavelength limit of Eq. (5.17), which states that the uniform density gauge curvature perturbation is conserved. [There is a subtlety here: it might appear that $\dot{\psi} = 0$ implies that $\delta\rho = 0$. However, only $\dot{\psi}$ (not ψ) appears in the homogeneous Einstein equation. Thus whenever $\dot{\psi} = 0$ all (constant) values of ψ are physically equivalent, and static curvature gauge is indistinguishable from uniform curvature gauge.]

To answer the final objection raised above, yes, $\delta\rho$ is simply a pure gauge mode. Hence this confirms the argument at the end of Section 5.1.1 that homogeneous adiabatic perturbations can be completely gauged away. Nevertheless, this reasoning will generalize to the inhomogeneous case, where it is non-trivial, and similarly to analogous arguments regarding the scalar field equations of motion.

To summarize, in the adiabatic case the background energy conservation equation specifies a unique one-dimensional trajectory in the ρ-$d\rho/dN$ phase plane. If we treat N as a rigid time parameter, which is equivalent to choosing static curvature gauge, then exactly homogeneous perturbations $\delta\rho$ correspond to a *constant* shift δN, and thus we are immediately led to the conserved quantities (5.33) and (5.35). Translating to uniform density gauge, we find the conserved curvature perturbation ψ_ρ.

Inhomogeneous case

I will begin the discussion of the inhomogeneous case by deriving a generalization of the energy conservation law Eq. (5.28) that applies exactly in the presence of spatial perturbations. The derivation begins with the completely general energy conservation law,

$$u_\nu T^{\mu\nu}{}_{;\mu} = -\rho_{,\mu}u^\mu - (\rho + P)u^\mu{}_{;\mu} - q^\mu{}_{;\mu} - q^\nu u^\mu u_{\nu;\mu} - \pi^{\mu\nu}u_{\nu;\mu} = 0, \tag{5.41}$$

which I derived in Section 3.3.2. Once a coordinate chart is chosen, the unit vector field u^μ is defined to be orthogonal to the constant coordinate time hypersurfaces, and q^μ and $\pi^{\mu\nu}$ are the momentum density and anisotropic stress, respectively, in that chart. As we are here considering arbitrary spacetimes, scalar modes no longer evolve independently from vectors and tensors, and hence q^μ and $\pi^{\mu\nu}$ must include vector and tensor parts. Recall that Eq. (5.41) is exact, and not based on any approximation about the closeness of the spacetime to FRW or whatever.

Now I choose the chart to be comoving (if this is possible; recall from Section 3.4.5 that comoving gauge can be ill-defined), so that $q^\mu = 0$. Thus u^μ is the unit comoving vector field, and $\rho_{,\mu}u^\mu = d\rho/d\tau$ is the derivative with respect to proper time τ along the comoving worldlines. Also, $u^\mu{}_{;\mu} = \theta_q$ is the expansion of this field. Finally, with the assumption that

$$\pi^{\mu\nu} u_{\nu;\mu} = 0, \tag{5.42}$$

which, with Eqs. (3.24) and (3.29), implies

$$\pi^{\mu\nu} \sigma_{\mu\nu} = 0, \tag{5.43}$$

I can write the energy conservation law (5.41) as

$$\frac{d\rho_q}{d\tau} + \theta_q(\rho_q + P_q) = 0. \tag{5.44}$$

This relation looks remarkably similar to the homogeneous FRW energy conservation law, Eq. (5.24) (and immediately reproduces it for the homogeneous case). To get this expression into its final form, I will change time variables from proper time τ to the number of e-folds of expansion, N, as I did in the homogeneous case. Here I define N through

$$dN = \frac{1}{3}\theta_q \, d\tau. \tag{5.45}$$

The integrated expansion

$$N = \frac{1}{3}\int \theta_q \, d\tau \tag{5.46}$$

now gives the exact number of e-folds of linear expansion of the comoving vector field along the path of integration. Finally, the energy conservation law becomes

$$\frac{d\rho_q}{dN} + 3(\rho_q + P_q) = 0. \tag{5.47}$$

This simple expression looks just like its homogeneous counterpart, Eq. (5.28), but is exact, subject only to the existence of the comoving chart and the vanishing of $\pi^{\mu\nu}\sigma_{\mu\nu}$.

Once again, in the adiabatic case $P = P(\rho)$ we can conclude that there is a unique solution $\rho_q(N)$ to Eq. (5.47). In the current inhomogeneous case, this means that the comoving densities along any two comoving worldlines differ at most by a constant shift δN, *i.e.* that ρ_q is of the form $\rho_q(N + \delta N)$, where δN depends only on the worldline. Thus the integrated expansion displacement $\Delta N(x^i)$ between any two particular values of ρ_q is the same for each comoving worldline, *i.e.*

$$\Delta N(x^i) = \text{const.} \tag{5.48}$$

Since the set of points where ρ_q takes on any particular value defines a hypersurface Σ_ρ of uniform density, this means that slices of uniform density are separated by uniform integrated expansion ΔN. But as shown in the paragraph surrounding Eq. (5.22), on sufficiently large scales uniform density and comoving hypersurfaces coincide. Therefore on large scales comoving slices are also separated by uniform integrated expansion ΔN.

But I can evaluate this integrated expansion separation ΔN between two comoving hypersurfaces Σ_{q_1} and Σ_{q_2} directly from the definition Eq. (5.46):

$$\Delta N = \frac{1}{3}\int_{\Sigma_{q_1}}^{\Sigma_{q_2}} \theta_q \, d\tau = \frac{1}{3}\int_{t_1}^{t_2} [3H(1-\phi_q) - 3\dot{\psi}_q](1+\phi_q)\, dt \tag{5.49}$$

$$= \int_{t_1}^{t_2} H\, dt - [\psi_q(t_2, x^i) - \psi_q(t_1, x^i)]. \tag{5.50}$$

Here I have used the explicit value of the expansion θ_q from Eq. (3.53) and have ignored the term $\nabla^2\sigma/a^2$, since we are considering large scales. I have also rewritten the integral in terms of the comoving coordinate time t, which is constant along the hypersurfaces Σ_q. I showed in the previous paragraph that $\Delta N = \text{const}$ for each comoving worldline, and clearly $\int_{t_1}^{t_2} H\, dt$ is independent of the worldline since H is unperturbed, so we can now infer that

$$\psi_q(t_2, x^i) - \psi_q(t_1, x^i) = C(t), \tag{5.51}$$

where $C(t)$ is an arbitrary function which is the same for each worldline. This means that ψ_q is constant up to at worst a spatially homogeneous part, $C(t)$. But in the previous subsection I established that a homogeneous adiabatic perturbation ψ_q must be constant (indeed such a perturbation can be gauged away). Therefore we can conclude that ψ_q (and hence ψ_ρ) is constant on large scales, as was to be shown.

The essence of the current argument is that the integrated expansion displacement ΔN between hypersurfaces Σ_ψ of uniform curvature is spatially constant. We can infer this from Eq. (5.51), when we recall that ψ_q is simply equal to the expansion displacement δN between uniform curvature and comoving hypersurfaces, and given that the conservation law Eq. (5.47) tells us that comoving hypersurfaces Σ_q are separated by uniform integrated expansion ΔN. If both the Σ_ψ and the Σ_q are separated by uniform ΔN, then their relative displacement ψ_q along any comoving worldline must be constant up to a spatially homogeneous part. Any such spatially uniform but time dependent shift between the Σ_ψ and the Σ_q must be pure gauge.

Note that this demonstration of the constancy of ψ_ρ, while similar in spirit to that for the homogeneous case above, differs in the important sense that here the gauge is fixed from the start. The density ρ_q is *exact* and hence it already contains the comoving gauge perturbation.

5.2 Long wavelength scalar field perturbations

I will now systematically discuss the evolution of long wavelength scalar field perturbations. I will begin with the homogeneous case, and display a very close connection between perturbations to the background dynamical system and homogeneous cosmological perturbations. I will also illustrate the types of solutions we expect (adiabatic, time-translation modes and physical, non-gauge modes). This will generalize readily to the long wavelength inhomogeneous case. A crucial point to make here is that for multi-field models, there is much more "freedom to perturb" the system than there is gauge freedom, so there are many physical modes.

In this section I consider exclusively a system of N scalar fields, so that for linear perturbations the anisotropic stress Π vanishes, as I showed in Section 3.6.1.

5.2.1 The homogeneous equations

It will be very useful to begin my discussion of the behaviour of long wavelength scalar field perturbations with a thorough treatment of the precisely homogeneous case, as I did for adiabatic conservation laws in the previous section. I will start with some general remarks on the nature of the exact solutions to the homogeneous equations of motion, which I can take to be the energy constraint or Friedmann equation, Eq. (2.83),

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho, \tag{5.52}$$

and the homogeneous Klein-Gordon equation, Eq. (2.126),

$$\ddot{\varphi}_A + 3H\dot{\varphi}_A + V_{,\varphi_A} = 0. \tag{5.53}$$

Note that while previously these equations had been considered as descriptions of the (fictitious) background, and inhomogeneous perturbations satisfied the equations of motion derived in Chapter 3, here any homogeneously perturbed universe must also satisfy Eqs. (5.52) and (5.53) exactly.

To consider how these equations could be solved, note that the scalar field energy-momentum tensor can be used to write the energy density ρ in terms of the scalar fields [recall Eq. (2.128)],

$$\rho = \frac{1}{2}\dot{\varphi}\cdot\dot{\varphi} + V(\varphi_A). \tag{5.54}$$

Also note that for any realistic cosmological model, we must have $H > 0$ at early times (of course at very late times we may have $H = 0$ momentarily when a closed universe begins to recollapse). In this case we can use the square root of (5.52) to write H uniquely in terms of the scalar fields, and hence we obtain a *closed* set of N coupled second order equations,

$$\ddot{\varphi}_A + 3H(\varphi_B, \dot{\varphi}_B)\dot{\varphi}_A + V_{,\varphi_A} = 0. \tag{5.55}$$

Thus we expect a unique solution $\varphi_A(t)$ which will depend on $2N$ arbitrary parameters (*e.g.* the initial values of φ_A and $\dot{\varphi}_A$). Once we have obtained such a solution (of course these are nonlinear equations and in practice it may be impossible to obtain an explicit solution without divine intervention) we can substitute it back into Eq. (5.52) and integrate to obtain the scale factor,

$$\frac{a}{a_0} = e^{\int H(t)\, dt}. \tag{5.56}$$

Note that there is no *physical* constant of integration here—the value a_0 is not observable. This is ultimately due to the homogeneity of the spacetime, since a constant rescaling of the scale factor a is equivalent to a spatial gauge transformation, which I explained in Section 3.4.3 does not affect the equations of motion. Therefore, we expect that the evolution of the scalar fields and metric is fully determined by $2N$ arbitrary parameters.

However, it should be clear that one of these parameters can be chosen to be a measure of the time at which the initial conditions are specified. That is, varying this parameter simply translates any solution along itself in the $2N$-dimensional scalar field phase space. As I will discuss in more detail below, this parameter corresponds to a pure adiabatic gauge mode. We then expect only $2N - 1$ physical parameters.

5.2.2 Exactly homogeneous perturbations

Next I will consider the evolution of exactly homogeneous perturbations from some homogeneous background solution. I will assume that the perturbations are small enough that they can be treated by linearized theory. The approach I take here was inspired by Sasaki and Tanaka [70].

Synchronous gauge and the time-translation mode

Perturbing the homogeneous Klein-Gordon equation, (5.53), I find

$$\delta\ddot{\varphi}_A + 3H\delta\dot{\varphi}_A + 3\delta H\dot{\varphi}_A + V_{,\varphi_A\varphi} \cdot \delta\varphi = 0. \tag{5.57}$$

Similarly, the linearization of the energy constraint Eq. (5.52) gives

$$2H\delta H = \frac{8\pi G}{3}\delta\rho. \tag{5.58}$$

The density perturbation can be evaluated explicitly in terms of the scalar field perturbations and the backgrounds using Eq. (5.54),

$$\begin{aligned}\delta\rho &= \frac{\partial\rho}{\partial\dot{\varphi}}\cdot\delta\dot{\varphi} + \frac{\partial\rho}{\partial\varphi}\cdot\delta\varphi &(5.59)\\ &= \dot{\varphi}\cdot\delta\dot{\varphi} + V_{,\varphi}\cdot\delta\varphi. &(5.60)\end{aligned}$$

Combining these expressions we obtain a closed set of equations for the scalar field perturbations,

$$\delta\ddot{\varphi}_A + 3H\delta\dot{\varphi}_A + \frac{4\pi G}{H}(\dot{\varphi}\cdot\delta\dot{\varphi} + V_{,\varphi}\cdot\delta\varphi)\dot{\varphi}_A + V_{,\varphi_A\varphi}\cdot\delta\varphi = 0. \tag{5.61}$$

To these linear equations we expect $2N$ independent solutions $\delta\varphi_A(t)$, once the background evolution $\varphi_A(t)$ has been specified.

Here, just like in the discussion of conservation laws in Section 5.1.3, there should appear to be serious problems with this simple "derivation" of the perturbation equations. In particular, what of general covariance? What gauge are $\delta\varphi_A$ and $\delta\rho$ to be specified in? Surely, given some solution to Eq. (5.61), I have the freedom to gauge transform it. Will the result of this transformation also be a solution to Eq. (5.61)? Also, what about the adiabatic time-translation mode that I promised?

The solution to these questions lies, as before, in the observation that in perturbing the background equations, I have implicitly assumed that the time coordinate t is rigid. That is, my method of perturbing Eqs. (5.53) and (5.54) was not fully generally covariant. At most, my method can accomodate a trivial reparametrization of time, $t \to t + T$, for constant T. But this is precisely the freedom that exists in synchronous gauge (where $\phi = 0$) in the homogeneous case (recall Section 3.4.4). Thus, in naively perturbing the homogeneous background equations as I have, I have not made any error; rather, I have implicitly restricted myself to synchronous gauge.

To verify this reasoning, we only need to evaluate the fully generally covariant perturbed Klein-Gordon equation, which I derived in Section 3.6.2 [Eq. (3.346)], in the homogeneous synchronous gauge case, with the result

$$\delta\ddot{\varphi}_A + 3H\delta\dot{\varphi}_A - 3\dot{\psi}_\phi\dot{\varphi}_A + V_{,\varphi_A\varphi}\cdot\delta\varphi = 0. \tag{5.62}$$

Next, the perturbed energy constraint Eq. (3.279) gives, in homogeneous synchronous gauge,

$$3\dot{\psi}_\phi = -\frac{4\pi G}{H}\delta\rho_\phi, \tag{5.63}$$

and the synchronous gauge perturbed energy density is [recall Eq. (3.324)]

$$\delta\rho_\phi = \dot{\varphi} \cdot \delta\dot{\varphi} + V_{,\varphi} \cdot \delta\varphi \tag{5.64}$$

[which matches Eq. (5.60)]. Combining these expressions, I recover Eq. (5.61) precisely. (In each of these expressions, the field perturbation $\delta\varphi_A$ must also be specified in synchronous gauge, but I have ignored the ϕ subscripts to reduce clutter.)

Next I will discuss the time-translation mode, which I argued in Section 5.2.1 must generally exist. First, we can write the homogeneous background equation (5.55) as an autonomous $2N$-dimensional dynamical system,

$$\dot{y}_i = F_i(y_j), \tag{5.65}$$

with the identifications

$$y_i = \varphi_i, \qquad y_{i+N} = \dot{\varphi}_i, \qquad i = 1, \ldots, N. \tag{5.66}$$

The flow vector F_i is a nonlinear function of the fields and velocities which I do not need to state explicitly. For any such autonomous system, we can trivially find the linearization of the time-translation mode. Any linear perturbation of Eq. (5.65) about some background solution $\bar{y}_i(t)$ must satisfy

$$\delta\dot{y}_i = \left.\frac{\partial F_i}{\partial y_j}\right|_{\bar{y}} \delta y_j. \tag{5.67}$$

But the time derivative of Eq. (5.65) gives

$$\ddot{y}_i = \frac{\partial F_i}{\partial y_j}\dot{y}_j. \tag{5.68}$$

Thus we can immediately conclude that

$$\delta y_i(t) = \dot{y}_i(t) T \tag{5.69}$$

is a solution to the linearized equations about any background solution $y_i(t)$ and for any *constant* T. Since $\dot{y}_i$ is the tangent vector to the $2N$-dimensional phase space trajectory, this perturbed solution indeed corresponds to a translation along the trajectory.

Note that this result again assumes that the time t is rigid. Hence we can apply it to the perturbed Klein-Gordon equation (5.61), which was derived under the same assumption, and conclude that

$$\delta\varphi_A(t) = \dot{\varphi}_A(t) T, \tag{5.70}$$

for constant T, is the time-translational mode of that equation. [Indeed it is not hard to verify that Eq. (5.70) *is* a solution of the perturbed Klein-Gordon equation (5.61).] Since T is constant, Eq. (5.70) implies that

$$(\delta\varphi_A(t), \delta\dot{\varphi}_A(t)) = T(\dot{\varphi}_A(t), \ddot{\varphi}_A(t)), \tag{5.71}$$

i.e. the perturbation is parallel to the background trajectory in the full $2N$-dimensional phase space. But this solution corresponds precisely to a pure synchronous gauge mode. Recall that in synchronous gauge we have the freedom to perform a temporal gauge transformation $t \to t + T$ for constant T. According to the scalar field gauge transformation law (3.328), such a transformation induces a scalar field perturbation which is exactly that given by Eq. (5.70). Furthermore, inserting expression (5.70) into Eq. (5.60) for the perturbed density and using the background equations of motion, it is very easy to show that for the time-translational mode

$$\delta\rho = \dot{\rho}T. \tag{5.72}$$

Similarly, I find

$$\delta P = \dot{P}T, \tag{5.73}$$

and using Eq. (5.63) I obtain

$$\dot{\psi} = -(\dot{H}T). \tag{5.74}$$

Recalling gauge transformations (3.148) to (3.152), each of these perturbed quantities is precisely that expected for a pure synchronous gauge mode (of course the momentum density q and shear σ vanish in this homogeneous case). Also, Eqs. (5.72) and (5.73) immediately tell us that this mode satisfies condition (5.2), and hence is adiabatic.

