Chapter 3

Cosmological models

George F R Ellis
Mathematics Department, University of Cape Town, South Africa

3.1 Introduction

The current standard models of the universe are the Friedmann–Lemaître (FL) family of models, based on the Robertson–Walker (RW) spatially homogeneous and isotropic geometries but with a much more complex set of matter constituents than originally envisaged by Friedmann and Lemaître. It is appropriate then to ask whether the universe is indeed well described by an RW geometry. There is reasonable evidence supporting these models on the largest observable scales, but at smaller scales they are clearly a bad description. Thus a better form of the question is: *On what scales and in what domains is the universe's geometry nearly RW? What are the best-fit RW parameters in the observable domain?*

Given that the universe is apparently well described by the RW geometry on the largest scales in the observable domain, the next question is: *Why is it RW? How did the universe come to have such an improbable geometry?* The predominant answer to this question at present is that it results from a very early epoch when inflation took place (a period of accelerating expansion through many e-folds of the scale of the universe). It is important to consider how good an answer this is. One can only do so by considering alternatives to RW geometries, as well as the models based on those geometries.

The third question is: *How did astronomical structure come to exist on smaller scales? Given a smooth structure on the largest scales, how was that smoothness broken on smaller scales?* Again, inflationary theory applied to perturbed FL models gives a general answer to that question: quantum fluctuations in the very early universe formed the seeds of inhomogeneities that could then grow, on scales bigger than the (time-dependent) Jeans' scale, by gravitational attraction. It is important to note, however, that not only do structure-formation effects depend in important ways on the background model, but also (and indeed, in consequence of this remark) many of the ways of estimating the model parameters depend on models of structure formation. Thus the previous questions and this one interact in a number of ways.

This review will look at the first two questions in some depth, and only briefly consider the third (which is covered in depth in Peacock's chapter). To examine these questions, we need to consider the family of cosmological solutions with observational properties like those of the real universe at some stage of their histories. Thus we are interested in the *full state space of solutions*, allowing us to see how realistic (lumpy) models are related to each other and to higher symmetry models, including, in particular, the FL models. This chapter develops general techniques for examining this family of models, and describes some specific models of interest. The first part looks at exact general relations valid in all cosmological models, the second part examines exact cosmological solutions of the field equations and the third part looks at the observational properties of these models and then returns to considering the previous questions. The chapter concludes by emphasizing some of the fundamental issues that make it difficult to obtain definitive answers if one tries to pursue the chain of cause and effect to extremely early times.

3.1.1 Spacetime

We will make the standard assumption that on large scales, physics is dominated by gravity, which is well described by general relativity (see, e.g. d'Inverno [19], Wald [129], Hawking and Ellis [68] or Stephani [117]), with gravitational effects resulting from spacetime curvature. The starting point for describing a spacetime is an atlas of local coordinates $\{x^i\}$ covering the four-dimensional spacetime manifold $\mathcal{M}$, and a Lorentzian metric tensor $g_{ij}(x^k)$ at each point of $\mathcal{M}$, representing the spacetime geometry near the point on a particular scale. This then determines the connection components $\Gamma^i{}_{jk}(x^s)$, and, hence, the spacetime curvature tensor R_{ijkl}, at that scale. The curvature tensor can be decomposed into its trace-free part (the Weyl tensor $C_{ijkl} : C^i{}_{jil} = 0$) and its trace (the Ricci tensor $R_{ik} \equiv R^s{}_{isk}$) by the relation

$$R_{ijkl} = C_{ijkl} - \tfrac{1}{2}(R_{ik}g_{jl} + R_{jl}g_{ik} - R_{il}g_{jk} - R_{jk}g_{il}) + \tfrac{1}{6}R(g_{ik}g_{jl} - g_{il}g_{jk}), \quad (3.1)$$

where $R \equiv R^a{}_a$ is the Ricci scalar. The coordinates may be chosen arbitrarily in each neighbourhood in $\mathcal{M}$. To be useful in an explanatory role, a cosmological model must be easy to describe—this means they have symmetries or special properties of some kind or other.

3.1.2 Field equations

The metric tensor is determined, at the relevant averaging scale, by the *Einstein gravitational field equations* ('EFEs')

$$(R_{ij} - \tfrac{1}{2}Rg_{ij}) + \lambda g_{ij} = \kappa T_{ij} \Leftrightarrow R_{ij} = \lambda g_{ij} + \kappa(T_{ij} - \tfrac{1}{2}Tg_{ij}) \quad (3.2)$$

where λ is the cosmological constant and κ the gravitational constant. Here T_{ij} (with trace $T = T^a_a$) is the total energy–momentum–stress tensor for all the matter and fields present, described at the relevant averaging scale. This covariant equation (a set of second-order nonlinear equations for the metric tensor components) shows that the Ricci tensor is determined pointwise by the matter present at each point, but the Weyl tensor is not so determined; rather it is fixed by suitable boundary conditions, together with the Bianchi identities for the curvature tensor:

$$\nabla_{[e}R_{ab]cd} = 0 \Leftrightarrow \nabla_{[e}R^e_{ab]cd} = 0 \tag{3.3}$$

(the equivalence of the full equations on the left with the first contracted equations on the right holding only for four dimensions or less). Consequently it is this tensor that enables gravitational 'action at a distance' (gravitational radiation, tidal forces, and so on). Contracting the right-hand of equation (3.3) and substituting into the divergence of equation (3.2) shows T_{ij} necessarily obeys the energy–momentum conservation equations

$$\nabla_j T^{ij} = 0 \tag{3.4}$$

(the divergence of λg_{ij} vanishes provided λ is indeed constant, as we assume). Thus matter determines the geometry which, in turn, determines the motion of the matter (see e.g. [132]). We can look for exact solutions of these equations, or approximate solutions obtained by suitable linearization of the equations; and one can also consider how the solutions relate to Newtonian theory solutions. Care must be exercised in the latter two cases, both because of the nonlinearity of the theory, and because there is no fixed background spacetime available in general relativity theory. This makes it essentially different from both Newtonian theory and special relativity.

3.1.3 Matter description

The total stress tensor T_{ij} is the sum of the N stress tensors T_{nij} for the various matter components labelled by index n (baryons, radiation, neutrinos, etc):

$$T_{ij} = \Sigma_n T_{nij} \tag{3.5}$$

each component being described by suitable equations of state which encapsulate their physics. The most common forms of matter in the cosmological context will often to a good approximation, each have a 'perfect fluid' stress tensor;

$$T_{nij} = (\mu_n + p_n)u_{ni}u_{nj} + p_n g_{ij} \tag{3.6}$$

with unit 4-velocity u^i_n ($u_{ni}\ u^i_n = -1$), energy density μ_n and pressure p_n, with suitable equations of state relating μ_n and p_n. In simple cases, they will be related by a barotropic relation $p_n = p_n(\mu_n)$; for example, for baryons, $p_\mathrm{b} = 0$ and for radiation, e.g. the cosmic background radiation ('CBR'), $p_r = \mu_r/3$,. However, in more complex cases there will be further variables determining p_n and μ_n; for example, in the case of a massless scalar field ϕ with potential $V(\phi)$, on choosing u^i as the unit vector normal to spacelike surfaces ϕ = constant, the stress tensor takes the form (3.6) with

$$4\pi p_\phi = \tfrac{1}{2}\dot{\phi}^2 - V(\phi), \qquad 4\pi\mu_\phi = \tfrac{1}{2}\dot{\phi}^2 + V(\phi). \tag{3.7}$$

It must be noted that, in general, different matter components will each have a different 4-velocity u^i_n, and the total stress tensor (3.5) of perfect fluid stress tensors (3.6) itself has the perfect fluid form if and only if the 4-velocities of all contributing matter components are the same, i.e. $u^i_n = u^i$ for all n; in that case,

$$T_{ij} = (\mu + p)u_i u_j + p g_{ij}, \qquad \mu \equiv \Sigma_n \mu_n, \qquad p \equiv \Sigma_n p_n \tag{3.8}$$

where μ is the total energy density and p the total pressure.

The individual matter components will each separately satisfy the conservation equation (3.4) if they are non-interacting with the other components; however this will no longer be the case if interactions lead to exchanges of energy and momentum between the different components. The key to a physically realistic cosmological model is the representation of suitable matter components, with realistic equations of state for each matter component and equations describing the interactions between the components. For reasonable behaviour of matter, irrespective of its constitution we require the 'energy condition'

$$\mu + p > 0 \tag{3.9}$$

on cosmological averaging scales (the vacuum case $\mu + p = 0$ can apply only to regions described on averaging scales less than or equal to that of clusters of galaxies).

3.1.4 Cosmology

A key feature of cosmological models, as contrasted with general solutions of the EFEs, is that in them, at each point a unique 4-velocity u^a is defined representing the preferred motion of matter there on a cosmological scale. Whenever the matter present is well described by the perfect fluid stress tensor (3.8), because of (3.9) there will be a unique timelike eigenvector of this tensor that can be used to define the vector u, representing the average motion of the matter, and conventionally referred to as defining the *fundamental world-lines* of the cosmology. Unless stated otherwise, we will assume that observers move with this 4-velocity. At late times, a unique frame is defined by choosing a 4-velocity such that the CBR anisotropy dipole vanishes; the usual assumption is that this is the same frame as defined locally by the average motion of matter [26]; indeed this assumption is what underlies studies of large-scale motions and the 'Great Attractor'.

The description of matter and radiation in a cosmological model must be sufficiently complete to determine the *observational relations* predicted by the model for both discrete sources and the background radiation, implying a well-developed theory of *structure growth* for very small and for very large physical scales (i.e. for light atomic nuclei and for galaxies and clusters of galaxies), and of *radiation absorbtion and emission.* Clearly an essential requirement for a viable cosmological model is that it should be able to reproduce current large-scale astronomical observations accurately.

I will deal with both the $1 + 3$ covariant approach [21, 26, 28, 91] and the orthonormal tetrad approach, which serves as a completion to the $1 + 3$ covariant approach [41].

3.2 $1 + 3$ covariant description: variables

3.2.1 Average 4-velocity of matter

The preferred 4-velocity is

$$u^a = \frac{\mathrm{d}x^a}{\mathrm{d}\tau}, \qquad u_a u^a = -1, \tag{3.10}$$

where τ is the proper time measured along the fundamental world-lines. Given u^a, unique *projection tensors* can be defined:

$$\begin{aligned} U^a{}_{\mathrm{b}} = -u^a u_{\mathrm{b}} &\Rightarrow U^a{}_c U^c{}_{\mathrm{b}} = U^a{}_{\mathrm{b}}, U^a{}_a = 1, U_{ab}u^b = u_a, \\ h_{ab} = g_{ab} + u_a u_{\mathrm{b}} &\Rightarrow h^a{}_c h^c{}_{\mathrm{b}} = h^a{}_{\mathrm{b}}, h^a{}_a = 3, h_{ab}u^b = 0. \end{aligned} \tag{3.11}$$

The first projects parallel to the velocity vector u^a, and the second determines the metric properties of the (orthogonal) instantaneous rest-spaces of observers moving with 4-velocity u^a. A *volume element* for the rest spaces:

$$\eta_{abc} = u^d \eta_{dabc} \Rightarrow \eta_{abc} = \eta_{[abc]}, \eta_{abc}u^c = 0, \tag{3.12}$$

where η_{abcd} is the four-dimensional volume element ($\eta_{abcd} = \eta_{[abcd]}$, $\eta_{0123} = \sqrt{|\det g_{ab}|}$) is also defined.

Furthermore, two derivatives are defined: the covariant time derivative '$\cdot$' along the fundamental world-lines, where for any tensor $T^{ab}{}_{cd}$

$$\dot{T}^{ab}{}_{cd} = u^e \nabla_e T^{ab}{}_{cd}, \tag{3.13}$$

and the fully orthogonally projected covariant derivative $\tilde{\nabla}$ where, for any tensor $T^{ab}{}_{cd}$,

$$\tilde{\nabla}_e T^{ab}{}_{cd} = h^a{}_f h^b{}_g h^p{}_c h^q{}_d h^r{}_e \nabla_r T^{fg}{}_{pq}, \tag{3.14}$$

with total projection on all free indices. The tilde serves as a reminder that if u^a has *non-zero* vorticity, $\tilde{\nabla}$ is *not* a proper three-dimensional covariant derivative (see equation (3.20)). The projected time and space derivatives of U_{ab}, h_{ab} and η_{abc} all vanish. Finally, following [91] we use angle brackets to denote orthogonal projections of vectors and the orthogonally projected symmetric trace-free part of tensors:

$$v^{\langle a\rangle} = h^a{}_b v^b, \qquad T^{\langle ab\rangle} = [h^{(a}{}_c h^{b)}{}_d - \tfrac{1}{3}h^{ab}h_{cd}]T^{cd}; \tag{3.15}$$

for convenience the angle brackets are also used to denote othogonal projections of covariant time derivatives along u^a (‘*Fermi derivatives*’):

$$\dot{v}^{\langle a\rangle} = h^a{}_b \dot{v}^b, \qquad \dot{T}^{\langle ab\rangle} = [h^{(a}{}_c h^{b)}{}_d - \tfrac{1}{3}h^{ab}h_{cd}]\dot{T}^{cd}. \tag{3.16}$$

3.2.2 Kinematic quantities

The orthogonal vector $\dot{u}^a = u^b\nabla_b u^a$ is the *acceleration vector*, representing the degree to which the matter moves under forces other than gravity plus inertia (which cannot be covariantly separated from each other in general relativity). The acceleration vanishes for matter in free fall (i.e. moving under gravity plus inertia alone).

We split the first covariant derivative of u_a into its irreducible parts, defined by their symmetry properties:

$$\nabla_a u_b = -u_a\dot{u}_b + \tilde{\nabla}_a u_b = -u_a\dot{u}_b + \tfrac{1}{3}\Theta h_{ab} + \sigma_{ab} + \omega_{ab} \tag{3.17}$$

where the trace $\Theta = \tilde{\nabla}_a u^a$ is the *(volume) rate of expansion* of the fluid (with $H = \Theta/3$ the Hubble parameter); $\sigma_{ab} = \tilde{\nabla}_{\langle a}u_{b\rangle}$ is the trace-free symmetric *shear* tensor ($\sigma_{ab} = \sigma_{(ab)}$, $\sigma_{ab}u^b = 0$, $\sigma^a{}_a = 0$), describing the rate of distortion of the matter flow; and $\omega_{ab} = \tilde{\nabla}_{[a}u_{b]}$ is the skew-symmetric *vorticity* tensor ($\omega_{ab} = \omega_{[ab]}$, $\omega_{ab}u^b = 0$), describing the rotation of the matter relative to a non-rotating (Fermi-propagated) frame. The meaning of these quantities follows from the evolution equation for a relative position vector $\eta^a_\perp = h^a{}_b\eta^b$, where η^a is a deviation vector for the family of fundamental world-lines, i.e. $u^b\nabla_b\eta^a = \eta^b\nabla_b u^a$. Writing $\eta^a_\perp = \delta\ell e^a$, $e_a e^a = 1$, we find the relative distance $\delta\ell$ obeys the propagation equation

$$\frac{(\delta\ell)^{\cdot}}{\delta\ell} = \tfrac{1}{3}\Theta + (\sigma_{ab}e^a e^b), \tag{3.18}$$

(the generalized Hubble law), and the relative direction vector e^a the propagation equation

$$\dot{e}^{\langle a\rangle} = (\sigma^a{}_b - (\sigma_{cd}e^c e^d)h^a{}_b - \omega^a{}_b)e^b, \tag{3.19}$$

giving the observed rate of change of position in the sky of distant galaxies [21, 26].

Each function f satisfies the important *commutation relation* for the $\tilde{\nabla}$-derivative [40]

$$\tilde{\nabla}_{[a}\tilde{\nabla}_{b]}f = \eta_{abc}\omega^c\dot{f}. \tag{3.20}$$

Applying this to the energy density μ shows that if $\omega^a \dot{\mu} \neq 0$ in an open set then $\tilde{\nabla}_a \mu \neq 0$ there, so non-zero vorticity implies anisotropic number counts in an expanding universe [61] (this is because there are then no 3-surfaces orthogonal to the fluid flow; see [21, 26]).

3.2.2.1 *Auxiliary quantities*

It is useful to define some associated kinematical quantities:

- the *vorticity vector* $\omega^a = \frac{1}{2}\eta^{abc}\omega_{bc} \Rightarrow \omega_a u^a = 0,\ \omega_{ab}\omega^b = 0$,
- the magnitudes $\omega^2 = \frac{1}{2}(\omega_{ab}\omega^{ab}) \geq 0$, $\sigma^2 = \frac{1}{2}(\sigma_{ab}\sigma^{ab}) \geq 0$, and
- the *average length scale* S determined by $\frac{\dot{S}}{S} = \frac{1}{3}\Theta$, so the volume of a fluid element varies along the fluid flow lines as S^3.

3.2.3 Matter tensor

Both the total matter *energy–momentum tensor* T_{ab} and each of its components can be decomposed relative to u^a in the form

$$T_{ab} = \mu u_a u_{\rm b} + q_a u_{\rm b} + u_a q_{\rm b} + p h_{ab} + \pi_{ab}, \tag{3.21}$$

where $\mu = (T_{ab}u^a u^b)$ is the *relativistic energy density* relative to u^a, $q^a = -T_{bc}u^b h^{ca}$ is the *relativistic momentum density* ($q_a u^a = 0$), which is also the energy flux relative to u^a, $p = \frac{1}{3}(T_{ab}h^{ab})$ is the *isotropic pressure*, and $\pi_{ab} = T_{cd}h^c{}_{\langle a}h^d{}_{b\rangle}$ is the trace-free *anisotropic pressure* ($\pi^a{}_a = 0$, $\pi_{ab} = \pi_{(ab)}$, $\pi_{ab}u^b = 0$). A different choice of u^a will result in a different splitting. The physics of the situation is in the equations of state relating these quantities; for example, the commonly imposed restrictions

$$q^a = \pi_{ab} = 0 \Leftrightarrow T_{ab} = \mu u_a u_{\rm b} + p h_{ab} \tag{3.22}$$

characterize a 'perfect fluid' moving with the chosen 4-velocity u_a as in equation (3.8) with, in general, an equation of state $p = p(\mu, s)$ where s is the entropy [21, 26].

3.2.4 Electromagnetic field

The *Maxwell field tensor* F_{ab} of an electromagnetic field is split relative to u^a into electric and magnetic parts by the relations (see [28])

$$E_a = F_{ab}u^b \Rightarrow E_a u^a = 0, \tag{3.23}$$

$$H_a = \tfrac{1}{2}\eta_{abc}F^{bc} \Rightarrow H_a u^a = 0. \tag{3.24}$$

Again, a different choice of u^a will result in a different split.

3.2.5 Weyl tensor

In analogy to F_{ab}, the *Weyl conformal curvature tensor* C_{abcd} defined by equation (3.1) is split relative to u^a into 'electric' and 'magnetic' *Weyl curvature* parts according to

$$E_{ab} = C_{acbd}u^c u^d \Rightarrow E^a{}_a = 0,\, E_{ab} = E_{(ab)},\, E_{ab}u^b = 0, \tag{3.25}$$

$$H_{ab} = \tfrac{1}{2}\eta_{ade}C^{de}{}_{bc}u^c \Rightarrow H^a{}_a = 0,\, H_{ab} = H_{(ab)},\, H_{ab}u^b = 0. \tag{3.26}$$

These influence the motion of matter and radiation through the *geodesic deviation equation* for timelike and null vectors, see, respectively, [107] and [120].

3.3 1 + 3 Covariant description: equations

There are three sets of equations to be considered, resulting from EFE (3.2) and its associated integrability conditions.

3.3.1 Energy–momentum conservation equations

We obtain from the conservation equations (3.4), on projecting parallel and perpendicular to u^a and using (3.21), the propagation equations

$$\dot{\mu} + \tilde{\nabla}_a q^a = -\Theta(\mu + p) - 2(\dot{u}_a q^a) - (\sigma^a{}_b \pi^b{}_a), \tag{3.27}$$

$$\dot{q}^{\langle a\rangle} + \tilde{\nabla}^a p + \tilde{\nabla}_b \pi^{ab} = -\tfrac{4}{3}\Theta q^a - \sigma^a{}_b q^b - (\mu + p)\dot{u}^a - \dot{u}_b \pi^{ab} - \eta^{abc}\omega_b q_c. \tag{3.28}$$

For perfect fluids, characterized by equation (3.8), these reduce to

$$\dot{\mu} = -\Theta(\mu + p), \tag{3.29}$$

the *energy conservation equation*, and the *momentum conservation equation*

$$0 = \tilde{\nabla}_a p + (\mu + p)\dot{u}_a \tag{3.30}$$

(which because of the perfect fluid assumption, has changed from a time-derivative equation for q^a to an algebraic equation for $\dot{u}_a$, and thus a time-derivative equation for u^a). These equations show that $(\mu + p)$ is both the inertial mass density and that it governs the conservation of energy. It is clear that if this quantity is zero (the case of an effective cosmological constant) or negative, the behaviour of matter will be anomalous; in particular velocities will be unstable if $\mu + p \to 0$, because the acceleration generated by a given force will diverge in this limit. If we assume a perfect fluid with a (linear) γ-law equation of state, then (3.29) shows that

$$p = (\gamma - 1)\mu,\, \dot{\gamma} = 0 \Rightarrow \mu = M/S^{3\gamma},\, \dot{M} = 0. \tag{3.31}$$

One can approximate ordinary matter in this way, with $1 \leq \gamma \leq 2$ in order that the causality and energy conditions are valid. Radiation corresponds to $\gamma = \frac{4}{3} \Rightarrow \mu = M/S^4$, so from Stefan's law ($\mu \propto T^4$) we find that $T \propto 1/S$. Another useful case is *pressure-free matter* (often described as 'baryonic' or 'cold dark matter (CDM)'); the momentum conservation: (3.30) shows that such matter moves geodesically (as expected from the equivalence principle):

$$\gamma = 1 \Leftrightarrow p = 0 \Rightarrow \dot{u}_a = 0,\, \mu = M/S^3. \tag{3.32}$$

This is the case of *pure gravitation*, without fluid dynamical effects. Another important case is that of a scalar field, see (3.7).

3.3.2 Ricci identities

The second set of equations arise from the *Ricci identities* for the vector field u^a, i.e.

$$2\nabla_{[a}\nabla_{b]}u^c = R_{ab}{}^c{}_d u^d. \tag{3.33}$$

On substituting from (3.17), using (3.2), and separating out the parallelly and orthogonally projected parts into a trace, symmetric trace-free and skew symmetric part, we obtain three propagation equations and three constraint equations. The *propagation equations* are the Raychaudhuri equation, the vorticity propagation equation and the shear propagation equation.