Static curvature gauge and the N-translational mode

To further illustrate these ideas, I will derive the perturbed homogeneous Klein-Gordon equation in another gauge, static curvature gauge, where $\dot{\psi} = 0$. Recall from Section 3.4.4 that, in the homogeneous case, this gauge choice contains the residual freedom to transform by

$$T = \frac{C}{H}, \tag{5.75}$$

where C is a constant. As I discussed in Section 5.1.3, this is equivalent to the freedom to shift the time variable N defined by

$$dN = H\,dt \tag{5.76}$$

by a constant amount $\delta N = C$. The variable N is a measure of the elapsed number of e-folds of expansion.

Based on my previous arguments, we should expect to be able to derive the static curvature gauge Klein-Gordon equation by naively perturbing the background equation written in terms of the time variable N, *i.e.* by treating N as rigid. To change variables from t to N, I need the relations

$$\dot{f} = Hf' \tag{5.77}$$

and

$$\ddot{f} = H^2 f'' + \dot{H} f', \tag{5.78}$$

for arbitrary function f, and where, *for this subsection only,* I define

$$f' = \frac{\partial f}{\partial N}. \tag{5.79}$$

The homogeneous Klein-Gordon equation (5.53) then becomes

$$\varphi_A'' + \left(3 + \frac{H'}{H}\right)\varphi_A' + \frac{V_{,\varphi_A}}{H^2} = 0. \tag{5.80}$$

Linearizing this equation, and then translating back to time t, I find

$$\delta\ddot{\varphi}_A + 3H\delta\dot{\varphi}_A + \left(\frac{\delta H}{H}\right)^{\cdot}\dot{\varphi}_A - 2\frac{\delta H}{H}V_{,\varphi_A} + V_{,\varphi_A\varphi}\cdot\delta\varphi = 0. \tag{5.81}$$

But, in static curvature gauge, the homogeneous energy constraint equation (3.279) gives

$$\frac{\delta H}{H} = \frac{4\pi G}{3}\frac{\delta\rho}{H^2} = -\phi. \tag{5.82}$$

Therefore, evaluating the general gauge perturbed Klein-Gordon equation, Eq. (3.346), in static curvature gauge, and substituting this expression for $\delta H/H$, we immediately confirm that Eq. (5.81) is correct. Again, we have a pure adiabatic gauge mode due to the residual gauge freedom, but in this case it is an N-translation mode,

$$\delta\varphi_A(N) = \varphi_A'(N)\delta N = \dot{\varphi}_A(t)T(t), \tag{5.83}$$

for constant δN and $T(t) = \delta N/H$. This mode corresponds to a constant translation δN along the trajectory (φ_A, φ_A') for solutions to Eq. (5.80).

Physical modes

So what have we accomplished in deriving the synchronous or static curvature gauge Klein-Gordon equation in this way, when we already had the general gauge equation? This certainly is an easy way to derive the gauge-fixed equations, but the real importance of this result is that it allows us to rigourously connect the behaviour of the low-dimensional background dynamical system (5.55) or (5.80) to the evolution of *physical* homogeneous cosmological perturbations. The dynamical system of homogeneous scalar fields (5.55) or (5.80) may be known to exhibit various behaviours, such as parametric resonance or dynamical chaos. This behaviour can be determined by the ordinary techniques for dynamical systems, without any concern for general covariance. Then, the evolution of homogeneous cosmological perturbations in synchronous or static curvature gauge is given *precisely* by the evolution of perturbations of the dynamical system.

To summarize, if we perturb the homogeneous equations of motion treating the time t (or N) as rigid, we obtain the perturbed equations in synchronous (or static curvature) gauge. The residual freedom in these gauges corresponds precisely to an adiabatic time- (or N-) translational mode, which must always exist. Fixing the (arbitrary) amplitude of this mode amounts to fixing the remaining freedom in the gauge. There are then $2N - 1$ *physical* modes $\delta\varphi_A$ which cannot be gauged away. The evolution of these physical modes is determined *precisely* by the evolution of perturbations of the background scalar field equation, treated like an ordinary dynamical system. The full usefulness of this approach will become more apparent in the next sections, where I show how to generate realistic inhomogeneous perturbations from homogeneous ones.

5.2.3 Inhomogeneous perturbations

My interest in this thesis is in the evolution during reheating of those perturbations which eventually became the cosmological scale modes that we observe in the CMB or in large-scale structure. These modes must have left the Hubble radius roughly 60 or 70 e-folds before the end of inflation in order to solve the flatness, horizon, and relic problems, as I discussed in Section 2.3.2. Therefore, during reheating we have $aH/k \sim 10^{25}$ or 10^{30} for these cosmological scales. The immensity of this ratio of mode wavelength to the only important physical scale, the Hubble length, should convince us that the behaviour of these modes is essentially identical to the homogeneous evolution described in the previous subsection. Nevertheless, it is important to consider what effects departures from homogeneity might have. Indeed, it is those departures from homogeneity that produced the observable structure today.

Recall the general gauge perturbed Klein-Gordon equation, Eq. (3.346),

$$\delta\ddot{\varphi}_A + 3H\delta\dot{\varphi}_A - \frac{1}{a^2}\nabla^2\delta\varphi_A - \left(\dot{\phi} + 3\dot{\psi} - \frac{1}{a^2}\nabla^2\sigma\right)\dot{\varphi}_A + 2\phi V_{,\varphi_A} + V_{,\varphi_A\varphi}\cdot\delta\varphi = 0. \qquad (5.84)$$

Above I discussed the homogeneous case, where both Laplacian terms can be ignored. In that case, the momentum constraint and off-diagonal Einstein equations both vanish, so the dynamics is completed by the perturbed energy constraint, Eq. (3.279). In order to obtain a closed set of equations for the homogeneous scalar field perturbation $\delta\varphi_A$ we must make a gauge choice, such as $\phi = 0$ or $\dot{\psi} = 0$ (or some linear combination of these), and then use the perturbed energy constraint to eliminate the remaining metric function. Whether we set $\phi = 0$ (synchronous gauge) or $\dot{\psi} = 0$ (static curvature gauge), we cannot fix the gauge completely, and hence the solutions $\delta\varphi_A$ must contain a residual gauge mode, which corresponds simply to time- or N-translations of the background. The remaining $2N-1$ modes, however, are true physical modes which cannot be gauged away.

Now, when we consider inhomogeneous perturbations, Eq. (5.84) becomes a partial differential equation. But because it is linear, when we Fourier-expand the perturbations in k-space, modes for different k do not couple. Thus, for any value of k, Eq. (5.84) again becomes a set of N ordinary differential equations. Of course it is, nevertheless, more complicated than in the homogeneous case. In particular, we can no longer ignore the shear σ. In addition, the momentum density does not in general vanish, and the momentum constraint and off-diagonal equations are non-trivial. To put the perturbed Klein-Gordon equation (5.84) into closed form we must fix the gauge completely. For example, we can specify zero shear gauge, where $\sigma = 0$. Then, as I discussed in Section 3.6.2, the off-diagonal Einstein equation tells us that $\psi = \phi$. Combining the energy and momentum constraint equations, we can solve for ψ in terms of the scalar field perturbations [recall Eq. (3.333)]. Substituting this expression back into the Klein-Gordon equation, we obtain a closed set of equations for $\delta\varphi_A$. There are thus $2N$ independent modes for each k, although one will be removed by a further constraint provided by the momentum constraint equation, leaving $2N-1$ modes. All of these are physical, as the gauge is fixed completely.

We do not expect to find a time-translation mode in this inhomogeneous case, since the Klein-Gordon equation for $k \neq 0$ is *not* a simple homogeneous perturbation of the background equation. For example, if the background fields oscillate at some frequency ω, we expect a mode k to oscillate at a frequency $\sqrt{\omega^2 + k^2}$. Thus if we set such a mode so that initially, at some position, it corresponds to a perturbation along the background trajectory, *i.e.* $(\delta\varphi_A, \delta\dot{\varphi}_A) \propto (\dot{\varphi}_A, \ddot{\varphi}_A)$, then the mode will not in general remain along the background trajectory.

Interestingly, we certainly *can* nevertheless always create an inhomogeneous pure gauge mode. If we begin with an exactly homogeneous universe and perform a gauge transformation $T_k(t)$ at any wavenumber k, then we trivially produce a scalar field perturbation $\delta\varphi_{Ak} = \dot{\varphi}_A T_k$. The reason this is possible is that general covariance demands that we also generate a corresponding shear $\sigma_k = T_k$ which *precisely* cancels the gradient term $\nabla^2\delta\varphi_A/a^2$ in the Klein-Gordon equation! Thus the gauge "perturbation" can evolve at the same frequency as the background and remain along it.

These remarks apply for arbitrary k. In the next subsection I will specialize to the long-wavelength case, and discuss a precise connection between homogeneous and long-wavelength perturbations.

5.2.4 Generating long wavelength solutions...

The situation regarding scalar field perturbations in the exactly homogeneous case was very simple: we found one adiabatic, pure gauge, time-translation mode, and $2N-1$ physical modes. I argued in the previous subsection that we again expect $2N-1$ physical modes for each k in the inhomogeneous case. I will now describe a method to generate long wavelength scalar field solutions from exactly homogeneous ones. This is based on the work of Kodama and Hamazaki [67], but see also [29, 68].

... from the homogeneous gauge mode

To start, consider the homogeneous time-translational mode in synchronous gauge,

$$\delta\varphi_A(t) = \dot{\varphi}_A T, \tag{5.85}$$

for constant T. I showed in Section 5.2.2 that for this mode we have

$$\delta\rho = \dot{\rho}T, \qquad \delta P = \dot{P}T, \qquad \psi = -HT - C, \tag{5.86}$$

for arbitrary constant C [recall Eqs. (5.72) to (5.74)]. Of course, we also have $\phi = \sigma = q = 0$ in this case. Now let me propose that each of the expressions (5.85) and (5.86) holds for an *inhomogeneous* mode at some (long) wavenumber k. That is, I will write

$$\delta\varphi_{Ak}(t) = \dot{\varphi}_A T_k, \qquad \delta\rho_k = \dot{\rho}T_k, \qquad \delta P_k = \dot{P}T_k, \qquad \psi_k = -HT_k - C_k, \tag{5.87}$$

where again T_k and C_k are constants. Next, I will choose the shear σ_k to exactly satisfy the off-diagonal Einstein equation, Eq. (3.283). Since the linear anisotropic stress vanishes for scalar fields, we can rewrite this equation as

$$\frac{(a\sigma_k)^{\cdot}}{a} = \phi_k - \psi_k. \tag{5.88}$$

Inserting the expression (5.87) for ψ_k, we can integrate this equation to give

$$\sigma_k = \frac{1}{a}\left[\int a(HT_k + C_k)\,dt + D_k\right] \tag{5.89}$$

$$= T_k + \frac{1}{a}\left(C_k \int a\,dt + D_k\right), \tag{5.90}$$

where D_k is an integration constant.

Now that I have my proposed inhomogeneous solution, I must check that it does in fact satisfy Einstein's equation in the long wavelength limit. First of all, for scalar fields the momentum density is given by $q_k = -\dot{\varphi}\cdot\delta\varphi_k$ [recall Eq. (3.325)]. Thus for my proposed solution (5.87), the momentum constraint equation (3.282) becomes

$$-\dot{H}T_k = 4\pi G\dot{\varphi}\cdot\dot{\varphi}T_k, \tag{5.91}$$

which is satisfied by virtue of the background equations. Next, the off-diagonal Einstein equation is satisfied because I chose σ_k such that it was. Third, the trace Einstein equation (3.284) is automatically satisfied, since it was satisfied for the homogeneous solution I began with, and it contains no spatial gradients or appearances of σ_k. Finally, the energy constraint equation (3.279) is satisfied for the same reason, when we ignore the Laplacian term $k^2(\psi_k + H\sigma_k)/a^2$. This should be valid as long as the constants C_k and D_k do not diverge like $1/k^2$ or faster as $k \to 0$. Thus my proposed solution is indeed a long wavelength solution.

Notice that my long wavelength solution (5.87) is always adiabatic, so according to the results of Section 5.1 the curvature perturbation ψ_ρ must be conserved. Indeed, it is easy to see that

$$\psi_{\rho k} = -C_k. \tag{5.92}$$

Note however that when $C_k = D_k = 0$, the solution becomes *exactly* a pure gauge mode. Thus when these constants do not vanish, the solution is physical and cannot be gauged away. Therefore *locally, i.e.* over regions much smaller than $1/k$ in size, such a physical mode can be considered as essentially a pure synchronous gauge mode, *i.e.* a time-translation of the background. However, over regions larger than $1/k$, it cannot be gauged away, as the shear becomes important. Locally the mode appears indistinguishable from the background, while the "physicality" of the mode is encoded in the long wavelength behaviour. Indeed this "long wavelength behaviour" becomes "short wavelength behaviour" when such a mode reenters the Hubble radius at late times and contributes to the clearly physical formation of structure.

However, there is something odd about this picture. When I checked that the components of Einstein's equation were satisfied, I (justifiably, apparently) ignored terms of order k^2 in the energy constraint, but I deemed that the off-diagonal equation, which is entirely of order k^2, be satisfied exactly. Why should I have ignored k^2 terms in one component, only to keep another component which is itself of order k^2? Would not *any* reasonably well-behaved shear σ_k, together with Eq. (5.87), satisfy all components of Einstein's equation to order k?

...from the homogeneous physical modes

I have now shown that we can generate a physical, adiabatic, long wavelength mode from the time-translational, pure gauge, homogeneous mode. Next I will try to duplicate this success with the *physical* homogeneous modes. In Section 5.2.2 I deduced that there were $2N-1$ such modes, each of which I can write in synchronous gauge as the (in general non-vanishing) set

$$(\delta\varphi_A, \delta\rho, \delta P, \psi), \tag{5.93}$$

together with

$$\phi = \sigma = q = 0. \tag{5.94}$$

The explicit forms of the non-vanishing perturbations are determined by the background evolution, and are not important here beyond that they do not constitute a pure gauge mode.

Once again, I propose that an inhomogeneous solution be the set

$$(\delta\varphi_{Ak}, \delta\rho_k, \delta P_k, \psi_k), \tag{5.95}$$

obtained by "promoting" the set (5.93) to wavenumber k. I maintain the synchronous condition $\phi_k = 0$. Again I deem that σ_k satisfy the off-diagonal equation, which gives

$$\sigma_k = -\frac{1}{a}\left(\int a\psi_k\, dt + D_k\right). \tag{5.96}$$

Exactly as before, this solution satisfies the off-diagonal and trace equations, and also the energy constraint equation when we ignore terms of order k^2.

However, things are different with the momentum constraint. Now, combining the momentum and energy constraints, I find that the scalar field perturbations in my proposed solution (5.95) must satisfy the constraint

$$4\pi G(\ddot{\varphi} \cdot \delta\varphi_k - \dot{\varphi} \cdot \delta\dot{\varphi}_k) = \frac{k^2}{a^2}(\psi_k + H\sigma_k) \tag{5.97}$$

[recall Eq. (3.333)]. For long wavelength solutions we can ignore the order k^2 terms and simply write

$$\ddot{\varphi} \cdot \delta\varphi_k - \dot{\varphi} \cdot \delta\dot{\varphi}_k = 0. \tag{5.98}$$

This single constraint on $\delta\varphi_A$ will reduce the number of these modes from the $2N-1$ homogeneous modes we began with to $2N-2$. This constraint actually has a very simple geometrical interpretation in the $2N$-dimensional scalar field phase space: if we write it in the form

$$(\delta\varphi_k, \delta\dot{\varphi}_k) \cdot (\ddot{\varphi}, -\dot{\varphi}) = 0, \tag{5.99}$$

we see that it simply means that the perturbation vector $(\delta\varphi_{Ak}, \delta\dot{\varphi}_{Ak})$ must have no component along the direction $(\ddot{\varphi}_A, -\dot{\varphi}_A)$, which is defined in terms of the background fields. That is, the perturbation vector is constrained to a $(2N-2)$-dimensional subspace in the tangent space at the background point.

The question of the significance of the direction $(\ddot{\varphi}_A, -\dot{\varphi}_A)$ to the background field dynamics is certainly an interesting one. For the simplest single-field case, $N = 1$, this direction is simply the orthogonal to the background trajectory $(\dot{\varphi}, \ddot{\varphi})$. In this case there is one pure gauge homogeneous solution, parallel to the background trajectory, and there is only $2N-1 = 1$ physical homogeneous mode, which must be orthogonal to the background trajectory since it is independent of the gauge mode. Therefore, in the single-field case, the constraint (5.99) eliminates the single mode generated from the physical homogeneous mode, and we are left with only the adiabatic mode. I do not know the significance of the direction $(\ddot{\varphi}_A, -\dot{\varphi}_A)$ in the case $N > 1$.