3.3.2.1 The Raychaudhuri equation

This equation

$$\dot{\Theta} = -\tfrac{1}{3}\Theta^2 + \nabla_a \dot{u}^a - 2\sigma^2 + 2\omega^2 - \tfrac{1}{2}(\mu + 3p) + \lambda, \tag{3.34}$$

the *basic equation of gravitational attraction* [21, 26, 28], shows the repulsive nature of a positive cosmological constant and leads to the identification of $(\mu + 3p)$ as the active gravitational mass density. Rewriting it in terms of the average scale factor S, this equation can be rewritten in the form

$$3\frac{\ddot{S}}{S} = -2(\sigma^2 - \omega^2) + \nabla_a \dot{u}^a - \frac{1}{2}(\mu + 3p) + \lambda, \tag{3.35}$$

showing how the curvature of the curve $S(\tau)$ along each world-line (in terms of proper time τ along that world-line) is determined by the shear, vorticity and acceleration; the total energy density and pressure in terms of the combination $(\mu + 3p)$—the *active gravitational mass*; and the cosmological constant λ. This gives the basic singularity theorem.

Singularity theorem. *[21, 26, 28] In a universe where the active gravitational mass is positive at all times,*

$$(\mu + 3p) > 0, \tag{3.36}$$

the cosmological constant vanishes (or is negative); $\lambda \leq 0$, and the vorticity and acceleration vanish; $\dot{u}^a = \omega^a = 0$ at all times, at any instant when $H_0 = \frac{1}{3}\Theta_0 > 0$, there must have been a time $t_0 < 1/H_0$ ago such that $S \to 0$ as $t \to t_0$; a spacetime singularity occurs there, where $\mu \to \infty$ and $T \to \infty$.

The further singularity theorems of Hawking and Penrose [68,69,124] utilize this result or its null version as an essential part of their proofs.

Closely related to this are three other results:

(1) a static universe model containing ordinary matter requires $\lambda > 0$ (Einstein's discovery of 1917);
(2) the Einstein static universe is unstable (Eddington's discovery of 1930);
(3) in a universe satisfying the requirements of the singularity theorem, at each instant t the age of the universe is less that $1/H(t)$, so for example the hot early stage of the universe takes place extremely rapidly.

Proofs follow directly from (3.35). The energy condition $(\mu + 3p) > 0$ will be satisfied by all ordinary matter but will not, in general, be satisfied by a scalar field, see (3.7).

3.3.2.2 The vorticity propagation equation

$$\dot{\omega}^{\langle a\rangle} - \tfrac{1}{2}\eta^{abc}\tilde{\nabla}_b\dot{u}_c = -\tfrac{2}{3}\Theta\omega^a + \sigma^a{}_b\omega^b. \tag{3.37}$$

If we have a barotropic perfect fluid:

$$q^a = \pi_{ab} = 0,\ p = p(\mu) \Rightarrow \eta^{abc}\tilde{\nabla}_b\dot{u}_c = 0, \tag{3.38}$$

then $\omega^a = 0$ is involutive: i.e. the statement

$$\omega^a = 0 \text{ initially} \Rightarrow \dot{\omega}^{\langle a\rangle} = 0 \Rightarrow \omega^a = 0 \text{ at later times}$$

follows from the vorticity conservation equation (3.37) (and it is also true in the special case $p = 0$). Thus non-trivial entropy dependence or an imperfect fluid is required to create vorticity.

When the vorticity vanishes $\Leftrightarrow \omega = 0$:

(1) The fluid flow is hypersurface-orthogonal, and there exists a cosmic time function t such that $u_a = -g(x^b)\nabla_a t$, allowing synchronization of the clocks of fundamental observers. If, in addition, the acceleration vanishes, we can set $g = 1$ and the time function can be proper time for all of them (whereas if the acceleration is non-zero, the coordinate time t will necessarily correspond to different proper times along different world-lines).

(2) The metric of the orthogonal 3-spaces t = constant formed by meshing together the tangent spaces orthogonal to u_a is h_{ab}.
(3) From the Gauss equation and the Ricci identities for u^a, the Ricci tensor of these 3-spaces is given by [21, 26]

$$^3R_{ab} = -\dot{\sigma}_{\langle ab\rangle} - \Theta\sigma_{ab} + \tilde{\nabla}_{\langle a}\dot{u}_{b\rangle} + \dot{u}_{\langle a}\dot{u}_{b\rangle} + \pi_{ab} + \tfrac{1}{3}h_{ab}\,{}^3R, \tag{3.39}$$

and their Ricci scalar is given by

$$^3R = 2\mu - \tfrac{2}{3}\Theta^2 + 2\sigma^2 + 2\lambda, \tag{3.40}$$

which is a generalized Friedmann equation, showing how the matter tensor determines the 3-space average curvature. These equations fully determine the curvature tensor $^3R_{abcd}$ of the orthogonal 3-spaces, and so show how the EFEs result in *spatial* curvature (as well as spacetime curvature) [21, 26].

3.3.2.3 *The shear propagation equation*

$$\dot{\sigma}^{\langle ab\rangle} - \tilde{\nabla}^{\langle a}\dot{u}^{b\rangle} = -\tfrac{2}{3}\Theta\sigma^{ab} + \dot{u}^{\langle a}\dot{u}^{b\rangle} - \sigma^{\langle a}{}_c\sigma^{b\rangle c} - \omega^{\langle a}\omega^{b\rangle} - (E^{ab} - \tfrac{1}{2}\pi^{ab}). \tag{3.41}$$

This shows how the tidal gravitational field E_{ab} directly induces shear (which then feeds into the Raychaudhuri and vorticity propagation equations, thereby changing the nature of the fluid flow), and that the anisotropic pressure term π_{ab} also generates shear in an imperfect fluid situation. Shear-free solutions are very special solutions, because (in contrast to the case of vorticity) a conspiracy of terms is required to maintain the shear zero if it is zero at any initial time (see later for a specific example).

The *constraint equations* are as follows:

(1) The (0α)-*equation*

$$0 = (C_1)^a = \tilde{\nabla}_b\sigma^{ab} - \tfrac{2}{3}\tilde{\nabla}^a\Theta + \eta^{abc}[\tilde{\nabla}_b\omega_c + 2\dot{u}_b\omega_c] + q^a, \tag{3.42}$$

shows how the momentum flux q^a (zero for a comoving perfect fluid) relates to the spatial inhomogeneity of the expansion.
(2) The *vorticity divergence identity*

$$0 = (C_2) = \tilde{\nabla}_a\omega^a - (\dot{u}_a\omega^a), \tag{3.43}$$

follows because ω^a is a curl.
(3) The H_{ab}-*equation*

$$0 = (C_3)^{ab} = H^{ab} + 2\dot{u}^{\langle a}\omega^{b\rangle} + (\mathrm{curl}\,\sigma)\omega^{ab} - (\mathrm{curl}\,\sigma)^{ab}, \tag{3.44}$$

characterizes the magnetic part of the Weyl tensor as being constructed from the 'curls' of the vorticity and shear tensors: $(\mathrm{curl}\,\omega)^{ab} = \eta^{cd\langle a}\tilde{\nabla}_c\omega^{b\rangle}{}_d$, $(\mathrm{curl}\,\sigma)^{ab} = \eta^{cd\langle a}\tilde{\nabla}_c\sigma^{b\rangle}{}_d$.

3.3.3 Bianchi identities

The third set of equations arises from the *Bianchi identities* (3.3). On using the splitting of R_{abcd} into R_{ab} and C_{abcd}, the $1+3$ splitting, (3.21),(3.25) of those quantities, and the EFE (3.2), these identities give two further propagation equations and two further constraint equations, which are similar in form to the Maxwell field equations for the electromagnetic field in an expanding universe (see [28]).

The *propagation equations* are:

$$\begin{aligned}(\dot{E}^{\langle ab\rangle} + \tfrac{1}{2}\dot{\pi}^{\langle ab\rangle}) = (\mathrm{curl}\, H)^{ab} - \tfrac{1}{2}\tilde{\nabla}^{\langle a}q^{b\rangle} - \tfrac{1}{2}(\mu+p)\sigma^{ab} - \Theta(E^{ab} + \tfrac{1}{6}\pi^{ab})\\ + 3\sigma^{\langle a}{}_c(E^{b\rangle c} - \tfrac{1}{6}\pi^{b\rangle c}) - \dot{u}^{\langle a}q^{b\rangle}\\ + \eta^{cd\langle a}[2\dot{u}_c H^{b\rangle}{}_d + \omega_c(E^{b\rangle}{}_d + \tfrac{1}{2}\pi^{b\rangle}{}_d)],\end{aligned} \tag{3.45}$$

the $\dot{E}$-*equation*, and

$$\begin{aligned}\dot{H}^{\langle ab\rangle} = &-(\mathrm{curl}\, E)^{ab} + \tfrac{1}{2}(\mathrm{curl}\,\pi)^{ab} - \Theta H^{ab} + 3\sigma^{\langle a}{}_c H^{b\rangle c}\\ &+ \tfrac{3}{2}\omega^{\langle a}q^{b\rangle} - \eta^{cd\langle a}[2\dot{u}_c E^{b\rangle}{}_d - \tfrac{1}{2}\sigma^{b\rangle}{}_c q_d - \omega_c H^{b\rangle}{}_d],\end{aligned} \tag{3.46}$$

the $\dot{H}$-*equation*, where we have defined the 'curls':

$$(\mathrm{curl}\, H)^{ab} = \eta^{cd\langle a}\tilde{\nabla}_c H^{b\rangle}{}_d, \qquad (\mathrm{curl}\, E)^{ab} = \eta^{cd\langle a}\tilde{\nabla}_c E^{b\rangle}{}_d. \tag{3.47}$$

These equations show how gravitational radiation arises: as in the electromagnetic case, taking the time derivative of the $\dot{E}$-equation gives a term of the form $(\mathrm{curl}\, H)\dot{}$; commuting the derivatives and substituting from the $\dot{H}$-equation eliminates H, and results in a term in $\ddot{E}$ and a term of the form $(\mathrm{curl}\,\mathrm{curl}\, E)$, which together give the wave operator acting on E [20, 66]. Similarly the time derivative of the $\dot{H}$-equation gives a wave equation for H, and associated with these is a wave equation for the shear σ.

The *constraint equations* are

$$\begin{aligned}0 = (C_4)^a = \tilde{\nabla}_b(E^{ab} + \tfrac{1}{2}\pi^{ab}) - \tfrac{1}{3}\tilde{\nabla}^a\mu + \tfrac{1}{3}\Theta q^a\\ - \tfrac{1}{2}\sigma^a{}_b q^b - 3\omega_b H^{ab} - \eta^{abc}[\sigma_{bd}H^d{}_c - \tfrac{3}{2}\omega_b q_c],\end{aligned} \tag{3.48}$$

the (div E)-*equation* with its source the spatial gradient of the energy density and

$$\begin{aligned}0 = (C_5)^a = \tilde{\nabla}_b H^{ab} + (\mu+p)\omega^a + 3\omega_b(E^{ab} - \tfrac{1}{6}\pi^{ab})\\ + \eta^{abc}[\tfrac{1}{2}\tilde{\nabla}_b q_c + \sigma_{bd}(E^d{}_c + \tfrac{1}{2}\pi^d{}_c)],\end{aligned} \tag{3.49}$$

the (div H)-*equation*, with its source the fluid vorticity. The (div E)-equation can be regarded as a (vector) analogue of the Newtonian Poisson equation [52], leading to the Newtonian limit and enabling tidal action at a distance. These equations respectively show that, generically, scalar modes will result in a non-zero divergence of E_{ab} (and hence a non-zero E-field) and vector modes in a non-zero divergence of H_{ab} (and hence a non-zero H-field).