To summarize, I have shown how the homogeneous modes I discussed in Section 5.2.2 can be used to generate physical inhomogeneous solutions. The homogeneous, pure gauge, time-translational mode can be "promoted" to a physical, adiabatic k-mode. This mode conserves the curvature perturbation ψ_ρ and looks locally like a time-translation of the background, but on large scales is not. The $2N-1$ homogeneous physical modes can be promoted to just $2N-2$ physical k-modes, as they are subject to a simple geometrical constraint. These modes do not look like translations of the background trajectory even locally, as they correspond to perturbations orthogonal to it. In the next section I will describe how these non-adiabatic modes might behave.

5.3 Non-adiabatic long wavelength modes

In the previous section I developed in some detail a description of long wavelength scalar field perturbations. I showed that in the single-field case, only an adiabatic mode is possible, which conserves the curvature perturbation ψ_ρ, looks locally like a time-translation of the background trajectory, and is responsible for the standard inflationary spectrum of perturbations. For a system with $N = 2$ or more scalar fields, however, I showed that we expect $2N-2$ physical modes (for each k) which can be generated from perturbations orthogonal to the background trajectory, and whose behaviour is determined only by the behaviour of the corresponding homogeneous scalar field dynamical system. It is finally time to examine how these modes might evolve and what consequences if any they may have on the spectrum of inflationary perturbations. Two important types of behaviour of the background dynamical system will be *periodic orbits,* which can lead to parametric resonance, and *dynamical chaos,* which leads to a more general type of instability.

5.3.1 Cosmological phase space flows

First I will demonstrate an important result regarding the phase space flow of a cosmological scalar field system. Consider an arbitrary $2N$-dimensional autonomous dynamical system

$$\dot{y}_i = F_i(y_j). \tag{5.100}$$

If V is some $2N$-dimensional volume in the phase space, then we can use Gauss's theorem to write

$$\frac{dV}{dt} = \int_V \boldsymbol{\nabla} \cdot \boldsymbol{F}\, d^{2N}y \tag{5.101}$$

for the rate of change of the volume V when it evolves under the flow F_i, where

$$\boldsymbol{\nabla} \cdot \boldsymbol{F} \equiv \frac{\partial F_i}{\partial y_i}. \tag{5.102}$$

Taking the limit of a small volume, we have

$$\lim_{V\to 0} \frac{1}{V}\frac{dV}{dt} = \boldsymbol{\nabla} \cdot \boldsymbol{F}, \tag{5.103}$$

or

$$\lim_{V\to 0} V = V_0 \exp\left(\int \boldsymbol{\nabla} \cdot \boldsymbol{F}\, dt\right) \tag{5.104}$$

for constant V_0.

Now, as before, I can write the scalar field homogeneous background equation of motion

$$\ddot{\varphi}_A + 3H(\varphi_B, \dot{\varphi}_B)\dot{\varphi}_A + V_{,\varphi_A} = 0 \tag{5.105}$$

in the form of the dynamical system (5.100) with the identifications

$$\left.\begin{aligned} y_i &= \varphi_i, & y_{i+N} &= \dot{\varphi}_i, \\ F_i &= \dot{\varphi}_i, & F_{i+N} &= -3H\dot{\varphi}_i - V_{,\varphi_i} \end{aligned}\right\} \qquad i = 1, \ldots, N. \tag{5.106}$$

To evaluate the divergence $\nabla \cdot \boldsymbol{F}$ I need

$$\frac{\partial H}{\partial \dot{\varphi}_A} = \frac{4\pi G}{3} \frac{\dot{\varphi}_A}{H}. \tag{5.107}$$

The result is then

$$\nabla \cdot \boldsymbol{F} = -3NH - 4\pi G \frac{\dot{\varphi} \cdot \dot{\varphi}}{H} \tag{5.108}$$

$$= -3NH + \frac{\dot{H}}{H}. \tag{5.109}$$

For all realistic models in an expanding universe we have $\dot{H} \leq 0$ and $H > 0$, so the divergence is strictly negative,

$$\nabla \cdot \boldsymbol{F} < 0. \tag{5.110}$$

In particular, during inflation we have $|\dot{H}| \ll H^2$ [recall Eq. (2.158)], so that

$$\nabla \cdot \boldsymbol{F} \simeq -3NH. \tag{5.111}$$

The result (5.110), with Eq. (5.101), implies that phase space volumes for the homogeneous scalar fields are always contracting in an expanding universe. The result (5.111) implies that phase space volumes contract very quickly during inflation:

$$V \simeq V_0 e^{-3NHt}. \tag{5.112}$$

Let us consider what this result means for the evolution of perturbations $(\delta\varphi, \delta\dot{\varphi})$ of the background dynamical system of scalar fields, Eq. (5.105), during single-field inflation. We know that one solution to the perturbed dynamical system is the adiabatic time-translation mode,

$$\delta\varphi_{\rm ad} = \dot{\varphi} T, \tag{5.113}$$

for constant T. But recall from Section 2.3.2 that during slow roll inflation, the inflaton φ undergoes overdamped decay, and hence the second time derivative can be ignored in the Klein-Gordon equation, *i.e.*

$$|\ddot{\varphi}| \ll H|\dot{\varphi}|. \tag{5.114}$$

Thus the adiabatic perturbation mode must decay slowly on the Hubble time scale,

$$|\delta\dot{\varphi}_{\rm ad}| \ll H|\delta\varphi_{\rm ad}|. \tag{5.115}$$

The second perturbation mode $(\delta\varphi_{\rm nad}, \delta\dot{\varphi}_{\rm nad})$ must contain a component orthogonal to the background trajectory $(\varphi, \dot{\varphi})$. But, since the adiabatic mode decays only very slowly, while the two-dimensional phase space volume decreases like $\exp(-3Ht)$, the non-adiabatic orthogonal mode *must* also decay very quickly, like

$$\delta\varphi_{\rm nad} \propto e^{-3Ht}. \tag{5.116}$$

We can readily visualize the phase space dynamics. If we imagine a disc of perturbations $(\delta\varphi, \delta\dot{\varphi})$ centred on some background trajectory at an initial time, then under the inflationary phase space flow, the area of the disc must decay like $\exp(-3Ht)$, while the extent of the disc along the trajectory, *i.e.* the amplitude of the adiabatic mode, must remain roughly constant. Therefore the disc becomes squeezed exponentially quickly into a very thin ellipse along the background trajectory. This, of course, is a description of a well-known phenomenon. The background field dynamics during inflation is sometimes said to exhibit *attractor solutions* [25], which in the current view corresponds to the rapid decay of perturbations orthogonal to any background trajectory onto that trajectory. Similarly, it is often pointed out that during slow roll, the field dynamics is essentially reduced to first order [recall Eq. (2.161)]. This means that solutions are, to good approximation, described by a single parameter, which corresponds to the amplitude of the adiabatic mode.

The important point here is that this straightforward and very easily visualized result concerning phase space flows and perturbations of the background dynamical system translates directly, via the arguments of the preceeding section, into a statement about the behaviour of long wavelength cosmological perturbations. In particular, we can conclude that after only a few e-folds of inflation, long wavelength cosmological perturbations for a single-field system are dominated by the adiabatic mode. Finally, I note that this behaviour is consistent with the result I demonstrated above, that the constraint Eq. (5.99) leads to the elimination of the non-adiabatic mode in the long wavelength limit.

5.3.2 Parametric resonance revisited

Now it is time to look again at the results of Chapter 4. There I discussed the behaviour of cosmological perturbations in a two-field system, with potential

$$V(\varphi_A) = \frac{\lambda}{4}\varphi^4 + \frac{g^2}{2}\varphi^2\chi^2. \tag{5.117}$$

The field φ is assumed to dominate over χ during inflation. I defined conformally transformed fields by $\tilde{\varphi}_A \equiv a\varphi_A$ ($\varphi_1 \equiv \varphi$, $\varphi_2 \equiv \chi$) and showed that, during the oscillatory phase following inflation, the transformed inflaton $\tilde{\varphi}$ performs elliptic cosine oscillations with period T,

$$\tilde{\varphi}(x) = \tilde{\varphi}_0 \,\mathrm{cn}\left(x - x_0, \frac{1}{\sqrt{2}}\right), \tag{5.118}$$

where $x \equiv \sqrt{\lambda}\tilde{\varphi}_0\eta$ is a rescaled conformal time. I showed that the linear evolution equations for the perturbations $\delta\tilde{\varphi}_A$ had periodic coefficients, which meant that solutions had to be in the Floquet form, Eq. (4.47),

$$\delta\tilde{\varphi}_A(x) = f(x)e^{\mu x}, \tag{5.119}$$

where $f(x)$ is periodic with period T, and the Floquet index μ can be real or imaginary. I showed that the Floquet index for $\delta\tilde{\chi}_k$ was real and positive in certain bands in the parameter space defined by comoving wavenumber k and the ratio g^2/λ (recall Fig. 4.3). The inflaton perturbation $\delta\tilde{\varphi}_k$ was only weakly unstable at small scales (large k) in the case of vanishing background χ.

My primary interest in Chapter 4 was in the behaviour of modes in the extremely long wavelength limit, $k \to 0$. Thus it will be interesting now to consider the evolution of the homogeneous dynamical system specified by Eq. (5.117) and to try to understand how it relates to parametric resonance in the $k \to 0$ limit of the perturbation equations. The solution

$$(\tilde{\varphi}, \tilde{\chi}) = (\tilde{\varphi}_0 \,\mathrm{cn}(x), 0) \tag{5.120}$$

(where I have abbreviated the argument of the elliptic cosine) is a *periodic orbit* in the four-dimensional background scalar field phase space. Therefore the question of the stability of long wavelength cosmological perturbations is precisely equivalent to the question of the stability of this periodic orbit. When we linearize perturbations about the periodic orbit, we obtain a set of equations whose coefficients depend on the background and hence are periodic. The solutions must take the Floquet form, and they correspond to the long wavelength modes studied in Chapter 4. There will be an adiabatic time-translation mode and three modes orthogonal to the background trajectory. When the ratio g^2/λ is such that $\delta\tilde{\chi}_k$, for $k \to 0$, is unstable, then the periodic orbit is unstable to perturbations in the $\tilde{\chi}$ directions. Since $\delta\tilde{\varphi}_k$ is always stable for $k \to 0$ and $\chi = 0$, the periodic orbit is stable to all perturbations that lie in the $(\tilde{\varphi}, \tilde{\varphi}')$ plane. This situation is illustrated in Fig. 5.2.

To summarize, it is the behaviour of the dynamical system of background scalar fields that determines the evolution of long wavelength cosmological

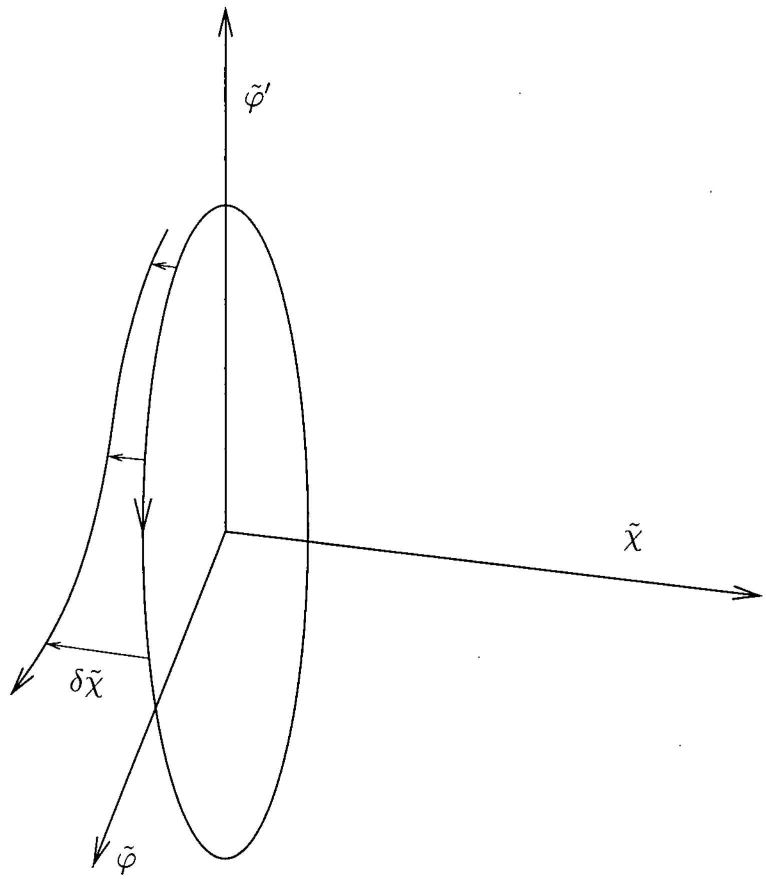

Figure 5.2: Periodic orbit in phase space of background fields $\tilde{\varphi}$ and $\tilde{\chi}$, with one dimension suppressed. The orbit lies in the $(\tilde{\varphi}, \tilde{\varphi}')$ plane. Perturbations orthogonal to that plane into some $\tilde{\chi}$ direction are unstable for appropriate g^2/λ, as illustrated by the short curve. This is the growth by parametric resonance described in Chapter 4. Perturbations in the plane of the orbit are always stable.

perturbations. During parametric resonance in the model (5.117), the background fields follow a periodic orbit in phase space, which is unstable in the $\tilde{\chi}$ directions and stable in the $\tilde{\varphi}$ directions. This gives us an easy to visualize picture of the evolution of long wavelength modes. Importantly, we can answer the question of whether long wavelength perturbations are unstable or not by examining only the background dynamical system. Of course, as I discussed in Chapter 4, the backreaction of perturbations on small scales can be important in eventually limiting the growth of the long wavelength modes. Thus we can use the background dynamics as an easy prefilter on models to determine if long wavelength modes are unstable. Once we have done so, we must always perform the much more difficult task of determining whether small scale perturbations limit the growth before it is significant.

It should be apparent that this model is very special in possessing an analytically describable periodic orbit in phase space. Actually, recall from Section 4.2.2 that I had to ignore a term of order $\varphi^2/m_{\rm P}^2$ in order to make analytical progress, though this was well justified. In addition I had to transform the fields and use a scaled conformal time coordinate. In more general cases, we must study the stability of arbitrary trajectories of the backgrounds. Fortunately much can be said about this from the theory of dynamical systems. A particular form of instability known as *dynamical chaos* will be my next topic.

5.3.3 Dynamical chaos

The qualitatively distinct types of behaviour possible for a two-dimensional dynamical system can be established exhaustively (see, *e.g.*, [69]). These include periodic orbits, fixed points, and *cycle graphs* (*e.g.* the trajectory of a particle in a potential well that has just enough energy to reach a local maximum of the potential). Each of these orbits can be stable or unstable to small perturbations. However, for orbits that are bounded in the phase space, a perturbation cannot grow without bound. (A trivial example to illustrate the caveat is the inverted oscillator, or potential hill, where clearly perturbations can grow without limit.)

The situation for dimension three or greater is far more complicated and the task of describing all qualitatively distinct types of behaviour is unsolved. However, it is known that a type of instability can exist which is not possible in two dimensions. For the general dynamical system

$$\dot{y}_i = F_i(y_j), \tag{5.121}$$

a sufficiently small perturbation from some background trajectory $\bar{y}(t)$ obeys the linearized equations

$$\delta\dot{y}_i = \left.\frac{\partial F_i}{\partial y_j}\right|_{\bar{y}} \delta y_j. \tag{5.122}$$

I will define the length of a perturbation δy_i by

$$|\delta y_i(t)| = \sqrt{\delta y_i \delta y^i}. \tag{5.123}$$

The Lyapunov exponent, h, for the system F_i and for initial conditions $\bar{y}(0)$ and $\delta y_i(0)$ is defined by

$$h(\bar{y}(0), \delta y_i(0)) = \lim_{t\to\infty} \frac{1}{t} \ln\left(\frac{|\delta y_i(t)|}{|\delta y_i(0)|}\right), \tag{5.124}$$

where the perturbation $\delta y_i(t)$ is evolved according to the *linearized* equations. Thus, the Lyapunov exponent is a logarithmic measure of the long time divergence rate of perturbations small enough to always evolve linearly. Had the *exact* perturbation dynamics been used to define the Lyapunov exponent instead of the linearized dynamics, then in a bounded system any perturbation is limited in size and hence h would always trivially vanish.

A dynamical system will exhibit a spectrum of Lyapunov exponents, for different directions of initial perturbation $\delta y_i(0)$. If, *for a bounded system* and for some background trajectory, at least one of the spectrum of exponents is positive, we call the system *chaotic.* (Again, the requirement of boundedness excludes trivially unstable systems.) In this case, it should be clear that generic perturbation initial conditions will contain some component along the direction $\delta y_i(0)$ corresponding to the largest Lyapunov exponent, and hence Eq. (5.124) will generically yield the largest Lyapunov exponent. Also, it is typically the case that the Lyapunov exponent will be independent of the choice of the background initial condition $\bar{y}(0)$, within some regions of the phase space [69]. Therefore I will henceforth drop both arguments on h and take it to represent the largest Lyapunov exponent for generic initial conditions within some chaotic region. For a background trajectory in such a chaotic region, and for a typical initial perturbation, we can then write

$$\lim_{t\to\infty} \frac{|\delta y_i(t)|}{|\delta y_i(0)|} = e^{ht}. \tag{5.125}$$

Necessary conditions for chaos in the dynamical system (5.121) are that it contain nonlinearities (clearly for an N-dimensional harmonic oscillator, $h = 0$), and, as already mentioned, that the dimensionality be three or more. In practice, chaos is a very common property of systems such as quartically coupled oscillators, of which many self-interacting multi-field inflationary models are examples. Recall our closed homogeneous background equations of motion for N scalar fields,

$$\ddot{\varphi}_A + 3H(\varphi_B, \dot{\varphi}_B)\dot{\varphi}_A + V_{,\varphi_A} = 0, \tag{5.126}$$

which I have pointed out can be written as a $2N$-dimensional dynamical system. Thus, for a system of $N = 2$ or more nonlinearly coupled homogeneous scalar fields, chaos is possible.