3.3.4 Implications

Altogether, we have six propagation equations and six constraint equations; considered as a set of evolution equations for the 1+3 covariant variables, they are a first-order system of equations. This set is determinate once the fluid equations of state are given; together they then form a dynamical system (the set closes up, but is essentially an infinite dimensional dynamical system because of the spatial derivatives that occur).

The *key issue* that arises is consistency of the constraints with the evolution equations. It is believed that they are *generally consistent* for physically reasonable and well-defined equations of state, i.e. they are consistent if no restrictions are placed on their evolution other than those implied by the constraint equations and the equations of state (this has been shown for irrotational dust [91]). It is this that makes consistent the overall hyperbolic nature of the equations with the 'instantaneous' action at a distance implicit in the Gauss-like equations (specifically, the (div E)-equation), the point being that the 'action at a distance' nature of the solutions to these equations is built into the initial data, which must be chosen so that the constraints are satisfied initially, and they then remain satisfied thereafter because the time evolution preserves these constraints (cf [49]).

3.3.5 Shear-free dust

One must be very cautious with imposing simplifying assumptions in order to obtain solutions: this can lead to major restrictions on the possible flows, and one can be badly misled if their consistency is not investigated carefully. A case of particular interest is *shear-free dust*, that is perfect-fluid solutions for which $\sigma_{ab} = 0$, $p = 0 \Rightarrow \dot{u}^a = 0$. In this case, careful study of the consistency conditions between all the equations [25] shows that necessarily $\omega\Theta = 0$: the solutions either do not rotate, or do not expand. This conclusion is of considerable importance, because if it were not true, there would be shear-free expanding and rotating solutions which would violate the Hawking–Penrose singularity theorems for cosmology [68,69] (integrating the vorticity equation along the fluid flow lines (3.37) gives $\omega = \omega_0/S^2$; substituting in the Raychaudhuri equation (3.34) and integrating, using the conservation equation (3.29), gives a first integral which is a generalized Friedmann equation, in which vorticity dominates expansion at early times and allows a bounce and singularity avoidance). The interesting point then is that *this result does not hold in Newtonian theory* [113], in which case there do indeed exist such solutions when suitable boundary conditions are imposed. If one uses these solutions as an argument against the singularity theorems, the argument is invalid; what they really do is point out the dangers of the Newtonian limit of cosmological equations.

3.4 Tetrad description

The 1+3 covariant equations are immediately transparent in terms of representing relations between 1 + 3 covariantly defined quantities with clear geometrical and/or physical significance. However, they do not form a complete set of equations guaranteeing the existence of a corresponding metric and connection. For that we need to use a full tetrad description. The equations determined will then form a complete set, which will contain as a subset all the 1 + 3 covariant equations just derived (albeit presented in a slightly different form) [53,55]. First we summarize a generic tetrad formalism, and then describe its application to cosmological models (cf [25,92]).

3.4.1 General tetrad formalism

A *tetrad* is a set of four linearly independent vector fields $\{\boldsymbol{e}_a\}$, $a = 0, 1, 2, 3$, which serves as a basis for spacetime vectors and tensors. It can be written in terms of a local coordinate basis by means of the *tetrad components* $e_a{}^i(x^j)$:

$$\boldsymbol{e}_a = e_a{}^i(x^j)\frac{\partial}{\partial x^i} \Leftrightarrow \boldsymbol{e}_a(f) = e_a{}^i(x^j)\frac{\partial f}{\partial x^i}, \qquad e_a{}^i \equiv \boldsymbol{e}_a(x^i), \tag{3.50}$$

(the latter stating that the ith component of the ath tetrad vector is just the directional derivative of the ith coordinate x^i in the direction $\boldsymbol{e}_a$). This relation can be thought of as just a change of vector basis, leading to a change of tensor components of the standard tensorial form:

$$T^{ab}{}_{cd} = e^a{}_i e^b{}_j e_c{}^k e_d{}^l T^{ij}{}_{kl}$$

with an obvious inverse, where the inverse components $e^a{}_i(x^j)$ (note the placing of the indices!) are defined by

$$e_a{}^i e^a{}_j = \delta^i{}_j \Leftrightarrow e_a{}^i e^b{}_i = \delta^b{}_a. \tag{3.51}$$

However, this is a change from an integrable basis to a non-integrable one, so the non-tensorial relations (specifically the form of the metric and connection components) differ slightly from when coordinate bases are used. A change of one tetrad basis to another will also lead to transformations of the standard tensor form for all tensorial quantities: if $\boldsymbol{e}_a = \lambda_a{}^{a'}(x^i)\boldsymbol{e}_{a'}$ is a change of tetrad basis with inverse $\boldsymbol{e}_{a'} = \lambda_{a'}{}^a(x^i)\boldsymbol{e}_a$ (in the case of orthonormal bases, each of these matrices representing a Lorentz transformation), then

$$T^{ab}{}_{cd} = \lambda_{a'}{}^a \lambda_{b'}{}^b \lambda_c{}^{c'} \lambda_d{}^{d'} T^{a'b'}{}_{c'd'}.$$

Again the inverse is obvious. The *commutation functions* related to the tetrad are the quantities $\gamma^a{}_{bc}(x^i)$ defined by the *commutators* $[\boldsymbol{e}_a, \boldsymbol{e}_b]$ of the basis vectors:

$$[\boldsymbol{e}_a, \boldsymbol{e}_b] = \gamma^c{}_{ab}(x^i)\boldsymbol{e}_c \Rightarrow \gamma^a{}_{bc}(x^i) = -\gamma^a{}_{cb}(x^i). \tag{3.52}$$

It follows (apply this relation to the coordinate x^i) that in terms of the tetrad components,

$$\gamma^a{}_{bc}(x^i) = e^a{}_i(e_b{}^j \partial_j e_c{}^i - e_c{}^j \partial_j e_b{}^i) = -2e_b{}^i e_c{}^j \nabla_{[i} e^a{}_{j]}. \tag{3.53}$$

These quantities vanish iff the basis $\{\boldsymbol{e}_a\}$ is a coordinate basis: that is, there exist coordinates x^i such that $\boldsymbol{e}_a = \delta_a{}^i \partial/\partial x^i$, iff

$$[\boldsymbol{e}_a, \boldsymbol{e}_b] = 0 \Leftrightarrow \gamma^a{}_{bc} = 0.$$

The *metric tensor components* in the tetrad form are given by

$$g_{ab} = g_{ij} e_a{}^i e_b{}^j = \boldsymbol{e}_a \cdot \boldsymbol{e}_b. \tag{3.54}$$

The inverse equation

$$g_{ij}(x^k) = g_{ab} e^a{}_i(x^k) e^b{}_j(x^k) \tag{3.55}$$

explicitly constructs the coordinate components of the metric from the (inverse) tetrad components $e^a{}_i(x^j)$. We can raise and lower tetrad indices by use of the metric g_{ab} and its inverse g^{ab}. In the case of an orthonormal tetrad,

$$g_{ab} = \text{diag}(-1, +1, +1, +1) = g^{ab}, \tag{3.56}$$

showing by (3.54) that the basis vectors are unit vectors orthogonal to each other. Such a tetrad is defined up to an arbitrary position-dependent Lorentz transformation.

The *connection components* $\Gamma^a{}_{bc}$ for the tetrad are defined by the relations

$$\nabla_{\boldsymbol{e}_b} \boldsymbol{e}_a = \Gamma^c{}_{ab} \boldsymbol{e}_c \Leftrightarrow \Gamma^c{}_{ab} = e^c{}_i e_b{}^j \nabla_j e_a{}^i, \tag{3.57}$$

i.e. it is the c-component of the covariant derivative in the b-direction of the a-vector. It follows that all covariant derivatives can be written out in tetrad components in a way completely analogous to the usual tensor form, for example

$$\nabla_a T_{bc} = \boldsymbol{e}_a(T_{bc}) - \Gamma^d{}_{ba} T_{dc} - \Gamma^d{}_{ca} T_{bd},$$

where for any function f, $\boldsymbol{e}_a(f) = e_a{}^i \partial f/\partial x^i$ is the derivative of f in the direction $\boldsymbol{e}_a$. In the case of an orthonormal tetrad, (3.56) shows that $\boldsymbol{e}_a(g_{bc}) = 0$; hence applying this relation to the metric tensor,

$$\nabla_a g_{bc} = 0 \Leftrightarrow \Gamma_{(ab)c} = 0, \tag{3.58}$$

—the connection components are skew in their first two indices, when we use the metric to raise and lower the first indices only, and are called 'Ricci rotation coefficients' or just *rotation coefficients*. We obtain from this and the assumption of vanishing torsion the relations for an orthonormal tetrad that are the analogues of the usual Christoffel relation:

$$\gamma^a{}_{bc} = -(\Gamma^a{}_{bc} - \Gamma^a{}_{cb}), \qquad \Gamma_{abc} = \tfrac{1}{2}(g_{ad}\gamma^d{}_{cb} - g_{bd}\gamma^d{}_{ca} + g_{cd}\gamma^d{}_{ab}). \quad (3.59)$$

This shows that the rotation coefficients and the commutation functions are each just linear combinations of the other.

Any set of vectors however must satisfy the *Jacobi identities*:

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0,$$

which follows from the definition of a commutator. Applying this to the basis vectors $\boldsymbol{e}_a$, $\boldsymbol{e}_\mathrm{b}$ and $\boldsymbol{e}_c$ gives the identities

$$\boldsymbol{e}_{[a}(\gamma^d{}_{bc]}) + \gamma^e{}_{[ab}\gamma^d{}_{c]e} = 0, \quad (3.60)$$

which are the integrability conditions that the $\gamma^a{}_{bc}(x^i)$ are the commutation functions for the set of vectors $\boldsymbol{e}_a$.

If we apply the Ricci identities to the tetrad basis vectors $\boldsymbol{e}_a$, we obtain the Riemann curvature tensor components in the form

$$R^a{}_{bcd} = \partial_c(\Gamma^a{}_{bd}) - \partial_d(\Gamma^a{}_{bc}) + \Gamma^a{}_{ec}\Gamma^e{}_{bd} - \Gamma^a{}_{ed}\Gamma^e{}_{bc} - \Gamma^a{}_{be}\gamma^e{}_{cd}. \quad (3.61)$$

Contracting this on a and c, one obtains the EFE in the form

$$R_{bd} = \partial_a(\Gamma^a{}_{bd}) - \partial_d(\Gamma^a{}_{ba}) + \Gamma^a{}_{ea}\Gamma^e{}_{bd} - \Gamma^a{}_{de}\Gamma^e{}_{ba} = T_{bd} - \tfrac{1}{2}T g_{bd} + \lambda g_{bd}. \quad (3.62)$$

It is not immediately obvious that this is symmetric, but this follows because (3.60) implies $R_{a[bcd]} = 0 \Rightarrow R_{ab} = R_{(ab)}$.