The importance of the presence of dynamical chaos in the background system of fields (5.126) is that, according to Eq. (5.125), for typical initial perturbations $(\delta\varphi_A, \delta\dot{\varphi}_A)$ those perturbations will grow exponentially in time with some Lyapunov exponent h. This means, as I discussed in Section 5.2.2, that typical physical homogeneous cosmological perturbations will grow exponentially as well. As I also explained, this statement applies in synchronous gauge when the background system is written for proper time (or, *e.g.*, static curvature gauge using integrated expansion N). However, remember that these modes are *physical* and cannot be gauged away. Furthermore, as I described in Section 5.2.4, these physical homogeneous modes can be straightforwardly "promoted" to realistic, long wavelength, inhomogeneous modes, which must also grow exponentially. Thus the presence of chaos in the background system directly implies that long wavelength perturbations will grow rapidly. Note the parallel with the argument for parametric resonance; only the nature of the unstable background orbit differs. Again, if some background system is found to be chaotic, the small scale behaviour must be examined carefully to determine if the large scale effect is important.

There have been a number of studies of chaos in systems of homogeneous fields in cosmology, although apparently none have made the connection with the growth of super-Hubble perturbations. Easther and Maeda [62] studied the chaotic dynamics of a two-field hybrid inflation system during reheating, although they did not include metric perturbations. They found two effects: the enhancement of defect formation, and a significant variation in the growth of the scale factor. However, they claim that only scales $k \sim aH$ at the time of reheating will be affected. Cornish and Levin [63] studied a single field model, and followed the evolution for several "cosmic cycles" of bang and crunch. Chaos is possible in such a simple system where $\dot{a}$ is no longer monotonic, since then we can no longer solve uniquely for H in terms of the scalar fields, and hence the full phase space is three-dimensional. Latora and Bazeia [64] studied a class of two-field quartically coupled systems which are chaotic in some regions of parameter space. See also [11, 12] for discussions of chaos in the context of reheating.

5.3.4 Hybrid inflation

I can illustrate the ideas of the preceeding subsection with a double scalar field model, which, as discussed above, is sufficiently complex for dynamical chaos to be possible. A popular class of two-field inflationary models is *hybrid* inflation [13, 71]. In these models inflation can be terminated by a symmetry breaking transition in one of the fields, and the subsequent oscillations can be chaotic [62]. I will consider the potential

$$V(\varphi, \chi) = \frac{1}{4\lambda}(M^2 - \lambda\chi^2)^2 + \frac{1}{2}m^2\varphi^2 + \frac{1}{2}g^2\varphi^2\chi^2. \qquad (5.127)$$

Inflation occurs at large φ, where the effective mass of the χ field, $m_\chi^2 = g^2\varphi^2 + \lambda\chi^2 - M^2$, is large and the χ field sits at the bottom of a potential valley at $\chi = 0$. Once the inflaton field drops below the critical value $\varphi_c = M/g$, the mass-squared m_χ^2 becomes negative (the potential valley becomes a ridge), and the fields undergo a symmetry breaking transition to one of the global minima at $\varphi = 0$, $\chi = \pm\chi_0$, where $\chi_0 = M/\sqrt{\lambda}$. I consider the "vacuum-dominated" regime, where the potential during the inflationary stage, $V(\varphi) = M^4/(4\lambda) + m^2\varphi^2/2$, is dominated by the false vacuum energy term $M^4/(4\lambda)$. I also consider the case where the Hubble parameter at the critical point,

$$H_0^2 = \frac{2\pi}{3\lambda}\frac{M^4}{m_{\mathrm{P}}^2}, \tag{5.128}$$

is much smaller than the oscillation frequencies about the global minima, which are $\overline{m}_\varphi = gM/\sqrt{\lambda}$ and $\overline{m}_\chi = \sqrt{2}M$ for small oscillations. This ensures that the fields will oscillate very many times after the critical point is reached before Hubble damping is significant.

Preheating has been studied in hybrid models for various parameter regimes in the absence of metric perturbations [14]. The behaviour of large-scale metric perturbations was studied in [42], where it was found that growth is possible on large scales. Oscillations in hybrid inflation were found to be chaotic in [62], although for a very different parameter regime than I examine here. Also, the connection with the growth of large-scale perturbations was not made in [62]. See also [15] for a discussion of chaotic dynamics in hybrid inflation.

Since my interest in this section is to establish a kinematical connection between dynamical chaos in the background fields and exponential growth of metric perturbations, I ignored the evolution of perturbations during the inflationary stage. It is important to note that for the parameters I consider, the large χ mass during inflation implies damping of large scale perturbations during inflation. Thus the question of whether the amplitude of metric perturbations produced in this model is consistent with the CMB normalization is not addressed here. A careful analysis, following the evolution of all important scales during inflation and preheating, and including the effects of backreaction, is required [58].

I considered the slice through parameter space specified by $M = 10^{-8}m_{\mathrm{P}}$, $m = 10^{-16}m_{\mathrm{P}}$, $\lambda = 10^{-3}$, and $g^2 = 10^{-2}$–10^{-4}. These parameters give an amplitude of cosmological density perturbations of the order 10^{-5} according to the standard inflationary calculation [71]. I evolved the homogeneous background fields according to Eqs. (5.126) with initial conditions $\varphi(t_0) = 0.999\varphi_c$ and $\chi(t_0) = 0.001\chi_0$. I followed the evolution of the comoving curvature perturbation ζ_k just as I did in Chapter 4, using the perturbed Klein-Gordon equation, Eq. (4.18), the momentum constraint, Eq. (4.17), and the expression (4.20) for ζ_k. I considered the super-Hubble value $k/a = 10^{-3}H_0$.

I calculated the largest Lyapunov exponent h for the background evolution using Eq. (5.125). The results, shown in Fig. 5.3, indicate the presence of rich

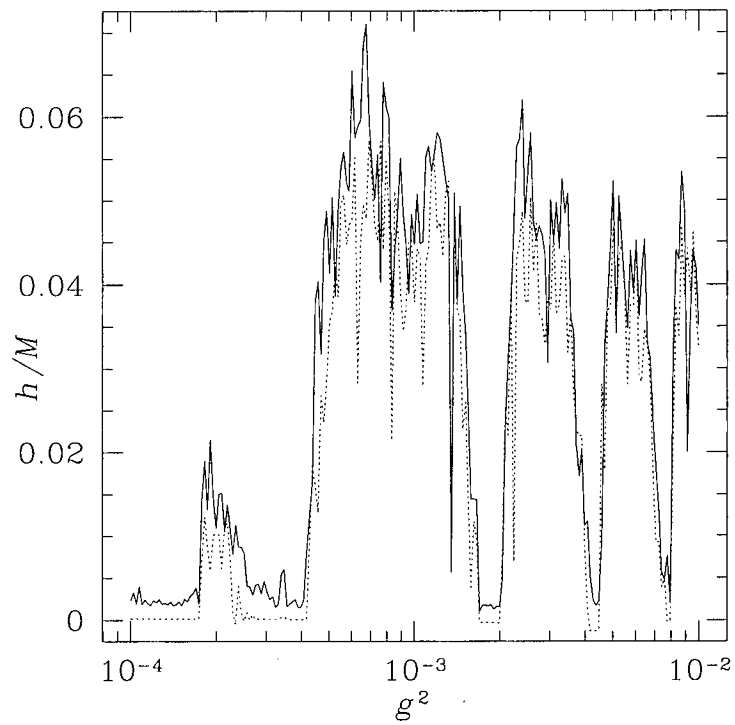

Figure 5.3: Largest Lyapunov exponent (solid line) for the homogeneous fields, and logarithmic growth rate of ζ_k (dotted line) for a scale $k/a = 10^{-3}H_0$. The homogeneous fields are oscillating about one of the hybrid model's global minima. The parameters are $M = 10^{-8}m_{\mathrm{P}}$, $m = 10^{-16}m_{\mathrm{P}}$, and $\lambda = 10^{-3}$. There is a clear correlation between the two curves.

structure as g^2 is varied. Regions of chaos with $h \simeq 0.05M$ are interspersed with regular stability bands where $h \simeq 0$. The most prominent stability band is near the supersymmetric point $g^2/\lambda = 2$. I also plot in Fig. 5.3 the logarithmic growth rate of ζ_k. The correlation between the Lyapunov exponent and the growth rate of large scale metric perturbations is strong numerical evidence in support of my arguments.

The growth at large scales is not simply due to the negative mass-squared (or "tachyonic") instability near the potential ridge at $\chi = 0$ [72]. To demonstrate this, I plot in Fig. 5.4 the logarithmic growth rate of ζ_k perturbations as a function of scale k. The solid line corresponds to the same parameters as in Fig. 5.3 (with $g^2 = 10^{-3}$), and shows a growth rate which is large at scales $k \simeq aM$ (due to the tachyonic instability), and approaches a constant (the chaotic Lyapunov exponent) as $k/(aH) \to 0$. The dotted line shows the

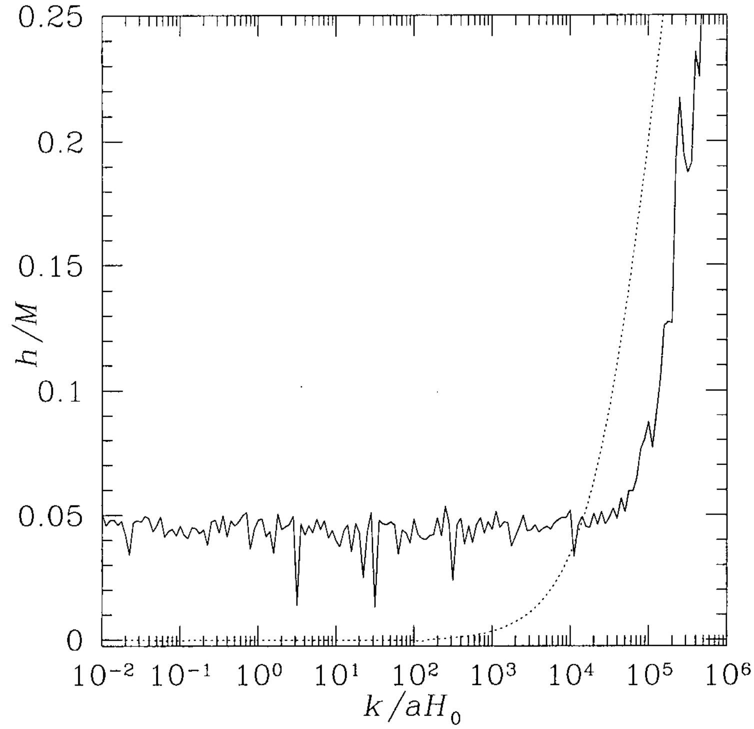

Figure 5.4: Logarithmic growth rate of ζ_k as a function of scale for the parameters of Fig. 5.3 and $g^2 = 10^{-3}$ (solid line) and $g^2 = 2 \times 10^{-3}$ (dotted line).

results for the same parameters except with $g^2 = 2\times 10^{-3}$. In this case, Fig. 5.3 indicates that the background trajectory is non-chaotic, and indeed in Fig. 5.4 the growth rate of ζ_k approaches zero at large scales. However, there is still strong growth at small scales for $g^2 = 2 \times 10^{-3}$, due again to the tachyonic instability. Note that both trajectories pass close to the critical point during their oscillations, and hence experience similar tachyonic effects. Thus, since only one exhibits growth on large scales, the tachyonic and chaotic effects are distinct.

To summarize, I have demonstrated a general link between instability in scalar field background evolution and the growth of super-Hubble metric perturbations. In particular, dynamical chaos in the fields can drive the growth—it is not necessary to have periodic motion and parametric resonance in the fields. Since chaos is common in multi-field systems, it is important to examine the super-Hubble evolution during preheating carefully, and also to follow the evolution during the inflationary period, in order to determine whether the model conflicts with CMB measurements.

Chapter 6

Conclusions

I have studied various aspects of the behaviour of long wavelength cosmological perturbations. I have presented a thorough description of the relevant theoretical background. Emphasis was made throughout on elucidating the physical meaning of the results, and on the important techniques and approximations that allow a tractable treatment of general relativistic perturbations.

I have examined the behaviour of super-Hubble modes in an inflationary model that exhibits parametric resonance. I performed a study of the growth of long wavelength perturbations in this model, and tried to determine whether the backreaction of small scale perturbations is sufficient to save the standard inflationary predictions based on the adiabatic conservation law. I concluded that, for certain parameter values, it is not.

I have also studied in considerable detail the connection between the behaviour of a homogeneous background dynamical system of scalar fields and the evolution of long wavelength cosmological perturbations. I showed how realistic, long wavelength modes can be constructed with only the knowledge of the background dynamical system. The presence of instabilities in that system implies the exponential growth of long wavelength metric perturbations.

As an example, it is the instability of a periodic orbit in the background phase space that is responsible for the growth of perturbations during parametric resonance as $k \to 0$. I found that dynamical chaos is a new route by which super-Hubble modes can be amplified during reheating. In this case, it is the instability of a background orbit that evolves in a much more general way than a simple periodic orbit which implies the amplification. I stress the importance of these results: dynamical chaos is expected to be common in multi-scalar-field models, and therefore any such model must be carefully examined to determine whether growth on large scales can be used to rule out the model. Of course, if a particular model *is* found to exhibit such growth, the behaviour must still be examined at small scales before a final conclusion is made.

There are many directions for future research. What other models exhibit resonant or chaotic growth of perturbations leading to a violation of the standard adiabatic predictions? What might a rigourous second order treatment of the perturbations reveal about the unbounded linear growth? What effects might even a limited period of growth, which does not violate observational constraints on the amplitude, have on the spectra?

Importantly, the behaviour of primordial cosmological modes will be increasingly subject to experimental test in the coming years. The satellite missions WMAP and later Planck will tightly constrain the spectral index of the cosmological power spectrum, and may detect a running index as well as a gravitational wave component in the CMB. These measurements are usually considered very important in constraining the inflationary model, and in particular the scalar field potential. Indeed the WMAP data already disfavour a $\lambda\varphi^4$ model [23]. However, the high precision of the experiments will require a correspondingly thorough understanding of the evolution of large scale modes after inflation. Indeed, the results to come may in fact provide a window on the post-inflationary physics.

Bibliography

[1] J. Traschen and R. Brandenberger, Phys. Rev. D **42**, 2491 (1990); A. Dolgov and D. Kirilova, Sov. J. Nucl. Phys. **51**, 172 (1990); L. Kofman, A. Linde, and A. Starobinsky, Phys. Rev. Lett. **73**, 3195 (1994).

[2] H. Kodama and T. Hamazaki, Prog. Theor. Phys. **96**, 949 (1996); Y. Nambu and A. Taruya, Prog. Theor. Phys. **97**, 83 (1997).

[3] M. Parry and R. Easther, Phys. Rev. D **59**, 061301 (1999); R. Easther and M. Parry, hep-ph/9910441.

[4] S. Tsujikawa, K. Maeda, and T. Torii, Phys. Rev. D **61**, 103501 (2000); S. Tsujikawa and B. A. Bassett, Phys. Rev. D **62**, 043510 (2000).

[5] B. A. Bassett and S. Liberati, Phys. Rev. D **58**, 021302 (1998); S. Tsujikawa, K. Maeda, and T. Torii, Phys. Rev. D **60**, 063515 (1999); S. Tsujikawa and H. Yajima, hep-ph/0007351.

[6] D. Boyanovsky, H. J. de Vega, and R. Holman, Phys. Rev. D **49**, 2769 (1994); S. Tsujikawa, K. Maeda, and T. Torii, Phys. Rev. D **61**, 103501 (2000); S.-J. Chang, Phys. Rev. D **12**, 1071 (1975).

[7] L. R. W. Abramo, R. H. Brandenberger, and V. F. Mukhanov, Phys. Rev. D **56**, 3248 (1997); V. F. Mukhanov, L. R. W. Abramo, and R. H. Brandenberger, Phys. Rev. Lett. **78**, 1624 (1997).

[8] S. Khlebnikov and I. Tkachev, Phys. Lett. B **390**, 80 (1997); S. Khlebnikov and I. Tkachev, Phys. Rev. Lett. **79**, 1607 (1997).

[9] J. Traschen and R. Brandenberger, Phys. Rev. D **42**, 2491 (1990); A. Dolgov and D. Kirilova, Sov. J. Nucl. Phys. **51**, 172 (1990); L. Kofman, A. Linde, and A. Starobinsky, Phys. Rev. Lett. **73**, 3195 (1994); L. Kofman, A. Linde, and A. Starobinsky, Phys. Rev. D **56**, 3258 (1997); Y. Shtanov, J. Traschen, and R. Brandenberger, Phys. Rev. D **51**, 5438 (1995).

[10] B. Bassett, D. Kaiser, and R. Maartens, Phys. Lett. B **455**, 84 (1999); F. Finelli and R. Brandenberger, Phys. Rev. Lett. **82**, 1362 (1999);

B. Bassett, F. Tamburini, D. Kaiser, and R. Maartens, Nucl. Phys. **B561**, 188 (1999); B. Bassett and F. Viniegra, Phys. Rev. D **62**, 043507 (2000); F. Finelli and S. Khlebnikov, Phys. Lett. B **504**, 309 (2001).