3.4.2 Tetrad formalism in cosmology

In detailed studies of families of exact non-vacuum solutions, it will usually be advantageous to use an orthonormal tetrad basis, because the tetrad vectors can be chosen in physically preferred directions. For a cosmological model we choose an orthonormal tetrad with the timelike vector $\boldsymbol{e}_0$ chosen to be either the fundamental 4-velocity field u^a, or the normals n^a to surfaces of homogeneity when they exist. This fixing implies that the initial six-parameter freedom of using Lorentz transformations has been reduced to a three-parameter freedom of rotations of the spatial frame $\{\boldsymbol{e}_\alpha\}$. The 24 algebraically independent rotation coefficients can then be split into (see [25,45,53]):

$$\Gamma_{\alpha 00} = \dot{u}_\alpha, \qquad \Gamma_{\alpha 0\beta} = \tfrac{1}{3}\Theta\delta_{\alpha\beta} + \sigma_{\alpha\beta} - \epsilon_{\alpha\beta\gamma}\omega^\gamma, \qquad \Gamma_{\alpha\beta 0} = \epsilon_{\alpha\beta\gamma}\Omega^\gamma \quad (3.63)$$

$$\Gamma_{\alpha\beta\gamma} = 2a_{[\alpha}\delta_{\beta]\gamma} + \epsilon_{\gamma\delta[\alpha}n^\delta{}_{\beta]} + \tfrac{1}{2}\epsilon_{\alpha\beta\delta}n^\delta{}_\gamma. \quad (3.64)$$

The first two sets contain the kinematical variables for the chosen vector field. The third is the rate of rotation Ω^α of the spatial frame $\{\boldsymbol{e}_\alpha\}$ with respect to a Fermi-propagated (physically non-rotating) basis along the fundamental flow lines. Finally, the quantities a^α and $n_{\alpha\beta} = n_{(\alpha\beta)}$ determine the nine spatial rotation coefficients. In terms of these quantities, the commutator equations (3.52) applied to any function f take the form

$$[\boldsymbol{e}_0, \boldsymbol{e}_\alpha](f) = \dot{u}_\alpha \boldsymbol{e}_0(f) - [\tfrac{1}{3}\Theta\delta_\alpha{}^\beta + \sigma_\alpha{}^\beta + \epsilon_\alpha{}^\beta{}_\gamma(\omega^\gamma + \Omega^\gamma)]\boldsymbol{e}_\beta(f), \quad (3.65)$$

$$[\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta](f) = 2\epsilon_{\alpha\beta\gamma}\omega^\gamma \boldsymbol{e}_0(f) + [2a_{[\alpha}\delta^\gamma{}_{\beta]} + \epsilon_{\alpha\beta\delta}n^{\delta\gamma}]\boldsymbol{e}_\gamma(f). \quad (3.66)$$

3.4.3 Complete set

The full set of equations for a gravitating fluid can be written in tetrad form, using the matter variables, the rotation coefficients (3.57) and the tetrad components (3.50) as the primary variables. The equations needed are the conservation equations (3.27), (3.28) and all the Ricci equations (3.61) and Jacobi identities (3.60) for the tetrad basis vectors, together with the tetrad equations (3.50) and the commutator equations (3.53). This gives a set of constraints and a set of first-order evolution equations, which include the tetrad form of all the $1+3$ covariant equations given earlier, based on the chosen vector field. For a prescribed set of equations of state, this gives the complete set of relations needed to determine the spacetime structure. One has the option of including or not including the tetrad components of the Weyl tensor as variables in this set; whether it is better to include them or not depends on the problem to be solved (if they are included, there will be more equations in the corresponding complete set, for we must then include the full Bianchi identities). The full set of equations is given in [41, 55], and see [25, 118] for the use of tetrads to study locally rotationally symmetric spacetimes, and [45, 128] for the case of Bianchi universes.

Finally, when tetrad vectors are chosen uniquely in an invariant way (e.g. as eigenvectors of a non-degenerate shear tensor), then—because they are uniquely defined from $1 + 3$ covariant quantities—all the rotation coefficients are covariantly defined scalars, so these equations are all equations for scalar invariants. The only times when it is not possible to define unique tetrads in this way is when the spacetimes are isotropic or locally rotationally symmetric (these concepts are discussed later).

3.5 Models and symmetries

3.5.1 Symmetries of cosmologies

Symmetries of a space or a spacetime (generically, 'space') are transformations of the space into itself that leave the metric tensor and all physical and geometrical properties invariant. We deal here only with continuous symmetries, characterized by a continuous group of transformations and associated vector fields [24].

3.5.1.1 Killing vectors

A space or spacetime *symmetry*, or *isometry*, is a transformation that drags the metric along a certain congruence of curves into itself. The generating vector field ξ_i of such curves is called a *Killing vector (field)* (or 'KV'), and obeys Killing's equations,

$$(L_\xi g)_{ij} = 0 \Leftrightarrow \nabla_{(i}\xi_{j)} = 0 \Leftrightarrow \nabla_i \xi_j = -\nabla_j \xi_i, \tag{3.67}$$

where L_X is the *Lie derivative*. By the Ricci identities for a KV, this implies the curvature equation:

$$\nabla_i \nabla_j \xi_k = R^m{}_{ijk}\xi_m, \tag{3.68}$$

and hence the infinite series of further equations that follows by taking covariant derivatives of this one, e.g.

$$\nabla_l \nabla_i \nabla_j \xi_k = (\nabla_l R^m{}_{ijk})\xi_m + R^m{}_{ijk}\nabla_l \xi_m. \tag{3.69}$$

The set of all KVs forms a Lie algebra with a basis $\{\xi_a\}$, $a = 1, 2, \ldots, r$, of dimension $r \leq \frac{1}{2}n(n-1)$. ξ^i_a denotes the components with respect to a local coordinate basis, a, b and c label the KV basis and i, j and k the coordinate components. Any KV can be written in terms of this basis, with *constant coefficients*. Hence, if we take the commutator $[\xi_a, \xi_\mathrm{b}]$ of two of the basis KVs, this is also a KV, and so can be written in terms of its components relative to the KV basis, which will be constants. We can write the constants as $C^c{}_{ab}$, obtaining

$$[\xi_a, \xi_\mathrm{b}] = C^c{}_{ab}\xi_c, \qquad C^a{}_{bc} = C^a{}_{[bc]}. \tag{3.70}$$

By the Jacobi identities for the basis vectors, these structure constants must satisfy

$$C^a{}_{e[b}C^e{}_{cd]} = 0 \tag{3.71}$$

(which is just equation (3.60) specialized to a set of vectors with constant commutation functions). These are the integrability conditions that must be satisfied in order that the Lie algebra exist in a consistent way. The transformations generated by the Lie algebra form a Lie group of the same dimension (see Eisenhart [24] or Cohn [11]).

Arbitrariness of the basis: We can change the basis of KVs in the usual way;

$$\xi_{a'} = \lambda_{a'}{}^a \xi_a \Leftrightarrow \xi^i_{a'} = \lambda_{a'}{}^a \xi^i_a, \tag{3.72}$$

where the $\lambda_{a'}{}^a$ are constants with $\det(\lambda_{a'}{}^a) \neq 0$, so unique inverse matrices $\lambda^{a'}{}_a$ exist. Then the structure constants transform as tensors:

$$C^{c'}{}_{a'b'} = \lambda^{c'}{}_c \lambda_{a'}{}^a \lambda_{b'}{}^b C^c{}_{ab}. \tag{3.73}$$

Thus the possible equivalence of two Lie algebras is not obvious, as they may be given in different bases.

3.5.1.2 Groups of isometries

The isometries of a space of dimension n must be a group, as the identity is an isometry, the inverse of an isometry is an isometry, and the composition of two isometries is an isometry. Continuous isometries are generated by the Lie algebra of KVs. The group structure is determined locally by the Lie algebra, in turn characterized by the structure constants [11]. The action of the group is characterized by the nature of its orbits in space; this is only partially determined by the group structure (indeed the same group can act as a spacetime symmetry group in quite different ways).

3.5.1.3 Dimensionality of groups and orbits

Most spaces have no KVs, but special spaces (with symmetries) have some. The group action defines orbits in the space where it acts and the dimensionality of these orbits determines the kind of symmetry that is present.

The *orbit* of a point p is the set of all points into which p can be moved by the action of the isometries of a space. Orbits are necessarily homogeneous (all physical quantities are the same at each point). An *invariant variety* is a set of points moved into itself by the group. This will be bigger than (or equal to) all orbits it contains. The orbits are necessarily invariant varieties; indeed they are sometimes called *minimum invariant varieties*, because they are the smallest subspaces that are always moved into themselves by all the isometries in the group. *Fixed points* of a group of isometries are those points which are left invariant by the isometries (thus the orbit of such a point is just the point itself). These are the points where all KVs vanish (however, the derivatives of the KVs there are non-zero; the KVs generate isotropies about these points). *General points* are those where the dimension of the space spanned by the KVs (that is, the dimension of the orbit through the point) takes the value it has almost everywhere; *special points* are those where it has a lower dimension (e.g. fixed points). Consequently, the dimension of the orbits through special points is lower than that of orbits through general points. The dimension of the orbit and isotropy group is the same at each point of an orbit, because of the equivalence of the group action at all points on each orbit.

The group is *transitive on a surface S* (of whatever dimension) if it can move any point of S into any other point of S. Orbits are the largest surfaces through each point on which the group is transitive; they are therefore sometimes referred to as *surfaces of transitivity*. We define their dimension as follows, and determine limits from the maximal possible initial data for KVs: *dimension of the surface of transitivity* $= s$, where in a space of dimension n, $s \leq n$.

At each point we can also consider the dimension of the isotropy group (the group of isometries leaving that point fixed), generated by all those KVs that vanish at that point: *dimension of an isotropy group* $= q$, where $q \leq \frac{1}{2}n(n-1)$.

The *dimension r of the group of symmetries* of a space of dimension n is $r = s + q$ (translations plus rotations). The dimension q of the isotropy group can vary over the space (but not over an orbit): it can be greater at special points (e.g. an axis centre of symmetry) where the dimension s of the orbit is less, but r (the dimension of the total symmetry group) must stay the same everywhere. From these limits , $0 \leq r \leq n + \frac{1}{2}n(n-1) = \frac{1}{2}n(n+1)$ (the maximal number of translations and of rotations). This shows the Lie algebra of KVs is finite dimensional.

Maximal dimensions: If $r = \frac{1}{2}n(n+1)$, we have a space(time) of constant curvature (maximal symmetry for a space of dimension n). In this case,

$$R_{ijkl} = K(g_{ik}g_{jl} - g_{il}g_{jk}), \tag{3.74}$$

with K a constant. One cannot get $q = \frac{1}{2}n(n-1) - 1$ so $r \neq \frac{1}{2}n(n+1) - 1$.

A group is *simply transitive* if $r = s \Leftrightarrow q = 0$ (no redundancy: dimensionality of group of isometries is just sufficient to move each point in a surface of transitivity into each other point). There is no continuous isotropy group.

A group is *multiply transitive* if $r > s \Leftrightarrow q > 0$ (there is redundancy in that the dimension of the group of isometries is larger than is necessary to move each point in an orbit into each other point). There exist non-trivial isotropies.

3.5.2 Classification of cosmological symmetries

We consider non-empty perfect fluid models, i.e. (3.6) holds with $(\mu + p) > 0$, implying u^a is the uniquely defined timelike eigenvector of the Ricci tensor.