[11] B. Bassett and F. Tamburini, Phys. Rev. Lett. **81**, 2630 (1998).

[12] D. Podolsky and A. Starobinsky, Grav. Cosmol. Suppl. **8N1**, 13 (2002).

[13] A. D. Linde, Phys. Lett. B **259**, 38 (1991); E. J. Copeland, A. R. Liddle, D. H. Lyth, E. D. Stewart, and D. Wands, Phys. Rev. D **49**, 6410 (1994).

[14] J. García-Bellido and A. Linde, Phys. Rev. D **57**, 6075 (1998); M. Bastero-Gil, S. F. King, and J. Sanderson, Phys. Rev. D **60**, 103517 (1999).

[15] R. Ramos, Phys. Rev. D **64**, 123510 (2001).

[16] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (W. H. Freeman and Company, New York, 1973).

[17] R. M. Wald, *General Relativity* (University of Chicago Press, Chicago, 1984).

[18] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Pergamon Press, Oxford, 1975).

[19] S. Weinberg, *Gravitation and Cosmology* (John Wiley and Sons, New York, 1972).

[20] E. W. Kolb and M. S. Turner, *The Early Universe* (Addison-Wesley, Reading, Mass., 1990).

[21] A. Linde, *Particle Physics and Inflationary Cosmology* (Harwood, Chur, 1990).

[22] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, Cambridge, 1982).

[23] D. N. Spergel *et al.*, astro-ph/0302209.

[24] D. H. Lyth and A. Riotto, Phys. Rep. **314**, 1 (1999).

[25] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger, Phys. Rep. **215**, 203 (1992).

[26] H. Kodama and M. Sasaki, Prog. Theor. Phys. Suppl. **No. 78**, 1 (1984).

[27] J. M. Stewart, Class. Quant. Grav. **7**, 1169 (1990).

[28] J. M. Bardeen, Phys. Rev. D **22**, 1882 (1980).

[29] W. Unruh, astro-ph/9802323.

[30] K. A. Malik and D. Wands, gr-qc/9804046.

[31] L. Kofman, A. Linde, and A. Starobinsky, Phys. Rev. D **56**, 3258 (1997).

[32] Y. Shtanov, J. Traschen, and R. Brandenberger, Phys. Rev. D **51**, 5438 (1995).

[33] B. Bassett, D. Kaiser, and R. Maartens, Phys. Lett. B **455**, 84 (1999).

[34] B. Bassett, F. Tamburini, D. Kaiser, and R. Maartens, Nucl. Phys. **B561**, 188 (1999).

[35] F. Finelli and R. Brandenberger, Phys. Rev. Lett. **82**, 1362 (1999).

[36] W. Lin, X. Meng, and X. Zhang, Phys. Rev. D **61**, 121301 (2000).

[37] A. Taruya and Y. Nambu, Phys. Lett. B **428**, 37 (1998).

[38] P. Ivanov, Phys. Rev. D **61**, 023505 (1999).

[39] K. Jedamzik and G. Sigl, Phys. Rev. D **61**, 023519 (1999).

[40] A. R. Liddle, D. H. Lyth, K. A. Malik, and D. Wands, hep-ph/9912473.

[41] B. Bassett, C. Gordon, R. Maartens, and D. Kaiser, Phys. Rev. D **61**, 061302 (2000).

[42] F. Finelli and R. Brandenberger, Phys. Rev. D **62**, 083502 (2000).

[43] B. Bassett and F. Viniegra, Phys. Rev. D **62**, 043507 (2000).

[44] P. Greene, L. Kofman, A. Linde, and A. Starobinsky, Phys. Rev. D **56**, 6175 (1997).

[45] S. Tsujikawa, B. A. Bassett, and F. Viniegra, hep-ph/0006354.

[46] E. F. Bunn, A. R. Liddle, and M. White, Phys. Rev. D **54**, 5917 (1996).

[47] M. Bastero-Gil, S. F. King, and J. Sanderson, Phys. Rev. D **60**, 103517 (1999).

[48] P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Scientists* (Springer-Verlag, Berlin, 1971).

[49] L. D. Landau and E. M. Lifshitz, *Mechanics* (Pergamon Press, Oxford, 1976).

[50] A. Erdelyi, *Higher Transcendental Functions Vol. 3* (McGraw-Hill, New York, 1955).

[51] A. R. Liddle and D. H. Lyth, Phys. Rep. **231**, 1 (1993).

[52] T. Prokopec and T. Roos, Phys. Rev. D **55**, 3768 (1997).

[53] V. Sahni and S. Habib, Phys. Rev. Lett. **81**, 1766 (1998).

[54] R. Allahverdi and B. A. Campbell, Phys. Lett. B **395**, 169 (1997).

[55] B. R. Greene, T. Prokopec, and T. G. Roos, Phys. Rev. D **56**, 6484 (1997).

[56] S. Khlebnikov and I. Tkachev, Phys. Rev. Lett. **77**, 219 (1996).

[57] J. P. Zibin, D. Scott, and M. White, Phys. Rev. D **60**, 123513 (1999).

[58] J. P. Zibin, R. Brandenberger, and D. Scott, Phys. Rev. D **63**, 043511 (2001).

[59] J. P. Zibin, hep-ph/0108008.

[60] D. Wands, K. A. Malik, D. H. Lyth, and A. R. Liddle, Phys. Rev. D **62**, 043527 (2000).

[61] C. Gordon, D. Wands, B. A. Bassett, and R. Maartens, Phys. Rev. D **63**, 023506 (2001).

[62] R. Easther and K. Maeda, Class. Quant. Grav. **16**, 1637 (1999).

[63] N. J. Cornish and J. J. Levin, Phys. Rev. D **53**, 3022 (1996).

[64] V. Latora and D. Bazeia, Int. J. Mod. Phys. **A14**, 4967 (1999).

[65] J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, Phys. Rev. D **28**, 679 (1983).

[66] D. H. Lyth and D. Wands, Phys. Rev. D **68**, 103515 (2003).

[67] H. Kodama and T. Hamazaki, Phys. Rev. D **57**, 7177 (1998).

[68] S. Weinberg, Phys. Rev. D **67**, 123504 (2003).

[69] E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, Cambridge, 1993).

[70] M. Sasaki and T. Tanaka, Prog. Theor. Phys. **99**, 763 (1998).

[71] A. D. Linde, Phys. Rev. D **49**, 748 (1994).

[72] G. Felder, L. Kofman, and A. Linde, hep-th/0106179.

[73] A. A. Starobinskii, Soviet Astr. Lett. **11**, 133 (1985).

[74] L. M. Krauss and M. White, Phys. Rev. Lett. **69**, 869 (1992).

[75] R. L. Davis *et al.*, Phys. Rev. Lett. **69**, 1856 (1992) [erratum **70**, 1733].

[76] D. H. Lyth and A. R. Liddle, Phys. Lett. B **291**, 391 (1992).

[77] T. Souradeep and V. Sahni, Mod. Phys. Lett. **A7**, 3541 (1992).

[78] D. S. Salopek, Phys. Rev. Lett. **69**, 3602 (1992).

[79] A. V. Markevich and A. A. Starobinsky, Astron. Lett. **22**, 431 (1996).

[80] M. Tegmark, Astrophys. J. **514**, L69 (1999).

[81] J. Lesgourgues, D. Polarski, and A. A. Starobinsky, astro-ph/9807019.

[82] A. Melchiorri, M. V. Sazhin, V. V. Shulga, and N. Vittorio, astro-ph/9901220.

[83] A. R. Liddle and D. H. Lyth, Phys. Rep. **231**, 1 (1993); J. E. Lidsey, *et al.*, Rev. Mod. Phys. **69**, 373 (1997); D. H. Lyth and A. Riotto, Phys. Rep. **314**, 1 (1999).

[84] M. S. Turner and M. White, Phys. Rev. D **53**, 6822 (1996).

[85] K. Coble, S. Dodelson, and J. A. Frieman, Phys. Rev. D **55**, 1851 (1997).

[86] R. R. Caldwell and P. J. Steinhardt, Phys. Rev. D **57**, 6057 (1998).

[87] A. D. Linde, Phys. Lett. B **129**, 177 (1983); A. D. Linde, *Particle Physics and Inflationary Cosmology* (Harwood Academic, Chur, Switzerland, 1990).

[88] A. A. Starobinsky and J. Yokoyama, gr-qc/9502002; J. García-Bellido and D. Wands, Phys. Rev. D **52**, 6739 (1995); M. Sasaki and E. D. Stewart, Prog. Theor. Phys. **95**, 71 (1996); J. García-Bellido and D. Wands, Phys. Rev. D **53**, 5437 (1996); T. T. Nakamura and E. D. Stewart, Phys. Lett. B **381**, 413 (1996).

[89] A. R. Liddle, in *From Quantum Fluctuations to Cosmological Structures: Proceedings of the First Moroccan School of Astrophysics*, edited by D. Valls-Gabaud *et al.*, Astronomical Society of the Pacific Conference Series **126**, 31 (1997).

[90] D. La and P. J. Steinhardt, Phys. Rev. Lett. **62**, 376 (1989).

[91] E. W. Kolb, Physica Scripta **T36**, 199 (1991).

[92] B. Whitt, Phys. Lett. B **145**, 176 (1984); K. Maeda, Phys. Rev. D **39**, 3159 (1989); D. Wands, Class. Quant. Grav. **11**, 269 (1994).

[93] S. Dodelson, W. H. Kinney, and E. W. Kolb, Phys. Rev. D **56**, 3207 (1997).

[94] D. H. Lyth, Phys. Rev. Lett. **78**, 1861 (1997); D. H. Lyth and A. Riotto, Phys. Rep. **314**, 1 (1999).

[95] L. F. Abbott and M. B. Wise, Nucl. Phys. **B244**, 541 (1984).

[96] F. Lucchin and S. Matarrese, Phys. Rev. D **32**, 1316 (1985).

[97] A. Linde, Phys. Lett. B **129**, 177 (1983).

[98] G. F. Smoot and D. Scott in the Review of Particle Properties, C. Caso *et al.*, The European Physical Journal **C3**, 127 (1998).

[99] J. R. Bond, A. H. Jaffe, and L. E. Knox, Phys. Rev. D **57**, 2117 (1998).

[100] J. R. Bond, A. H. Jaffe, and L. E. Knox, astro-ph/9808264.

[101] M. White, Phys. Rev. D **53**, 3011 (1996).

[102] U. Seljak and M. Zaldarriaga, Astrophys. J. **469**, 437 (1996).

[103] E. F. Bunn and M. White, Astrophys. J. **480**, 6 (1997).

[104] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, Astrophys. J. **304**, 15 (1986).

[105] N. Sugiyama, Astrophys. J. Supp. **100**, 281 (1995).

[106] J. A. Peacock and S. J. Dodds, Mon. Not. R. Astron. Soc. **267**, 1020 (1994).

[107] R. J. Scherrer and D. H. Weinberg, Astrophys. J. **504**, 607 (1998).

[108] J. A. Peacock, Mon. Not. R. Astron. Soc. **284**, 885 (1997).

[109] P. T. P. Viana and A. R. Liddle, Mon. Not. R. Astron. Soc. **303**, 535 (1999).

[110] A. R. Liddle, D. H. Lyth, P. T. P. Viana, and M. White, Mon. Not. R. Astron. Soc. **282**, 281 (1996).

[111] E. F. Bunn, A. R. Liddle, and M. White, Phys. Rev. D **54**, 5917 (1996).

[112] R. A. C. Croft, D. H. Weinberg, M. Pettini, L. Hernquist, and N. Katz, astro-ph/9809401.

[113] S. M. Carroll, W. H. Press, and E. L. Turner, Annu. Rev. Astron. Astrophys. **30**, 499 (1992).

[114] D. A. Vandenberg, P. B. Stetson, and M. Bolte, Annu. Rev. Astron. Astrophys. **34**, 461 (1996).

[115] B. Chaboyer, Phys. Rep. **307**, 23 (1998).

[116] D. A. White and A. Fabian, Mon. Not. R. Astron. Soc. **273**, 72 (1995).

[117] A. V. Filippenko and A. G. Riess, Phys. Rep. **307**, 31 (1998).

[118] J. R. Gott, Nature **295**, 304 (1982); M. Sasaki, T. Tanaka, K. Yamamoto, and J. Yokoyama, Phys. Lett. B **317**, 510 (1993); M. Bucher, A. S. Goldhaber, and N. Turok, Phys. Rev. D **52**, 3314 (1995); A. D. Linde and A. Mezhlumian, Phys. Rev. D **52**, 6789 (1995).

[119] M. Tanaka and M. Sasaki, Prog. Theor. Phys. **97**, 243 (1997).

[120] M. Bucher and J. D. Cohn, Phys. Rev. D. **55**, 7461 (1997).

[121] M. White and J. Silk, Phys. Rev. Lett. **77**, 4704 (1996) [erratum **78**, 3799].

[122] M. White and D. Scott, Comments Astrophys. **18**, 289 (1996).

[123] W. Hu and M. White, Astrophys. J. **486**, L1 (1997).

[124] M. White, Phys. Rev. D **46**, 4198 (1992).

[125] M. S. Turner, M. White, and J. E. Lidsey, Phys. Rev. D **48**, 4613 (1993).

[126] M. S. Turner and M. White, Phys. Rev. D **53**, 6822 (1996).

[127] A. R. Liddle, Phys. Rev. D **49**, 3805 (1994).

[128] M. S. Turner, Phys. Rev. D **55**, R435 (1997).

[129] R. R. Caldwell, M. Kamionkowski, and L. Wadley, Phys. Rev. D **59**, 027101 (1999).

[130] K. S. Thorne, in *Particle and Nuclear Astrophysics and Cosmology in the Next Millennium, Snowmass 94*, edited by E. W. Kolb and R. D. Peccei (World Scientific, Singapore, 1995), p. 398.

[131] K. Danzmann, Ann. N. Y. Acad. Sci. **759**, 481 (1995).

[132] V. M. Kaspi, J. H. Taylor, and M. F. Ryba, Astrophys. J. **428**, 713 (1994).

[133] L. Knox and M. S. Turner, Phys. Rev. Lett. **73**, 3347 (1994); L. Knox, Phys. Rev. D **52**, 4307 (1995).

[134] M. Zaldarriaga, D. N. Spergel, and U. Seljak, Astrophys. J. **488**, 1 (1997).

[135] J. R. Bond, G. Efstathiou, and M. Tegmark, Mon. Not. R. Astron. Soc. **291**, L33 (1997).

[136] W. Hu and M. White, New Astron. **2**, 323 (1997).

[137] M. Zaldarriaga and U. Seljak, Phys. Rev. D **55**, 1830 (1997); M. Zaldarriaga and U. Seljak, Phys. Rev. D **58**, 023003 (1998).

[138] M. Kamionkowski and A. Kosowsky, Phys. Rev. D **57**, 685 (1998).

[139] W. H. Kinney, Phys. Rev. D **58**, 123506 (1998).

[140] M. Kamionkowski, A. Kosowsky, and A. Stebbins, Phys. Rev. D **55**, 7368 (1997).

[141] W. Hu and M. White, Phys. Rev. D **56**, 596 (1997).

Appendix A

Limits on the Tensor Contribution to Microwave Anisotropies

Here I attach earlier work in collaboration with Martin White which lies somewhat outside the scope of the main part of the thesis. This work was published in [57]. While the content of this appendix is perhaps better suited to an entire chapter, it does not fit well into the flow of the main part of this thesis and would require considerable background material. Thus I have decided to include a concise description of the work here.[1]

A.1 Introduction

The presence of a primordial gravitational wave perturbation spectrum was an early prediction of inflationary models of the big bang [73]. However, it was not until the results of the *COBE* satellite mission that it became possible to begin to meaningfully constrain the tensor contribution to the overall perturbation spectrum [74–78]. In an early result, Salopek [78] found that, assuming power-law inflation, tensors must contribute less than about 50% of the cosmic microwave background (CMB) fluctuations at the 10° scale.

Since that time, ground- and balloon-based experiments have begun to fill in the smaller-scale regions of the CMB power spectrum. These scales are crucial for constraining the gravity wave contribution because the tensor spectrum is expected to be negligible on scales finer than $\sim 1°$, and therefore large-scale power greater than that expected for scalars can be attributed to tensors. Markevich and Starobinsky [79] have set some stringent limits on the tensor contribution. For example, they found that the ratio of tensor to scalar components of the CMB spectrum is $T/S < 0.7$ at 97.5% confidence for a flat, cosmological-constant-free universe with $H_0 = 50\,\mathrm{kms^{-1}Mpc^{-1}}$. However, their analysis used a limited CMB data set and considered only a restricted

[1] A version of this appendix has been published. Zibin, J. P., Scott, D., and White, M. (1999) Limits on the gravity wave contribution to microwave anisotropies, Phys. Rev. D **60**: 123513.

set of values for the cosmological parameters. Recently Tegmark [80] has performed an analysis using a compilation of CMB data, and found a 68% upper confidence limit of 0.56 on the tensor to scalar ratio. In this work specific inflationary models were not considered, but a number of parameters were allowed to vary freely. In another recent study Lesgourgues *et al.* [81] analysed a particular broken-scale-invariance model of inflation with a steplike primordial perturbation spectrum, and found that the tensor to scalar ratio can reach unity. Melchiorri *et al.* [82] placed limits on tensors allowing for a blue scalar spectral index, and indeed found that blue spectra and a large tensor component are most consistent with CMB observations.

Our aim here is to provide a more comprehensive answer (or set of answers) to the question: how big can T/S be? We present constraints on tensors for specific models of inflation as well as for freely varying parameters. In all cases we marginalize over the important, but as yet undetermined, cosmological parameters. We use both *COBE* and small-scale CMB data, as well as information about the matter power spectrum from galaxy correlation, cluster abundance, and Lyman α forest measurements. We refer to these various measurements of the power spectra as "data sets". We additionally consider the effect, for each data set, of various observational constraints on the cosmological parameters, such as the age of the universe, cluster baryon density, and recent supernova measurements. We refer to these constraints as "parameter constraints" (this separation between "data sets" and "parameter constraints" is somewhat subjective, but dealt with consistently in our Bayesian approach; it is conceptually simpler to consider power spectrum constraints as measurements with some Gaussian error, while regarding allowed limits on cosmological parameters as restrictions on parameter space). Finally we consider what implications our results have for the direct detection of primordial gravity waves.

A.2 Inflation models

Our goal is to provide limits on the tensor contribution to the primordial perturbation spectra using a variety of recent observations. In models of inflation, the scalar (density) and tensor (gravity wave) metric perturbations produced during inflation are specified by two spectral functions, $A_{\rm S}(k)$ and $A_{\rm T}(k)$, for wave number k. These spectra are determined by the inflaton potential $V(\phi)$ and its derivatives [83]. However, when comparing model predictions with actual observations of the CMB, it is more useful to translate the inflationary spectra into the predicted multipole expansions of the CMB temperature field: $\Delta T/T(\theta,\phi) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\theta,\phi)$, where $Y_{\ell m}(\theta,\phi)$ are the spherical harmonics.

The spectrum $C_\ell \equiv \langle |a_{\ell m}|^2 \rangle$ can be decomposed into scalar and tensor parts, $C_\ell = C_\ell^{\mathrm{S}} + C_\ell^{\mathrm{T}}$. In the literature the tensor to scalar ratio is conventionally specified either at $\ell = 2$ or in the spectral plateau at $\ell \simeq 10 - 20$. Here we have chosen the $\ell = 2$ or quadrupole moments of the temperature field, and write $S \equiv 5C_2^{\mathrm{S}}/4\pi$ and $T \equiv 5C_2^{\mathrm{T}}/4\pi$ as usual [84].

In order to constrain the tensor contribution T/S, we need to specify the particular model of inflation under consideration. This is because the model may provide a specific relationship between the ratio T/S and the scalar spectral index n_{S}. Except in Sec. A.6 we only consider spatially flat inflation models (*i.e.* $\Omega_0 + \Omega_\Lambda = 1$, where Ω_0 and $\Omega_\Lambda = \Lambda/(3H_0^2)$ are the fractions of critical density due to matter and a cosmological constant, respectively). In addition, we do not consider the "quintessence" models [85, 86], where a significant fraction of the critical density is currently in the form of a scalar field with equation-of-state different from that of matter, radiation, or cosmological constant (although it would not be difficult to extend our results for explicit models with recent epoch dynamical fields). We will also restrict ourselves to models which use the slow-roll approximation, and incorporate only a single *dynamical* field – a class of models sometimes called "chaotic inflation" [87]. This is not as restrictive as it might sound, since most viable inflationary models are of this form. Although some genuinely two-field models are known [88], many multi-field models, the "hybrid" class, have only one field dynamically important and in these cases we effectively regain the single field case [89]. In addition, theories which modify general relativity (*e.g.* "extended" inflation [90, 91]) can often be recast as ordinary general relativity with a single effective scalar field [89, 92].

It is often convenient to classify inflationary models as either "small-field", "large-field", or the already mentioned hybrid models [93]. Small-field models are characterized by an inflaton field which rolls from a potential maximum towards a minimum at $\langle \phi \rangle \neq 0$. These models generally produce negligible tensor contribution, but may result in the spectral index n_{S} differing significantly from scale invariance [83]. In hybrid models, the important scalar field rolls towards a potential minimum with non-zero vacuum energy. These models also typically have very small T/S, and the scalar index can be greater than unity [83]. The large-field models involve so-called "chaotic" initial conditions, where an inflaton initially displaced from the potential minimum rolls towards the origin. Large-field models can produce large T/S and $1 - n_{\mathrm{S}}$, and these are the models considered in this paper. This is not to say that small-field and hybrid models are not interesting; on the contrary, current views of inflation in the particle physics context suggest that T/S is expected to be small [94]. However, large-field models must be considered when examining the observational evidence for a large tensor contribution.

It is also worth pointing out that we could construct models with a dip in the scalar power spectrum at large scales which compensates for the tensor contribution. Although we have not explored detailed models, we imagine that in principle models could be constructed with arbitrarily high T/S. We consider all such models with features at relevant scales to be unappealing unless there are separate physical arguments for them.

In addition to considering models with free scalar index and tensor contribution T/S, we shall thus focus on two classes of inflationary models which can be considered representative of those predicting large gravity wave contributions. Both are restricted to "red" spectral tilts, $n_S \leq 1$. The first, "power-law inflation" (PLI) [95, 96], is characterized by exponential inflaton potentials of the form

$$V(\phi) \propto \exp\sqrt{\frac{16\pi\phi^2}{qm_{\rm Pl}^2}}, \tag{A.1}$$

and results in a scale factor growth $a(t) \propto t^q$, hence the name. For PLI the tensor-to-scalar ratio in k-space can be calculated exactly as a function of n_S:

$$\frac{A_T^2(k)}{A_S^2(k)} = \frac{1-n_S}{3-n_S} \quad . \tag{A.2}$$

Note the tensor contribution is directly related to the scalar spectral index n_S, which is further related to the tensor spectral index $n_T = n_S - 1$ in this model. Converting from k-space to the observed anisotropy spectrum introduces a dependence on the cosmological constant which can be approximated by [84]

$$T/S = -7\widetilde{n}\left[0.97 + 0.58\widetilde{n} + 0.25\Omega_\Lambda - \left(1 + 1.1\widetilde{n} + 0.28\widetilde{n}^2\right)\Omega_\Lambda^2\right], \tag{A.3}$$

where $\widetilde{n} \equiv n_S - 1 = n_T$. The dependence on Λ arises because of different evolution for scalars and tensors when Λ dominates at late times. The dependence on other cosmological parameters is negligible [84].

We also consider the large-field polynomial potentials,

$$V(\phi) \propto \phi^p, \tag{A.4}$$

for integral $p > 1$ [97]. In this case both n_S and n_T are determined by the exponent p [83]:

$$n_S = 1 - \frac{2p+4}{p+200}, \tag{A.5}$$

$$n_T = \frac{-2p}{p+200}. \tag{A.6}$$

The tensor index may be related to T/S through the consistency relation [84]

$$\frac{T}{S} = -7\frac{f_T^{(0)}}{f_S^{(0)}}n_T, \tag{A.7}$$

where the cosmological parameter dependence, again dominated by Ω_Λ, can be approximated by

$$f_{\rm S}^{(0)} = 1.04 - 0.82\Omega_\Lambda + 2\Omega_\Lambda^2, \tag{A.8}$$
$$f_{\rm T}^{(0)} = 1.0 - 0.03\Omega_\Lambda - 0.1\Omega_\Lambda^2. \tag{A.9}$$

As a third possibility, we will also consider models with scalar index varying over the range $n_{\rm S} = 0.8 - 1.2$, but with an independently varying tensor contribution T/S.

A.3 Microwave background anisotropies

In order to evaluate likelihoods and confidence limits for T/S based on CMB measurements, we performed χ^2 fits of model C_ℓ spectra to CMB data. We did this for a set of "band-power" estimates of anisotropy at different scales, and separately for the *COBE* data themselves. For our first approach, we used a collection of binned data to represent the anisotropies as a function of ℓ. Specifically we took the flat-spectrum effective quadrupole values listed in Smoot and Scott [98] and binned them into nine intervals separated logarithmically in ℓ. We chose this simplified approach since we anticipated a large computational effort in covering a reasonably large parameter space. The use of binned data has been shown elsewhere [99] to give similar results to more thorough methods. If anything, there is a bias towards lowering the height of any acoustic peak, inherent in the simplifying assumption of symmetric Gaussian error bars [100]; for placing upper limits on T/S our approach is therefore conservative. We are also erring on the side of caution by using the binned data only up to the first acoustic peak, neglecting constraints from detections and upper limits at smaller angular scales.

We ignored the effect of reionization on the C_ℓ spectra. Reionization to optical depth τ reduces the power of small-scale anisotropies by $e^{-2\tau}$. Thus, in placing upper limits on T/S, it is conservative to set $\tau = 0$.

A fitting function for the spectrum, valid up to the first peak at $l \simeq 220$, has been provided by White [101]:

$$C_l(\nu) = \left(\frac{l}{10}\right)^\nu C_l(\nu = 0), \tag{A.10}$$

where ν is the (nearly) degenerate combination of cosmological parameters

$$\nu \equiv n_{\rm S} - 1 - 0.32\ln(1+0.76r) + 6.8(\Omega_{\rm B}h^2 - 0.0125) - 0.37\ln(2h) - 0.16\ln(\Omega_0). \tag{A.11}$$

Here $r \equiv 1.4C_{10}^{\mathrm{T}}/C_{10}^{\mathrm{S}}$ is the tensor to scalar ratio at $\ell = 10$, normalized to provide $r = T/S$ for $\Omega_0 = 1$ and $n_{\mathrm{S}} \to 1$. The parameter h is defined through $H_0 = 100h\ \mathrm{kms^{-1}Mpc^{-1}}$, and Ω_{B} is the fraction of the critical density in baryons. Thus the standard CDM (sCDM) spectrum is specified by $\nu = 0$. We found that the Ω_Λ dependence of r can be well captured by introducing the rescaled variable r', defined by

$$r' = \frac{r}{0.94 + 1.105\Omega_\Lambda^{3.75}}, \tag{A.12}$$

and setting $r' \equiv T/S$.

We fitted the model spectra of Eq. (A.10) to the binned data as follows. For each combination of parameters $(h, \Omega_{\mathrm{B}}h^2, \Omega_0, n_{\mathrm{S}}, T/S)$ we normalized the model spectrum to the binned data, and evaluated the likelihood $\mathcal{L}(h, \Omega_{\mathrm{B}}h^2, \Omega_0, n_S, T/S) \propto \exp(-\chi^2/2)$. Next this likelihood was integrated, uniformly in the parameter, over the ranges of $h = 0.5 - 0.8$, $\Omega_{\mathrm{B}}h^2 = 0.007 - 0.024$, and $\Omega_0 = 0.25 - 1$, subject to the constraints of Eq. (A.3) for PLI and Eqs. (A.5), (A.6), and (A.7) for polynomial potentials. For the case of free T/S, the scalar index was varied in the range $n_{\mathrm{S}} = 0.8 - 1.2$. Finally the resultant $\mathcal{L}(T/S)$ was normalized to a peak value of unity and the 95% confidence limits evaluated. We tried to choose reasonable ranges for the prior probability distributions of the "nuisance parameters", guided by the current weight of evidence. We checked that mild departures from our adopted ranges lead to only small modifications to our results. However, we caution that our conclusions will not necessarily be applicable for models which lie significantly outside the parameter space we considered. In addition, note that according to Eq. (A.11) we can crudely estimate an upper limit on T/S by combining the observational lower limit on ν with the maximal baryon density and minimal Ω_0 and h from our parameter ranges. However, this turns out to be an overly conservative estimate: for example, for $n_{\mathrm{S}} = 1$, and using a lower limit of $\nu = -0.2$, Eq. (A.11) gives an upper limit of $T/S = 3.8$, compared with the limit $T/S = 1.6$ from Sec. A.7.

For the separate constraint from the *COBE* data, we used the software package CMBFAST [102] to calculate likelihoods based only on the *COBE* results at large scales. CMBFAST calculates the spectrum using a line-of-sight integration technique. It then calculates likelihoods by finding a quadratic approximation to the large scale spectrum and using the *COBE* fits of Bunn and White [103]. These likelihoods were integrated and 95% limits calculated as above, except that the baryon density was fixed at $\Omega_{\mathrm{B}} = 0.05$ to save computation time (and since Ω_{B} has negligible effect at these scales). The results of this procedure are presented in Sec. A.7.

A.4 Large-scale structure

A.4.1 Galaxy correlations

We next applied observations of galaxy correlations to constrain T/S indirectly through the power spectrum of the density fluctuations, $\Delta^2(k)$. The power spectrum $\Delta^2(k)$ is expressed, following Bunn and White [103], by

$$\Delta^2(k) = \delta_{\rm H}^2 \left(\frac{ck}{H_0}\right)^{3+n_{\rm S}} T^2(k). \tag{A.13}$$

Here $\delta_{\rm H}$ is the (Ω_0, $n_{\rm S}$, and r dependent) normalization described in Sec. A.4.2, and $T(k)$ is the transfer function which describes the evolution of the spectrum from its primordial form to the present.

We explicitly used for the transfer function the fit of Bardeen *et al.* [104],

$$T(q) = \frac{\ln(1+2.34q)}{2.34q}\left[1+3.89q+(16.1q)^2+(5.46q)^3+(6.71q)^4\right]^{-1/4}, \tag{A.14}$$

with the scaling of Sugiyama [105]

$$q = \frac{k(T_{\gamma 0}/2.7\,{\rm K})^2}{\Omega_0 h^2 \exp\left(-\Omega_{\rm B} - \sqrt{h/0.5}\,\Omega_{\rm B}/\Omega_0\right)}. \tag{A.15}$$

Here $T_{\gamma 0}$ is the temperature of the CMB radiation today.

We performed χ^2 fits of the (unnormalized) model power spectrum given by Eqs. (A.13) - (A.15) to the compilation of data provided in Table I of Peacock and Dodds [106], excluding their four smallest scale data points. These points were omitted because, while there are theoretical reasons [107] to expect that the galaxy bias approaches a constant on large scales, at the smallest scales the assumption of a linear bias appears to break down [108]. Here we are fitting for the *shape* of the matter power spectrum, ignoring the overall amplitude, since the normalization is complicated by the ambiguities of galaxy biasing.

The fitting was performed in exactly the same way as was described in Sec. A.3 for the binned microwave anisotropies. Namely the model curves were normalized to the Peacock and Dodds data, the integrated likelihood was calculated, and the 95% confidence limits for $n_{\rm S}$ were evaluated. Since the shape of the power spectrum [Eq. (A.15)] is independent of the tensor amplitude, this technique can only provide limits on T/S when T/S is determined by the spectral index. That is, the galaxy correlation data can only constrain T/S for our PLI and ϕ^p cases, using the relationships [Eq. (A.3) or Eqs. (A.5), (A.6), and (A.7)] between T/S and $n_{\rm S}$.

A.4.2 Cluster abundance

A very useful quantity for constraining the *amplitude* of the power spectrum is the dispersion of the density field smoothed on a scale R, defined by

$$\sigma^2(R) = \int_0^\infty W^2(kR)\Delta^2(k)\frac{dk}{k}. \tag{A.16}$$

Here $W(kR)$ is the smoothing function, which we take to be a spherical top-hat specified by

$$W(kR) = 3\left[\frac{\sin(kR)}{(kR)^3} - \frac{\cos(kR)}{(kR)^2}\right]. \tag{A.17}$$

Traditionally the dispersion is quoted at the scale $8h^{-1}$ Mpc, and given the symbol σ_8. For our experimental value we used the result of Viana and Liddle [109], who analysed the abundance of large galaxy clusters to obtain

$$\sigma_8 = 0.56\,\Omega_0^{-0.47}, \tag{A.18}$$

with relative 95% confidence limits of $-18\,\Omega_0^{0.2\log_{10}\Omega_0}$ and $+20\,\Omega_0^{0.2\log_{10}\Omega_0}$ percent. Several other estimates have been published; the one we used is fairly representative, and with a more conservative error bar than most.

To compare this experimental result with the model value predicted by Eq. (A.16), we must fix the normalization $\delta_{\rm H}$. We used the result of Liddle *et al.* [110] who fitted $\delta_{\rm H}$ using the *COBE* large scale normalization to obtain

$$10^5\,\delta_{\rm H}(n_{\rm S},\Omega_0) = 1.94\,\Omega_0^{-0.785-0.05\ln\Omega_0}\exp[f(n_{\rm S})], \tag{A.19}$$

where

$$f(n_{\rm S}) = \begin{cases} -0.95\widetilde{n} - 0.169\widetilde{n}^2, & \text{No tensors,} \\ 1.00\widetilde{n} + 1.97\widetilde{n}^2, & \text{PLI.} \end{cases} \tag{A.20}$$

For the case of non-PLI tensors, we used the fitting form of Bunn *et al.* [111]:

$$\begin{aligned} 10^5\,\delta_{\rm H} = {} & 1.91\,\Omega_0^{-0.80-0.05\ln\Omega_0}\frac{\exp(-1.01\widetilde{n})}{\sqrt{1+(0.75-0.13\Omega_\Lambda^2)\,r}} \\ & \times(1+0.18\widetilde{n}\Omega_\Lambda - 0.03r\Omega_\Lambda)\,. \end{aligned} \tag{A.21}$$

We calculated likelihoods for our model σ_8 using a Gaussian with peak and 95% limits specified by Eq. (A.18), and then integrated $\mathcal{L}$ and found limits for T/S as in the binned microwave case.

A.4.3 Lyman α absorption cloud statistics

Another measure of the amplitude of the matter power spectrum has been obtained recently by Croft *et al.* [112], who analysed the Lyman α (Ly α) absorption forest in the spectra of quasars at redshifts $z \simeq 2.5$. These results apply at smaller comoving scales than the cluster abundance σ_8 measurements, and hence are potentially more constraining. Croft *et al.* found

$$\Delta^2(k_p) = 0.57\,^{+0.26}_{-0.18} \tag{A.22}$$

at 1σ confidence, where the effective wavenumber $k_p = 0.008(\text{km s}^{-1})^{-1}$ at $z = 2.5$.