Spacetime is four-dimensional, so the possibilities for the dimension of the surface of transitivity are $s = 0, 1, 2, 3, 4$. Because u^a is invariant, the isotropy group at each point has to be a sub-group of the rotations $O(3)$ acting orthogonally to u^a, but there is no two-dimensional subgroup of $O(3)$. Thus the possibilities for isotropy at a general point are:

(1) *Isotropic*: $q = 3$, the matter is a perfect fluid, the Weyl tensor vanishes, all kinematical quantities vanish except Θ. All observations (at every point) are isotropic. This is the RW family of geometries.
(2) *Local rotational symmetry* ('LRS'): $q = 1$, the Weyl tensor is of algebraic Petrov type D, kinematical quantities are rotationally symmetric about a preferred spatial direction. All observations at every general point are rotationally symmetric about this direction. All metrics are known in the case of dust [25] and a perfect fluid [51, 118].
(3) *Anisotropic*: $q = 0$; there are no rotational symmetries. Observations in each direction are different from observations in each other direction.

Putting this together with the possibilities for the dimensions of the surfaces of transitivity, we have the following possibilities (see table 3.1).

Table 3.1. Classification of cosmological models (with $(\mu + p) > 0$) by isotropy and homogeneity.

Dimension, Isotropy group	Dim invariant variety: $s = 2$ Inhomogeneous	$s = 3$ Spatially homogeneous	$s = 4$ Spacetime homogeneous
$q = 0$ anisotropic	Generic metric form known. Spatially self-similar, Abelian G_2 on 2D spacelike surfaces, non-Abelian G_2	Bianchi: orthogonal, tilted	Osvath/Kerr
$q = 1$ LRS	Lemaître–Tolman–Bondi family	Kantowski–Sachs, LRS Bianchi	Gödel
$q = 3$ isotropic	None (cannot happen)	Friedmann	Einstein static
	Two non-ignorable coordinates	One non-ignorable coordinate	Algebraic EFE (no redshift)

Dimension Isotropy group	Dim invariant variety: $s = 0$ Inhomogeneous	$s = 1$ Inhomogeneous/ no isotropy group
$q = 0$	Szekeres–Szafron, Stephani–Barnes, Oleson type N	General metric form independent of one coord; KV h.s.o./not h.s.o.
	The real universe!	

3.5.2.1 Spacetime homogeneous models

These models with $s = 4$ are unchanging in space and time, hence μ is a constant, so by the energy conservation equation (3.29) they cannot expand: $\Theta = 0$. They cannot produce an almost isotropic redshift, and are not useful as models of the real universe. Nevertheless they are of some interest for their geometric properties.

The *isotropic case* $q = 3$ ($\Rightarrow r = 7$) is the Einstein static universe, the non-expanding FL model that was the first relativistic cosmological model found. It is not a viable cosmology because it has no redshifts, but it laid the foundation for the discovery of the expanding FLRW models.

The *LRS case* $q = 1$ ($\Rightarrow r = 5$) is the Gödel stationary rotating universe [60], also with no redshifts. This model was important because of the new understanding it brought as to the nature of time in general relativity (see [68, 124]). It is a model in which causality is violated (there exist closed timelike lines through each spacetime point) and there exists no cosmic time function whatsoever.

The anisotropic models $q = 0$ ($\Rightarrow r = 4$) are all known, but are interesting only for the light they shed on Mach's principle; see [101].

3.5.2.2 Spatially homogeneous universes

These models with $s = 3$ are the major models of theoretical cosmology, because they express mathematically the idea of the 'cosmological principle': all points of space at the same time are equivalent to each other [6].

The *isotropic case* $q = 3$ ($\Rightarrow r = 6$) is the family of FL models, the standard models of cosmology, with the comoving RW metric form:

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + S^2(t)(\mathrm{d}r^2 + f^2(r)(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)), \qquad u^a = \delta^a_0. \tag{3.75}$$

Here the space sections are of constant curvature $K = k/S^2$ and

$$f(r) = \sin r, r, \sinh r \tag{3.76}$$

if the normalized spatial curvature k is $+1, 0, -1$ respectively. The space sections are necessarily closed if $k = +1$.

The *LRS case* $q = 1$ ($\Rightarrow r = 4$) is the family of Kantowski–Sachs universes [13,80] plus the LRS orthogonal [45] and tilted [77] Bianchi models. The simplest are the Kantowski–Sachs family, with comoving metric form

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + A^2(t)\,\mathrm{d}r^2 + B^2(t)(\mathrm{d}\theta^2 + f^2(\theta)\,\mathrm{d}\phi^2), \qquad u^a = \delta^a_0, \tag{3.77}$$

where $f(\theta)$ is given by (3.76).

The *anisotropic case* $q = 0$ ($\Rightarrow r = 3$) is the family of Bianchi universes with a group of isometries G_3 acting simply transitively on spacelike surfaces. They can be orthogonal or tilted. The simplest class is the Bianchi type I family, with an Abelian isometry group and metric form:

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + A^2(t)\,\mathrm{d}x^2 + B^2(t)\,\mathrm{d}y^2 + C^2(t)\,\mathrm{d}z^2, \qquad u^a = \delta^a_0. \tag{3.78}$$

The family as a whole has quite complex properties; these models are discussed in the following section.

3.5.2.3 Spatially inhomogeneous universes

These models have $s \leq 2$. The *LRS cases* ($q = 1 \Rightarrow s = 2, r = 3$) are the spherically symmetric family with metric form:

$$ds^2 = -C^2(t,r)\,\mathrm{d}t^2 + A^2(t,r)\,\mathrm{d}r^2 + B^2(t,r)(\mathrm{d}\theta^2 + f^2(\theta)\,\mathrm{d}\phi^2), \qquad u^a = \delta^a_0, \tag{3.79}$$

where $f(\theta)$ is given by (3.76). In the dust case, we can set $C(t,r) = 1$ and can integrate the EFE analytically; for $k = +1$, these are the ('LTB') spherically symmetric models [5,87]. They may have a centre of symmetry (a timelike world-line), and can even allow two such centres, but they cannot be isotropic about a general point (because isotropy everywhere implies spatial homogeneity).

Solutions with no symmetries at all have $r = 0 \Rightarrow s = 0$, $q = 0$. The real universe, of course, belongs to this class; all the others are intended as approximations to this unique universe. Remarkably, we know some exact solutions without any symmetries, specifically (a) the Szekeres quasi-spherical models [121, 122], (b) Stephani's conformally flat models [84, 116], and (c) Oleson's type-N solutions (for a discussion of these and all the other inhomogeneous models, see Krasiński [85] and Kramer *et al* [83]). One further interesting family without global symmetries are the 'Swiss-cheese' models, made by cutting and pasting segments of spherically symmetric models [23,112].

Because of the nonlinearity of the equations, it is helpful to have exact solutions at hand as models of structure formation as well as studies of linearly perturbed FL models (briefly discussed later). The dust (Tolman–Bondi) and perfect fluid spherically symmetric models are useful here, in particular in terms of relating the time evolution to self-similar models. However, in the fully nonlinear regime numerical solutions of the full equations are needed.

3.6 Friedmann–Lemaître models

The FL models are discussed in detail in other chapters, so here I will only briefly mention some interesting properties of these solutions (and see also [33]). These models are perfect fluid solutions with metric form (3.75), characterized by

$$\dot{u} = 0 = \omega = \sigma = 0, \qquad \theta = 3\frac{\dot{S}}{S} \tag{3.80}$$

$$\Rightarrow \tilde{\nabla}_e \mu = \tilde{\nabla}_e p = \tilde{\nabla}_e \theta = 0, \qquad E_{ab} = H_{ab} = 0. \tag{3.81}$$

They are isotropic about every point ($q = 3$) and consequently are spatially homogeneous ($s = 3$). The equations that apply are the covariant equations (3.80), (3.83) with restrictions (3.80). The dynamical equations are the energy equation (3.29)

$$\dot{\mu} = -3\frac{\dot{S}}{S}(\mu + p), \tag{3.82}$$

the Raychaudhuri equation (3.34):

$$3\frac{\ddot{S}}{S} = -\tfrac{1}{2}(\mu + 3p) + \lambda \tag{3.83}$$

and the Friedmann equation (3.40), where ${}^3R = 6k/S^2$,

$$3\frac{\dot{S}^2}{S^2} - \kappa\mu - \lambda = -\frac{3k}{S^2}. \tag{3.84}$$

The Friedmann equation is a first integral of the other two when $\dot{S} \neq 0$. The solutions, of course, depend on the equation of state; for the currently favoured universe models, going backward in time there will be

(1) a cosmological-constant-dominated phase,
(2) a matter-dominated phase,
(3) a radiation-dominated phase,
(4) a scalar-field-dominated inflationary phase and
(5) a pre-inflationary phase where the physics is speculative (see the last section of this chapter). The normalized density parameter is $\Omega \equiv \kappa\mu/3H^2$, where as usual $H = \dot{S}/S$.

3.6.1 Phase planes and evolutionary paths

From these equations, one can obtain phase planes

(i) for the density parameter Ω against the deceleration parameter q, see [115];
(ii) for the density parameter Ω against the Hubble parameter H, see [128] for the case $\lambda = 0$; and
(iii) for the density parameter Ω against the scale parameter S, see [94], showing how Ω changes in inflationary and non-inflationary universes.

It is a consequence of the equations that the spatial curvature parameter k is a constant of the motion. In particular, flatness cannot change as the universe evolves: either $k = 0$ or not, depending on the initial conditions, and this is independent of any inflation that may take place. Thus while inflation can drive the spatial curvature $K = k/S^2$ very close indeed to zero, it cannot set $K = 0$.

If one has a scalar field matter source ϕ with potential $V(\phi)$, one can obtain essentially arbitrary functional forms for the scale function $S(t)$ by using the arbitrariness in the function $V(\phi)$ and running the field equations backwards, see [46].

3.6.2 Spatial topology

The Einstein field equations determine the time evolution of the metric and its spatial curvature, but they do not determine its spatial topology. Spatially closed

FL models can occur even if $k = 0$ or $k = -1$, for example with a toroidal topology [27]. These universes can be closed on a small enough spatial scale that we could have seen all the matter in the universe already, and indeed could have seen it many times over; see the discussion on 'small universes' later.

3.6.3 Growth of inhomogeneity

This is studied by looking at linear perturbations of the FL models, as well as by examining inhomogeneous models. The geometry and dynamics of perturbed FL models is described in detail in other talks, so I will again just make a few remarks. In dealing with perturbed FL models, one runs into the *gauge issue*: the background model is not uniquely defined by a realistic (lumpy) model and the definition of the perturbations depends on the choice of background model (the gauge chosen). Consequently it is advisable to use gauge-invariant variables, either coordinate-based [2] or covariant [39]. When dealing with multiple matter components, it is important to take carefully into account the separate velocities needed for each matter component, and their associated conservation equations. The CBR can best be described by kinetic theory, which again can be presented in a covariant and gauge invariant way [10, 59].

3.7 Bianchi universes ($s = 3$)

These are the models in which there is a group of isometries G_3 simply transitive on spacelike surfaces, so they are spatially homogeneous. There is only one essential dynamical coordinate (the time t) and the EFE reduce to ordinary differential equations, because the inhomogeneous degrees of freedom have been 'frozen out'. They are thus quite special in geometrical terms; nevertheless, they form a rich set of models where one can study the exact dynamics of the full nonlinear field equations. The solutions to the EFE will depend on the matter in the spacetime. In the case of a fluid (with uniquely defined flow lines), we have two different kinds of models:

(1) *Orthogonal models*, with the fluid flow lines orthogonal to the surfaces of homogeneity (Ellis and MacCallum [45], see also [128]). In this case the fluid 4-velocity u^a is *parallel* to the normal vectors n^a so the matter variables will be just the fluid density and pressure. The fluid flow is necessarily irrotational and geodesic.