These results cannot be directly compared with the model predictions of Eq. (A.13), because Eq. (A.13) provides its predictions for the current time, *i.e.* $z = 0$. To translate to $z = 2.5$, we must first convert the model k from the comoving Mpc^{-1} units conventionally used in discussions of the matter power spectrum to $(\text{km s}^{-1})^{-1}$ at $z = 2.5$, using

$$k[(\text{km s}^{-1})^{-1}] = \frac{1+z}{H(z)}\,k[\text{Mpc}^{-1}], \tag{A.23}$$

where

$$H(z) = H_0\sqrt{\Omega_0(1+z)^3 + \Omega_\Lambda} \tag{A.24}$$

for flat universes.

Next, we must consider the growth of the perturbations themselves. In a critical density universe (and assuming linear theory), the growth law is simply $\Delta^2(k,z) = \Delta^2(k,0)(1+z)^{-2}$. As Ω_Λ increases, the growth is suppressed, and this can be accounted for by writing

$$\Delta^2(k,z) = \Delta^2(k,0)\frac{g^2[\Omega(z)]}{g^2(\Omega_0)}\frac{1}{(1+z)^2}, \tag{A.25}$$

where the growth suppression factor $g(\Omega)$ can be accurately parametrized by [113]

$$g(\Omega) = \frac{5}{2}\Omega\left(\frac{1}{70} + \frac{209\Omega}{140} - \frac{\Omega^2}{140} + \Omega^{4/7}\right)^{-1}, \tag{A.26}$$

and the redshift dependence of Ω is given by

$$\Omega(z) = \Omega_0\,\frac{(1+z)^3}{1-\Omega_0+(1+z)^3\Omega_0}, \tag{A.27}$$

all for spatially flat universes.

We calculated likelihoods using the normalized model predictions of Eq. (A.13), translated to $z = 2.5$ as described above, and then obtained limits for T/S as in the cluster abundance case.

A.5 Parameter constraints

A.5.1 Age of the universe

In flat Λ models, the age of the universe is [113]

$$t_0 = \frac{2}{3H_0} \frac{\sinh^{-1}\left(\sqrt{\Omega_\Lambda/\Omega_0}\right)}{\sqrt{\Omega_\Lambda}}. \tag{A.28}$$

During the integration of the likelihoods, we investigated the effect of imposing a constraint on the parameters h and Ω_0, so that regions of parameter space corresponding to ages below various limits were excluded. This simply corresponds to a more complex form for the priors on the parameters. The precise limit on the age of the universe is a matter of on-going debate (*e.g.* [114, 115]). A lower limit of around 11 Gyr now seems to be the norm, so we considered this case explicitly. We also considered the effect of a more constraining limit of 13 Gyr, still preferred by some authors.

A.5.2 Baryons in clusters

Recent measurements of the baryon density in clusters have suggested low Ω_0 for consistency with nucleosythesis. We chose to use the results of White and Fabian [116] for the baryon density

$$\frac{\Omega_B}{\Omega_0} = (0.056 \pm 0.014)h^{-3/2}, \tag{A.29}$$

where the errors are at the 1σ level. We explored the implications of applying this constraint during the likelihood integrations, by adding a term

$$\left(\frac{h^{3/2}\Omega_B/\Omega_0 - 0.056}{0.014}\right)^2 \tag{A.30}$$

to each value of χ^2.

A.5.3 Supernova constraints

Measurements of high-z Type-Ia supernovae (SNe Ia) are in principle well-suited to constraining Ω_0 on the assumption of a flat Λ universe, since such measurements are sensitive to (roughly) the difference between Ω_0 and Ω_Λ. We used the experimental results of Filippenko and Riess of the High-z Supernova Search team [117], who found for flat Λ models

$$\Omega_0 = 0.25 \pm 0.15 \tag{A.31}$$

at 1σ confidence. We also investigated the effect of applying this constraint as above.

A.6 Open models

For models with open geometry the situation is more complicated, and so we restrict ourselves to a brief discussion here. In addition to the added technical complexity involved in working in hyperbolic spaces, the presence of an additional scale, the curvature scale, renders ambiguous the meaning of scale-invariant fluctuations. For the most obvious scale-invariant spectrum of gravity wave modes, the quadrupole anisotropy actually diverges! For this reason one requires a definite calculation of the fluctuation spectrum from a well realized open model. The advent of open inflationary models [118] has allowed, for the first time, a *calculation* of the spectrum of primordial fluctuations in an open universe. As with all inflationary models, a nearly scale-invariant spectrum of gravitational waves (tensor modes) is produced [119, 120]. The size of these modes in k-space, and their relation to the spectral index, is not dissimilar to the flat space models we have been considering. In the inflationary open universe models the spectrum of perturbations is cut-off at large spatial scales, leading to a finite gravity wave spectrum. However, the exact scale of the cutoff depends on details of the model, introducing further model dependence into the ℓ-space predictions.

Since gravitational waves provide anisotropies but no density fluctuations, their presence will in general lower the normalization of the matter power spectrum (for a fixed large angle CMB normalization). Open models already have quite a low normalization [121, 122], so the most conservative limits on gravity waves come from models which produce the *minimal* tensor anisotropies, i.e. where the cutoff operates as efficiently as possible. The *COBE* normalization for such models with PLI is [123]

$$10^5\,\delta_{\rm H} = 1.95\,\Omega_0^{-0.35-0.19\ln\Omega_0+0.15\widetilde{n}}\exp\left(1.02\widetilde{n}+1.70\widetilde{n}^2\right)\,. \qquad \text{(A.32)}$$

Combining this normalization with the cluster abundance gives a strong constraint on T/S. We show in Fig. A.1 the 95% CL upper limit on T/S as a function of Ω_0 in these models.

A.7 Results

Figure A.2 presents likelihoods, integrated over the parameter ranges described above, and plotted versus T/S, for the various data sets, and specifically for

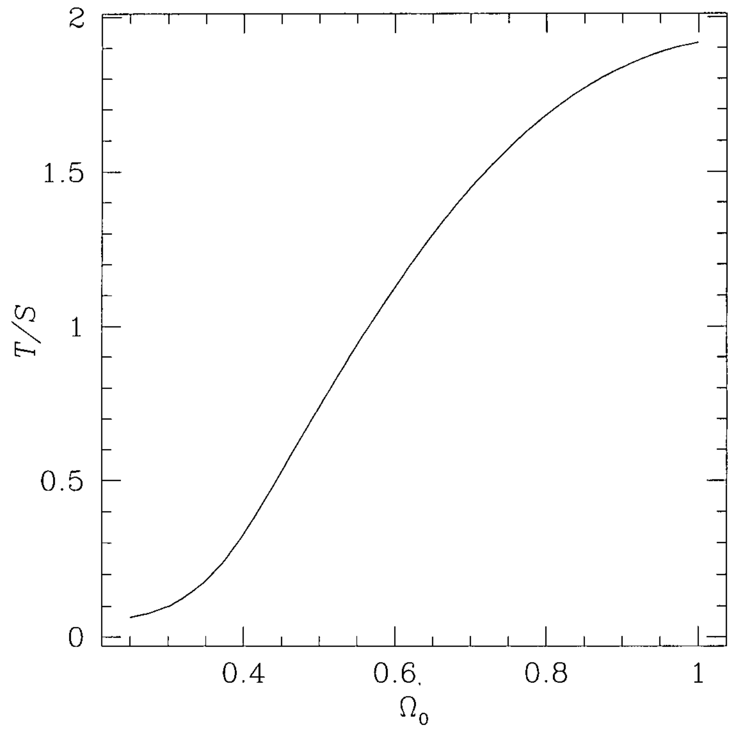

Figure A.1: 95% upper confidence limits on T/S for the open models described in the text. The cluster abundance data set was used with no parameter constraints.

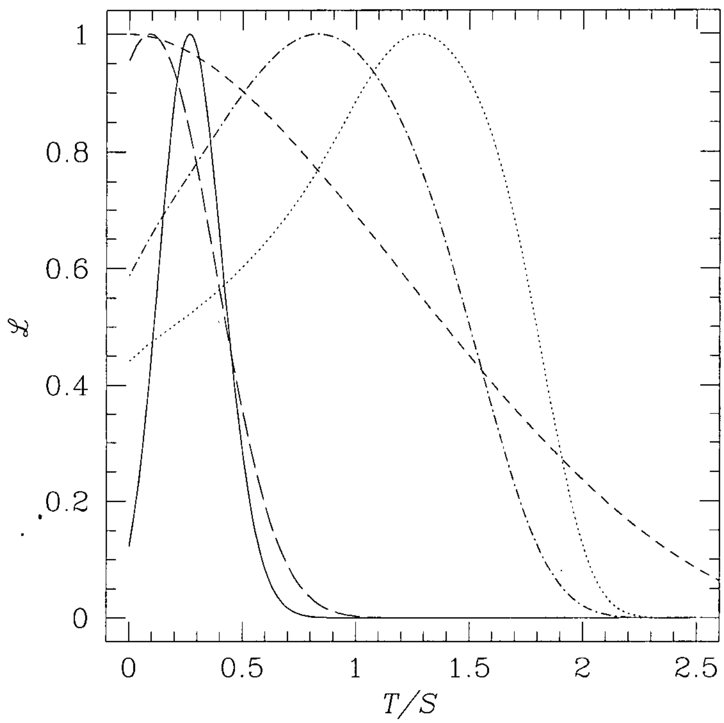

Figure A.2: Integrated likelihoods versus T/S for the various data sets: short-dashed, long-dashed, dotted, dash-dotted, and solid curves represent respectively *COBE*, binned CMB, cluster abundance, Ly α absorption, and combined data. All curves are for PLI inflation, using the priors discussed in Sec. A.3, with no additional parameter constraints.

PLI models. For the curve labelled "combined", likelihoods for each data set (except the *COBE* data) were multiplied together before integration. (Including the *COBE* data would have been redundant, since the binned CMB set already contains the *COBE* results.) Thus the "combined" values represent joint likelihoods for the relevant data sets, on the assumption of independent data. Note that the combined data curve of Fig. A.2 differs significantly from the product of the already marginalized curves for the different data sets, which indicates that parameter covariance is important here. Also, the maximum joint likelihood in Fig. A.2 corresponds to $\chi^2 \simeq 9$, which indicates a good fit for the 15 degrees of freedom involved.

Figure A.3 displays integrated likelihoods versus T/S for each data set and for the combined data, again on the assumption of PLI. The effect of each parameter constraint is illustrated. The *COBE* data shape constraint is very weak, and exhibits essentially no cosmological parameter dependence, as expected. Thus the parameter constraints have little effect on the likelihoods, and the curves are not shown here. Cluster abundance is not much more constraining than the *COBE* shape, but exhibits considerably stronger cosmological parameter dependence, and hence is affected substantially by the various parameter constraints. The matter power spectrum shape constraint is so weak that we do not plot it here. The strongest constraint comes from the binned CMB data, and indeed these data dominate the joint results.

We can understand the general features of the large parameter dependence exhibitted by the likelihoods for the matter spectrum data sets as follows. Near sCDM parameter values, it is well known that the matter power spectrum contains too much small-scale power when *COBE*-normalized at large scales. The presence of tensors improves the fit at small scales by decreasing the *scalar* normalization at *COBE* scales. Reducing h or Ω_0, however, also decreases the power at small scales, improving the fit over sCDM, and thus reducing the need for tensors. When an age constraint is applied, we force the model towards lower h and Ω_0 according to Eq. (A.28), and hence towards lower T/S, as is seen in Fig. A.3. The cluster baryon and supernova constraints similarly move us to smaller Ω_0.

Figure A.4 displays likelihoods versus p for ϕ^p inflation, while Fig. A.5 presents likelihoods versus T/S for the case of free tensor contribution and $n_S = 1$. In all plots, curves have been omitted for the very weakly constraining data sets. The curves of Fig. A.4 closely resemble those of Fig. A.3. This is because, for $p \gg 2$, Eqs. (A.5), (A.6), and (A.7) give $T/S \simeq -6.85\,\widetilde{n}$, which is similar to the PLI result of Eq. (A.3).

In Fig. A.5 we see that the data are considerably less constraining, compared with the PLI case, when we allow T/S to vary freely. This was expected, since in the PLI case, the lowering of n_S tends to enhance the effect of increasing T/S. We also expect that a blue scalar tilt would oppose the effect of tensors on the spectrum and hence allow larger T/S. Figure A.6 illustrates this effect by plotting the 95% upper confidence limits on T/S versus scalar index with T/S free. Note that we cannot meaningfully constrain T/S here by marginalizing over n_S, since the best fits to the spectrum remain good even for very blue tilts and very large T/S.

Our confidence limits are summarized in Table A.1 for the case of power-law inflation, Table A.2 for polymomial potentials, and Table A.3 for free T/S and $n_S = 1$. In all cases 95% upper limits on T/S are presented, after integrating over the ranges of parameter space specified in Sec. A.3. The row gives the data set used, while the column specifies the type of parameter constraint applied, if any.

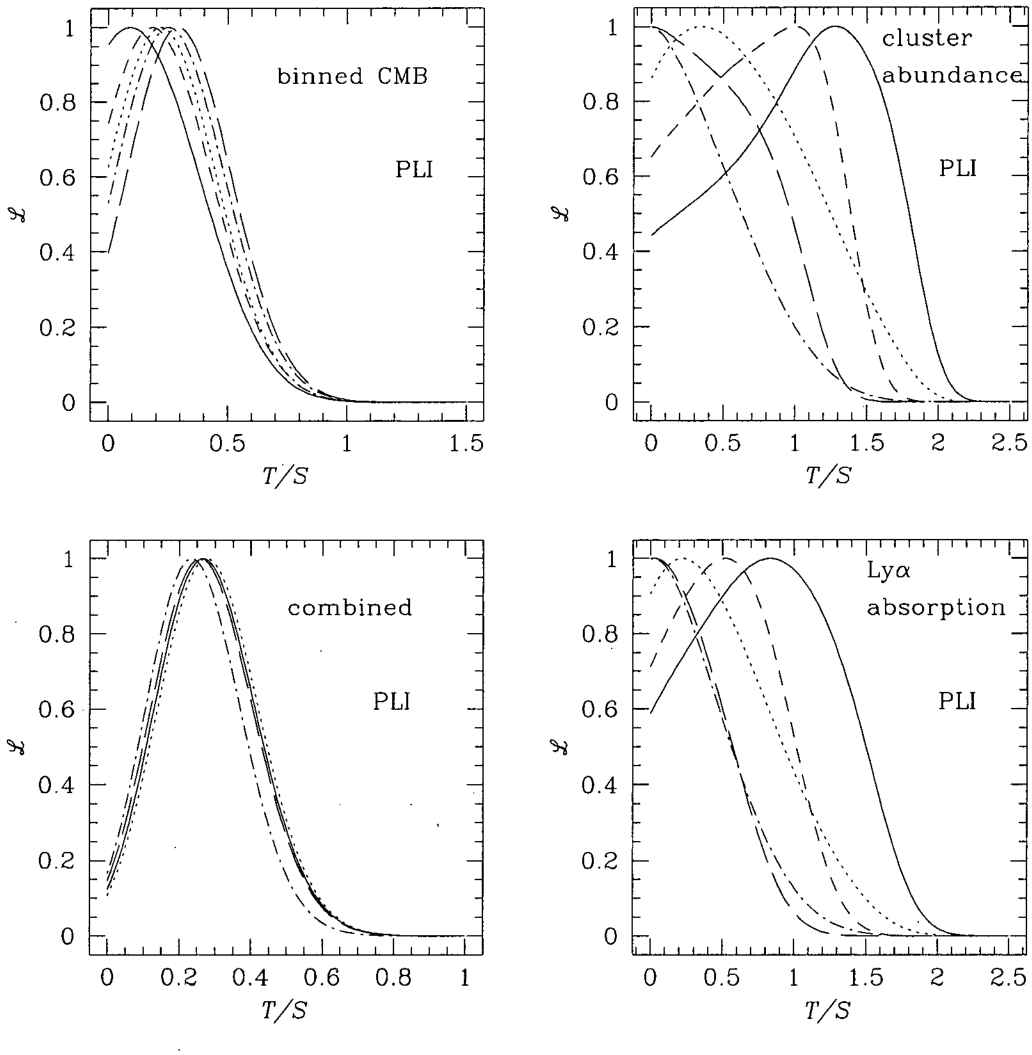

Figure A.3: Integrated likelihoods versus T/S for PLI and for the various data sets; clockwise from upper left: binned CMB, cluster abundance, Ly α absorption, and combined data. Solid, short-dashed, long-dashed, dotted, and dot-dashed curves represent no constraint, $t_0 > 11$ Gyr, $t_0 > 13$ Gyr, cluster baryon fraction, and SNe Ia parameter constraints, respectively.

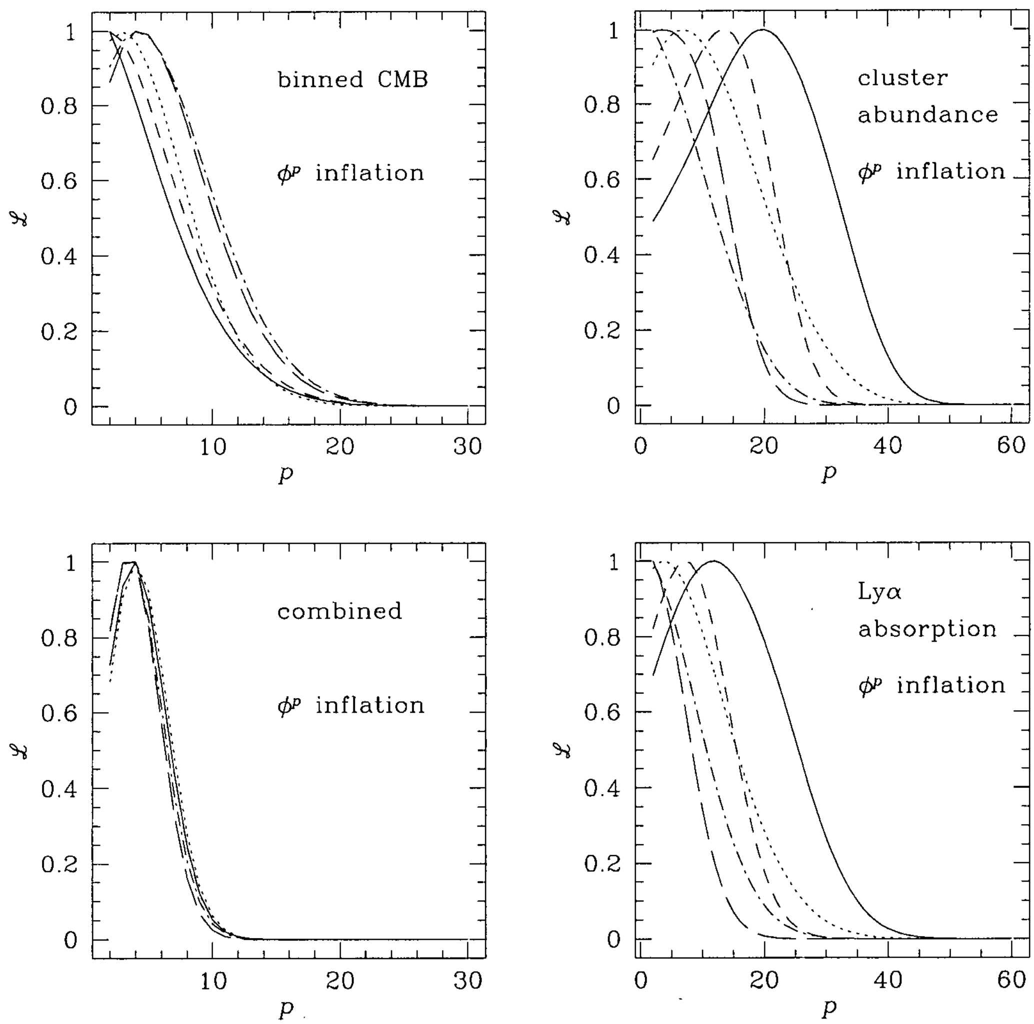

Figure A.4: Integrated likelihoods versus p for ϕ^p inflation and for various data sets; clockwise from upper left: binned CMB, cluster abundance, Ly α absorption, and combined data. Solid, short-dashed, long-dashed, dotted, and dot-dashed curves represent no constraint, $t_0 > 11$ Gyr, $t_0 > 13$ Gyr, cluster baryon fraction, and SNe Ia parameter constraints, respectively.

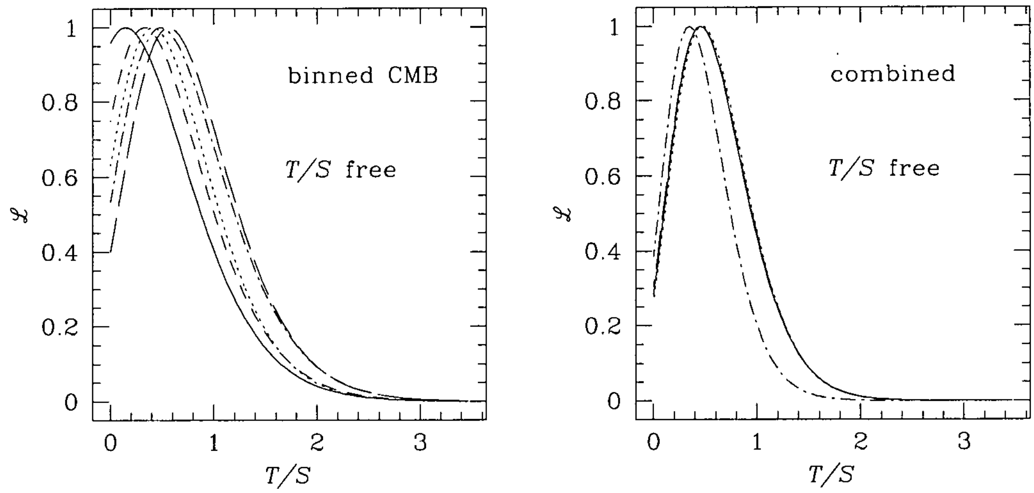

Figure A.5: Integrated likelihoods versus T/S for T/S free and $n_S = 1$, for binned CMB (left) and combined data (right). Solid, short-dashed, long-dashed, dotted, and dot-dashed curves represent no constraint, $t_0 > 11$ Gyr, $t_0 > 13$ Gyr, cluster baryon fraction, and SNe Ia parameter constraints, respectively.

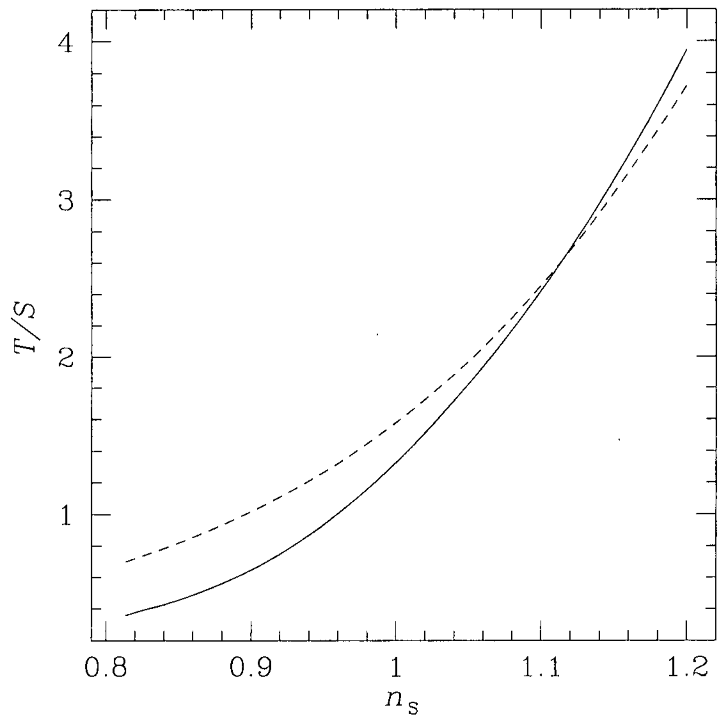

Figure A.6: 95% confidence limits on T/S versus n_S, for T/S free and for binned CMB (dashed) and combined data (solid). No parameter constraints have been applied.

Table A.1: 95% confidence limits on T/S for various data sets and parameter constraints, and for power-law inflation. "2.5+" represents no constraint (for values as high as this we do not expect our approximations to be adequate in any case).

Data set	No constr.	$t_0 > 11$ Gyr	$t_0 > 13$ Gyr	Baryon	SN
COBE	2.1	2.1	2.1	2.1	2.1
binned CMB	0.60	0.62	0.67	0.63	0.65
galaxy correl.	2.5+	2.5+	2.5+	2.5+	2.5+
cluster abund.	1.8	1.4	1.1	1.5	1.1
Ly α	1.6	1.1	0.82	1.3	0.96
combined	0.52	0.52	0.51	0.53	0.47

Table A.2: 95% confidence limits on T/S as for Table A.1 but for ϕ^p inflation.

Data set	No constr.	$t_0 > 11$ Gyr	$t_0 > 13$ Gyr	Baryon	SN
binned CMB	0.63	0.64	0.67	0.64	0.65
galaxy correl.	2.5+	2.5+	2.5+	2.5+	2.5+
cluster abund.	2.1	1.5	1.2	1.8	1.2
Ly α	1.8	1.2	0.87	1.5	1.1
combined	0.49	0.49	0.49	0.50	0.45

Table A.3: 95% confidence limits on T/S as for Table A.1 but for $n_S = 1$ and T/S free.

Data set	No constr.	$t_0 > 11$ Gyr	$t_0 > 13$ Gyr	Baryon	SN
COBE	2.5+	2.5+	2.5+	2.5+	2.5+
binned CMB	1.6	1.6	1.8	1.6	1.8
cluster abund.	2.5+	2.5+	2.5+	2.5+	2.5+
Ly α	2.5+	2.5+	2.5+	2.5+	2.5+
combined	1.3	1.3	1.3	1.3	1.0

A.8 The Future

The discovery of a nearly scale-invariant spectrum of long wavelength gravity waves would be tremendously illuminating. Inflation is the only known mechanism for producing an almost scale-invariant spectrum of adiabatic scalar fluctuations, a prediction which is slowly gaining observational support. In the simplest, "toy", models of inflation a potentially large amplitude almost scale-invariant spectrum of gravity waves is also predicted. For monomial inflation models within the slow-roll approximation, detailed characterization of this spectrum could in principle allow a reconstruction of the inflaton potential [83]. This surely is a window onto physics at higher energies than have ever been probed before.

Inflation models based on particle physics, rather than "toy" potentials, predict a very small tensor spectrum [94]. However, essentially nothing is known about particle physics above the electroweak scale, and extrapolations of our current ideas to arbitrarily high energies could easily miss the mark. We must be guided then by observations. We have argued that observational support for a large gravity wave component is weak. Indeed observations definitely require the tensor anisotropy to be subdominant for large angle CMB anisotropies. One the other hand, it is still possible to have $T/S \simeq 0.5$, and since it would be so exciting to discover any tensor signal at all we are led to ask: how small can a tensor component be and still be detectable? What are the best ways to look for a tensor signal?

A.8.1 Direct detection

The feasibility of the direct detection of inflation-produced gravitational waves has been addressed by a number of authors [74, 124–129], with pessimism expressed by most.

The ground-based laser interferometers LIGO and VIRGO [130] will operate in the $f \sim 100\,\mathrm{Hz}$ frequency band, while the European Space Agency's proposed space-based interferometer LISA [131] would operate in the $f \sim 10^{-4}\,\mathrm{Hz}$ band. Millisecond pulsar timing is sensitive to waves with periods on the order of the observation time, *i.e.* frequencies $f \sim 10^{-7} - 10^{-9}\,\mathrm{Hz}$ [130]. These instruments probe regions of the tensor perturbation spectrum which entered during the radiation dominated era. Expressions for the fraction of the critical density due to gravity waves per logarithmic frequency interval can be found in [124–128]. Assuming that $\Omega_0 = 1$ in a PLI model, with the only relativistic particles being photons and 3 neutrino species, and taking the *COBE* quadrupole $Q = T + S \simeq 4.4 \times 10^{-11}$, one finds [128]

$$\Omega_{\mathrm{GW}}(f)h^2 = 5.1 \times 10^{-15} \frac{n_{\mathrm{T}}}{n_{\mathrm{T}} - 1/7} \exp\left[n_{\mathrm{T}} N + \frac{1}{2} N^2 (dn_{\mathrm{T}}/d\ln k)\right], \quad \text{(A.33)}$$

where $N \equiv \ln(k/H_0)$ and $n_{\mathrm{T}} = -(T/S)/7$ is the tensor spectral index.

Using Eq. (A.33) Turner [128] found that the local energy density in gravity waves is maximized at $T/S = 0.18$ for $f \sim 10^{-4}\,\mathrm{Hz}$. At this maximum, the local energy density is in the range $\Omega_{\mathrm{GW}} h^2 \simeq 10^{-15} - 10^{-16}$, which lies a couple of orders of magnitude below the expected sensitivity of LISA, and several orders below that of LIGO/VIRGO [130]. This is also well below the current upper limit of $\Omega_{\mathrm{GW}} h^2 < 6 \times 10^{-8}$ (at 95% confidence) from pulsar timing [132]. As T/S increases above 0.18, $\Omega_{\mathrm{GW}}(f \sim 10^{-4}\mathrm{Hz})h^2$ begins to *decrease* due to the increasing magnitude of the tensor spectral index.

Recall that our joint data constraint for PLI gives $T/S \lesssim 0.5$, so our results predict that the inflationary spectrum of gravity waves from PLI is not amenable to direct detection.

A.8.2 Limits from the CMB

With the advent of *WMAP* and especially the *Planck* Surveyor, with its higher sensitivity, detailed maps of the CMB are just around the corner. What do we expect will be possible from these missions? This question has been dealt with extensively before. Assuming a cosmic variance limited experiment capable of determining only the anisotropy in the CMB but with all other parameters known, one can measure T/S only if it is larger than about 10% [133]. A more realistic assessment for *MAP* and *Planck* suggests this limit is rarely reached in practice [134, 135].

However the ability to measure linear polarization in the CMB anisotropy offers the prospect of improving the sensitivity to tensor modes (for a recent review of polarization see [136]). In addition to the temperature anisotropy, two components of the linear polarization can be measured. It is convenient to split the polarization into parity even (E-mode) and parity odd (B-mode) combinations – named after the familiar parity transformation properties of the electric and magnetic fields, but not to be confused with the E and B fields of the electromagnetic radiation.

Polarization offers two advantages over the temperature. First, with more observables the error bars on parameters are tightened. In addition the polarization breaks the degeneracy between reionization and a tensor component, allowing extraction of smaller levels of signal [137]. Model dependent constraints on a tensor mode as low as 1% appear to be possible with the *Planck* satellite [134, 135, 138, 139]. Extensive observations of patches of the sky from the ground (or satellites even further into the future) could in principle push the sensitivity even deeper.

There is a further handle on the tensor signal however. Since scalar modes have no "handedness" they generate only parity even, or E-mode polarization [137, 140]. A definitive detection of B-mode polarization would thus indicate the presence of other modes, with tensors being more likely since vector modes decay cosmologically. Moreover a comparison of the B-mode, E-mode and temperature signals can definitively distinguish tensors from other sources of perturbation (*e.g.* [141]).

Unfortunately the detection of a B-mode polarization will prove a formidable experimental challenge. The level of the signal, shown in Fig. A.7 for $T/S = 0.01$, 0.1 and 1.0, is very small. As an indicative number, with $T/S = 0.5$, our upper limit, the *total* rms B-mode signal, integrated over ℓ, is $0.24\,\mu$K in a critical density universe. These sensitivity requirements, coupled with our current poor state of knowledge of the relevant polarized foregrounds make it seem unlikely a B-mode signal will be detected in the near future.

A.9 Conclusions

We have examined the current experimental limits on the tensor-to-scalar ratio. Using the *COBE* results, as well as small-scale CMB observations, and measurements of galaxy correlations, cluster abundances, and Ly α absorption we have obtained conservative limits on the tensor fraction for some specific inflationary models. Importantly, we have considered models with a wide range of cosmological parameters, rather than fixing the values of Ω_0, H_0, etc. For power-law inflation, for example, we find that $T/S < 0.52$ at the 95% con-

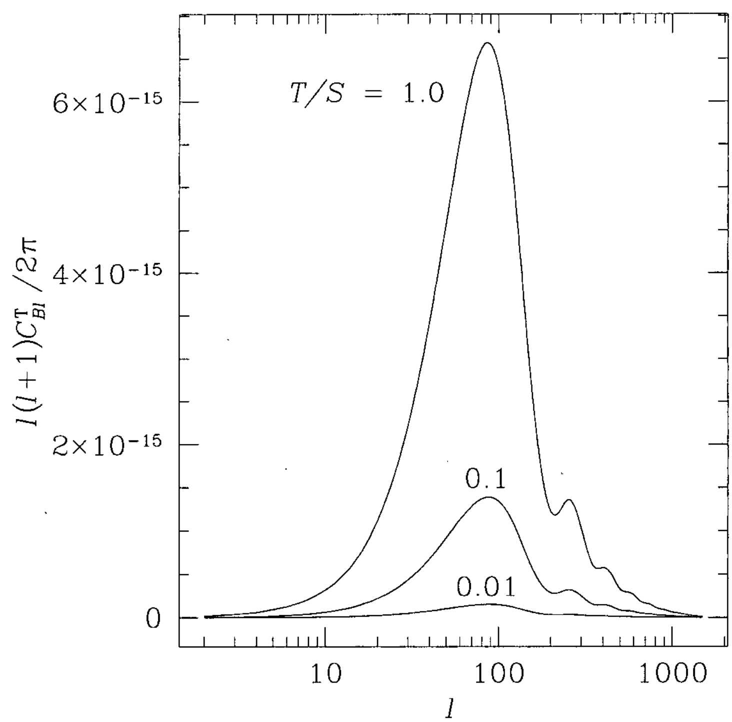

Figure A.7: B-mode tensor polarization signal $C^{\mathrm{T}}_{B\ell}$ for T/S = 0.01, 0.1, and 1.0, with the remaining parameters specified as standard CDM.

fidence level. Similar constraints apply to ϕ^p inflaton models, corresponding to approximately $p < 8$. Much of this constraint on the tensor-to-scalar ratio comes from the relation between T/S and the scalar spectral index n_S in these theories. For models with tensor amplitude unrelated to *scalar* spectral index it is still possible to have $T/S > 1$. Currently the tightest constraint is provided by the combined CMB data sets. Since the quality of such data are expected to improve dramatically in the near future, we expect much tighter constraints (or more interestingly a real detection) in the coming years.