(2) *Tilted models*, with the fluid flow lines not orthogonal to the surfaces of homogeneity. Thus the fluid 4-velocity is *not parallel* to the normals, and the components of the fluid peculiar velocity enter as further variables (King and Ellis [15, 77]). They determine the fluid energy–momentum tensor components relative to the normal vectors (a perfect fluid will appear as an imperfect fluid in that frame). Rotating models *must* be tilted, and are much more complex than non-rotating models.

3.7.1 Constructing Bianchi universes

The approach of Ellis and MacCallum [45]) uses an orthonormal tetrad based on the normals to the surfaces of homogeneity (i.e. $\boldsymbol{e}_0 = \boldsymbol{n}$, the unit normal vector to these surfaces). The tetrad is chosen to be invariant under the group of isometries, i.e. the tetrad vectors commute with the KVs. Then we have an orthonormal basis $\boldsymbol{e}_a$, $a = 0, 1, 2, 3$, such that equation (3.52) becomes

$$[\boldsymbol{e}_a, \boldsymbol{e}_{\mathrm{b}}] = \gamma^c{}_{ab}(t)\boldsymbol{e}_c \tag{3.85}$$

and all dynamic variables are function of time t only. The matter variables—$\mu(t)$, $p(t)$, and $u_\alpha(t)$ in the case of tilted models—and the commutation functions $\gamma^a{}_{bc}(t)$, which by (3.59) are equivalent to the rotation coefficients, are chosen to be these variables. The EFE (3.2) are first-order equations for these quantities, supplemented by the Jacobi identities for the $\gamma^a{}_{bc}(t)$, which are also first-order equations. Thus the equations needed are just the tetrad equations mentioned in section 3.3, for the case

$$\dot{u}^\alpha = \omega^\alpha = 0 = \boldsymbol{e}_\alpha(\gamma^a{}_{bc}). \tag{3.86}$$

The spatial commutation functions $\gamma^\alpha{}_{\beta\gamma}(t)$ can be decomposed into a time-dependent matrix $n_{\alpha\beta}(t)$ and vector $a^\alpha(t)$, see (3.66), and are equivalent to the structure constants $C^\alpha{}_{\beta\gamma}$ of the symmetry group at each point. In view of (3.86), the Jacobi identities (3.60) for the spatial vectors now take the simple form

$$n^{\alpha\beta}a_\beta = 0. \tag{3.87}$$

The tetrad basis can be chosen to diagonalize $n_{\alpha\beta}$ at all times, to attain $n_{\alpha\beta} = \mathrm{diag}(n_1, n_2, n_3)$, $a^\alpha = (a, 0, 0)$, so that the Jacobi identities are then simply $n_1 a = 0$. Consequently we define two major classes of structure constants (and so Lie algebras):

Class A: $a = 0$; and
Class B: $a \neq 0$.

Following Schücking's extension of Bianchi's work, the classification of G_3 group types used is as in table 3.2. Given a specific group type at one instant, this type will be preserved by the evolution equations for the quantities $n_\alpha(t)$ and $a(t)$. This is a consequence of a generic property of the EFE: they will always preserve symmetries in initial data (within the Cauchy development of that data); see Hawking and Ellis [68].

In some cases, the Bianchi groups allow higher symmetry subcases, i.e. they are compatible with isotropic (FL) or LRS models, see [45] for details. For us the interesting point is that $k = 0$ FL models are compatible with groups of type I and VII$_0$, $k = -1$ models with groups of types V and VII$_h$, and $k = +1$ models with groups of type IX.

Table 3.2. Canonical structure constants for different Bianchi types. The parameter $h = a^2/n_2n_3$.

Class	Type	n_1	n_2	n_3	a	
A	I	0	0	0	0	Abelian
	II	+ve	0	0	0	
	VI_0	0	+ve	−ve	0	
	VII_0	0	+ve	+ve	0	
	VIII	−ve	+ve	+ve	0	
	IX	+ve	+ve	+ve	0	
B	V	0	0	0	+ve	
	IV	0	0	+ve	+ve	
	VI_h	0	+ve	−ve	+ve	$h < 0$
	III	0	+ve	−ve	n_2n_3	same as VI_1
	VII_h	0	+ve	+ve	+ve	$h > 0$

The set of tetrad equations (section 3.3) with restrictions (3.86) will determine the evolution of all the commutation functions and matter variables and, hence, determine the metric and also the evolution of the Weyl tensor. One can relate these equations to variational principles and a Hamiltonian, thus expressing them in terms of a potential formalism that gives an intuitive feel for what the evolution will be like [92, 93]. They are also the basis of dynamical systems analyses.

3.7.2 Dynamical systems approach

The most illuminating description of the evolution of families of Bianchi models is a dynamical systems approach based on the use of orthonormal tetrads, presented in detail in Wainwright and Ellis [128]. The main variables used are essentially the commutation functions mentioned earlier, but rescaled by a common time-dependent factor.

3.7.2.1 Reduced differential equations

The basic idea [12, 126] is to write the EFE in a way that enables one to study the evolution of the various physical and geometrical quantities *relative to the overall rate of expansion of the universe*, as described by the rate of expansion scalar Θ or, equivalently, *the Hubble parameter* $H = \frac{1}{3}\Theta$. The remaining freedom in the choice of orthonormal tetrad needs to be eliminated by specifying the variables Ω^α implicitly or explicitly (for example by specifying the basis as eigenvectors of the $\sigma_{\alpha\beta}$). This also simplifies other quantities (for example the choice of a shear eigenframe will result in the tensor $\sigma_{\alpha\beta}$ being represented by two diagonal terms). One hence obtains a reduced set of variables, consisting of H and the remaining commutation functions, which we denote symbolically by $\boldsymbol{x} = (\gamma^a{}_{bc}|_{\text{reduced}})$. The physical state of the model is thus described by the vector $(H, \boldsymbol{x})$. The details of this reduction differ for classes A and B in the latter case, there is an algebraic constraint of the form $g(\boldsymbol{x}) = 0$, where g is a homogeneous polynomial.

The idea is now to normalize $\boldsymbol{x}$ with the Hubble parameter H. Denoting the resulting variables by a vector $\boldsymbol{y} \in R^n$, we write

$$\boldsymbol{y} = \frac{\boldsymbol{x}}{H}. \tag{3.88}$$

These new variables are *dimensionless*, and will be referred to as *expansion-normalized variables*. It is clear that each dimensionless state $\boldsymbol{y}$ determines a one-parameter family of physical states $(\boldsymbol{x}, H)$. The evolution equations for the $\gamma^a{}_{bc}$ lead to evolution equations for H and $\boldsymbol{x}$ and hence for $\boldsymbol{y}$. In order that the evolution equations define a flow, it is necessary, in conjunction with the rescaling of the variables, to introduce a *dimensionless time variable* τ according to

$$S = S_0 \mathrm{e}^{\tau}, \tag{3.89}$$

where S_0 is the value of the scale factor at some arbitrary reference time. Since S assumes values $0 < S < +\infty$ in an ever-expanding model, τ assumes all real values, with $\tau \to -\infty$ at the initial singularity and $\tau \to +\infty$ at late times. It follows that

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \frac{1}{H} \tag{3.90}$$

and the evolution equation for H can be written

$$\frac{\mathrm{d}H}{\mathrm{d}\tau} = -(1+q)H, \tag{3.91}$$

where the *deceleration parameter* q is defined by $q = -\ddot{S}S/\dot{S}^2$, and is related to $\dot{H}$ by $\dot{H} = -(1+q)H^2$. Since the right-hand side of the evolution equations for the $\gamma^a{}_{bc}$ are homogeneous of degree 2 in the $\gamma^a{}_{bc}$, the change (3.90) of the time variable results in H cancelling out of the evolution equation for $\boldsymbol{y}$, yielding an autonomous differential equation (DE):

$$\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}\tau} = \boldsymbol{f}(\boldsymbol{y}), \qquad \boldsymbol{y} \in R^n. \tag{3.92}$$

The constraint $g(\boldsymbol{x}) = 0$ translates into a constraint

$$g(\boldsymbol{y}) = 0, \tag{3.93}$$

which is preserved by the DE. The functions $\boldsymbol{f} : R^n \to R^n$ and $g : R^n \to R$ are polynomial functions in $\boldsymbol{y}$. An essential feature of this process is that the evolution equation for H, namely (3.91), decouples from the remaining equations (3.92) and (3.93). Thus the DE (3.92) describes the evolution of the non-tilted Bianchi cosmologies, the transformation of variables essentially scaling away the effects of the overall expansion. An important consequence is that the new variables are bounded near the initial singularity.

3.7.2.2 *Equations and orbits*

Since τ assumes all real values (for models which expand indefinitely), the solutions of (3.92) are defined for all τ and hence define a *flow* $\{\phi_\tau\}$ on R^n. The evolution of the cosmological models can thus be analysed by studying the orbits of this flow in the physical region of state space, which is a subset of R^n defined by the requirement that the matter energy density μ be non-negative, i.e.

$$\Omega(\boldsymbol{y}) = \frac{\kappa\mu}{3H^2} \geq 0, \tag{3.94}$$

where the density parameter Ω is a dimensionless measure of μ.

The *vacuum boundary*, defined by $\Omega(\boldsymbol{y}) = 0$, describes the evolution of vacuum Bianchi models, and is an invariant set which plays an important role in the qualitative analysis because vacuum models can be asymptotic states for perfect fluid models near the big bang or at late times. There are other invariant sets which are also specified by simple restrictions on $\boldsymbol{y}$ which play a special role: the subsets representing each Bianchi type (table 3.2), and the subsets representing higher-symmetry models, specifically the FLRW models and the LRS Bianchi models (table 3.1).

It is desirable that the dimensionless state space D in R^n is a *compact set*. In this case each orbit will have non-empty future and past limit sets, and hence there will exist a past attractor and a future attractor in state space. When using expansion-normalized variables, compactness of the state space has a direct physical meaning for ever-expanding models: if the state space is compact, then at the big bang no physical or geometrical quantity diverges more rapidly than the appropriate power of H, and at late times no such quantity tends to zero less rapidly than the appropriate power of H. This will happen for many models; however, the state space for Bianchi type VII_0 and type VIII models is non-compact. This lack of compactness manifests itself in the behaviour of the Weyl tensor at late times.

3.7.2.3 *Equilibrium points and self-similar cosmologies*

Each ordinary orbit in the dimensionless state space corresponds to a one-parameter family of physical universes, which are conformally related by a constant rescaling of the metric. However, for an equilibrium point $\boldsymbol{y}^*$ of the DE (3.92), which satisfies $\boldsymbol{f}(\boldsymbol{y}^*) = \boldsymbol{0}$, the deceleration parameter q is a constant, i.e. $q(\boldsymbol{y}^*) = q^*$, and we find

$$H(\tau) = H_0 \mathrm{e}^{(1+q^*)\tau}.$$

In this case the parameter H_0 is no longer essential, since it can be set to unity by a translation of τ, $\tau \to \tau +$ constant; then (3.90) implies that

$$Ht = \frac{1}{1+q^*}, \tag{3.95}$$

so that the commutation functions are of the form (constant) $\times\, t^{-1}$. It follows that the resulting cosmological model is *self-similar*. It thus turns out that *to each equilibrium point of the DE (3.92) there corresponds a unique self-similar cosmological model*. In such a model the physical states at different times differ only by an overall change in the length scale. Such models are expanding, but in such a way that their dimensionless state does not change. They include the flat FLRW model ($\Omega = 1$) and the Milne model ($\Omega = 0$). All vacuum and non-tilted perfect fluid self-similar Bianchi solutions have been given by Hsu and Wainwright [73]. The equilibrium points determine the asymptotic behaviour of other more general models.

3.7.2.4 Phase planes

Many phase planes can be constructed explicitly. The reader is referred to Wainright and Ellis [128] for a comprehensive presentation and survey of results. Several interesting points emerge

(1) *Variety of singularities*. Various types of singularity can occur in Bianchi universes: cigar, pancake and oscillatory in the orthogonal case. In the case of tilted models, one can, in addition get non-scalar singularities, associated with a change in the nature of the spacetime symmetries—a horizon occurs where the surfaces of homogeneity change from being timelike to being spacelike, so the model changes from being spatially homogeneous to spatially inhomogeneous [15, 42]. The fluid can then run into timelike singularities, quite unlike the spacelike singularities in FL models. Thus the singularity structure can be quite unlike that in a FL model, even in models that are arbitrarily similar to a FL model today and indeed since the time of decoupling.
(2) *Relation to lower dimensional spaces*. It seems that the lower dimensional spaces, delineating higher symmetry models, may be skeletons guiding the development of the higher dimensional spaces (the more generic models). This is one reason why study of the exact higher symmetry models is of significance.
(3) *Identification of models in state space*. The analysis of the phase planes for Bianchi models shows that the procedure sometimes adopted of identifying all points in state space corresponding to the same model, is not a good idea. For example the Kasner ring that serves as a framework for evolution of many other Bianchi models contains multiple realizations of the same Kasner model. To identify them as the same point in state space would make the

evolution patterns very difficult to follow. It is better to keep them separate, but to learn to identify where multiple realizations of the same model occur (which is just the *equivalence problem* for cosmological models).

3.7.3 Isotropization properties

An issue of importance is whether these models tend to isotropy at early or late times. An important paper by Collins and Hawking [16] shows that for ordinary matter, at late times, types I, V, VII, isotropize but other Bianchi models become anisotropic at very late times, even if they are very nearly isotropic at present. Thus isotropy is unstable in this case. However, a paper by Wald [130] showed that Bianchi models will tend to isotropize at late times if there is a positive cosmological constant present, implying that an inflationary era can cause anisotropies to die away. The latter work, however, while applicable to models with non-zero tilt angle, did not show this angle dies away, and indeed it does not do so in general (Goliath and Ellis [62]). Inflation also only occurs in Bianchi models if there is not too much anisotropy to begin with (Rothman and Ellis [111]), and it is not clear that shear and spatial curvature are in fact removed in all cases [109]. Hence, some Bianchi models isotropize due to inflation, but not all.

An important idea that arises out of this study is that of *intermediate isotropization*: namely, models that become very like a FLRW model for a period of their evolution but start and end quite unlike these models. It turns out that many Bianchi types allow intermediate isotropization, because the FLRW models are saddle points in the relevant phase planes. This leads to the following two interesting results:

Bianchi evolution theorem 1. *Consider a family of Bianchi models that allow intermediate isotropization. Define an ϵ-neighbourhood of a FLRW model as a region in state space where all geometrical and physical quantities are closer than ϵ to their values in a FLRW model. Choose a time scale L. Then no matter how small ϵ and how large L, there is an open set of Bianchi models in the state space such that each model spends longer than L within the corresponding ϵ-neighbourhood of the FLRW model.*

This follows because the saddle point is a fixed point of the phase flow; consequently the phase flow vector becomes arbitrarily close to zero at all points in a small enough open region around the FLRW point in state space. Consequently, although these models are quite unlike FLRW models at very early and very late times, there is an open set of them that are observationally indistinguishable from a FLRW model (choose L long enough to encompass from today to last coupling or nucleosynthesis, and ϵ to correspond to current observational bounds). Thus there exist many such models that are viable as models of the real universe in terms of compatibility with astronomical observations.

Bianchi evolution theorem 2. *In each set of Bianchi models of a type admitting intermediate isotropization, there will be spatially homogeneous models that are linearizations of these Bianchi models about FLRW models. These perturbation modes will occur in any almost-FLRW model that is generic rather than fine-tuned; however, the exact models approximated by these linearizations will be quite unlike FLRW models at very early and very late times.*

Proof is by linearizing the previous equations (see the following section) to obtain the Bianchi equations linearized about the FLRW models that occur at the saddle point leading to the intermediate isotropisation. These modes will be the solutions in a small neighbourhood about the saddle point permitted by the linearized equations (given existence of solutions to the nonlinear equations, linearization will not prevent corresponding linearized solutions existing).

The point is that these modes can exist as linearizations of the FLRW model; if they do not occur, then the initial data have been chosen to set these modes precisely to zero (rather than being made very small), which requires very special initial conditions. Thus these modes will occur in almost all almost-FLRW universes. Hence, if one believes in generality arguments, they will occur in the real universe. When they occur, they will, at early and late times grow until the model is very far from a FLRW geometry (while being arbitrarily close to an FLRW model for a very long time, as per the previous theorem).

3.8 Observations and horizons

The basic observational problem is that, because of the enormous scale of the universe, we can effectively only see it from one spacetime point, 'here and now' [26, 29]. Consequently what we are able to see is a projection onto a 2-sphere ('the sky') of all the objects in the universe, and our fundamental problem is determining the distances of the various objects we see in the images we obtain. In the standard universe models, redshift is a reliable zero-order distance indicator, but is unreliable at first order because of local velocity perturbations. Thus we need the array of other distance indicators (Tully–Fisher for example). Furthermore, to test cosmological models we need at least two reliable measurable properties of the objects we see, that we can plot against each other (magnitude and redshift, for example), and most of them are unreliable both because of intrinsic variation in source properties, and because of evolutionary effects associated with the inevitable lookback-time involved when we observe distant objects.

3.8.1 Observational variables and relations: FL models

The basic variables underlying direct observations of objects in the spatially homegenous and isotropic FL models are:

(1) the *redshift*, basically a time-dilation effect for all measurements of the source (determined by the radial component of velocity);
(2) the *area distance*, equivalent to the angular diameter distance in RW geometries, and also equivalent (up to a redshift factor) to the luminosity distance in all relativistic cosmological models, because of the reciprocity theorem [26]—this can best be calculated from the geodesic deviation equation [54]; and
(3) *number counts*, determined by (i) the number of objects in a given volume, (ii) the relation of that volume to increments in distance measures (determined by the spacetime geometry) and (iii) the selection and detection effects that determine which sources we can actually identify and measure (difficulties in detection being acute in the case of dark matter).

Thus to determine the spacetime geometry in these models, we need to correlate at least two of these variables against each other. Further observational data which must be consistent with the other observations comes from the following sources:

(4) *background radiation spectra* at all wavelengths particularly the 3K blackbody relic radiation ('CBR'); and
(5) *the 'relics' of processes taking place in the hot big-bang era*, for example the primeval element abundances resulting from baryosynthesis and nucleosynthesis in the hot early universe (and the CBR anisotropies and present large-scale structures can also be regarded in this light, for they are evidence about density fluctuations, which are one form of such relic).

The observational relations in FL models are covered in depth in other reports to this meeting (and see also [26, 33]), so I will just comment on two aspects here.

Selection/detection effects: The way we detect objects from available images depends on both their surface brightness and their apparent size. Thus we essentially need *two variables* to adequately characterize selection and detection effects; it simply is not adequate to discuss such effects on the basis of apparent magnitude or flux alone [47]. Hence one should regard with caution any catalogues that claim to be magnitude limited, for that cannot be an adequate criteria for detection limits; such catalogues may well be missing out many low surface-brightness objects.

Minimum angles and trapping surfaces: For ordinary matter, there is a redshift $z_\star$ such that apparent sizes of objects of fixed linear size reach a minimum at $z = z_\star$, and for larger redshift look larger again. What is happening here is that the universe as a whole is acting as a giant gravitational lens, refocusing our past light cone as a whole [26]; in an Einstein–de-Sitter universe, this happens at $z_\star = 5/4$; in a low density universe, it happens at about $z = 4$. This refocusing means that closed trapped surfaces occur in the universe, and hence via the Hawking–Penrose singularity theorems, leads to the prediction of the existence of a spacetime singularity in our past [68].

3.8.2 Particle horizons and visual horizons

For ordinary equations of state, because causal influences can travel at most at the speed of light, there is both a *particle horizon* [110, 124], limiting causal communication since the origin of the universe and a *visual horizon* [50], limiting visual communication since the decoupling of matter and radiation. The former depends on the equation of state of matter at early times, and can be changed drastically by an early period of inflation; however the latter depends only on the equation of state since decoupling, and is unaffected by whether inflation took place or not. From (3.75), at an arbitrary time of observation t_0, the radial comoving coordinate values corresponding to the particle and event horizons, respectively, of an observer at the origin of coordinates are:

$$u_{\mathrm{ph}}(t_0) = \int_0^{t_0} \frac{\mathrm{d}t}{S(t)}, \qquad u_{\mathrm{vh}}(t_0) = \int_{t_d}^{t_0} \frac{\mathrm{d}t}{S(t)}, \tag{3.96}$$

where we have assumed the initial singularity occurred at $t = 0$ and decoupling at $t = t_{\mathrm{d}}$. We cannot have had causal contact with objects lying at a coordinate value r greater than $u_{\mathrm{ph}}(t_0)$, and cannot have received any type of electromagnetic radiation from objects lying at a coordinate value r greater than $u_{\mathrm{vh}}(t_0)$.

It is fundamental to note, then, that no object can leave either of these horizons once it has entered it: *once two objects are in causal or visual contact, that contact cannot be broken, regardless of whether inflation or an accelerated expansion takes place or not.* This follows immediately from (3.96): $t_1 > t_0 \Rightarrow u_{\mathrm{ph}}(t_1) > u_{\mathrm{ph}}(t_0)$ (the integrand between t_0 and t_1 is positive, so $\mathrm{d}u_{\mathrm{ph}}(t)/\mathrm{d}t = 1/S(t) > 0$.) Furthermore the physical scales associated with these horizons cannot decrease while the universe is expanding. These are

$$D_{\mathrm{ph}}(t) = S(t)u_{\mathrm{ph}}(t), \qquad D_{\mathrm{vh}}(t) = S(t)u_{\mathrm{vh}}(t)$$

respectively, at time t; hence for example $\mathrm{d}(D_{\mathrm{ph}}(t))/\mathrm{d}t = 1 + H(t)D_{\mathrm{ph}}(t) > 0$. Much of the literature on inflation is misleading in this regard.

3.8.3 Small universes

The one case where visual horizons do not occur is when the universe has compact spatial sections whose physical size is less than the Hubble radius; consider, for example, the case of a $k = 0$ model universe of toroidal topology, with a length scale of identification of, say, 300 Mpc. In that case we can see right round the universe, with many images of each galaxy, and indeed many images of our own galaxy [48]. There are some philosophical advantages in such models [32], but they may or may not correspond to physical reality. If this is indeed the case, it would show up in multiple images of the same objects [48, 81], identical circles in the CBR anisotropy pattern across the sky [18], and altered CBR power spectra predictions [17]. A complete cosmological observational programme should test for the possibility of such small alternative universe topologies, as well as determining the fundamental cosmological parameters.

3.8.4 Observations in anisotropic and inhomogeneous models

In anisotropic models, new kinds of observations become possible. First, each of these relations will be anisotropic and so will vary with direction in the sky. In particular,

(6) *background radiation anisotropies* will occur and provide important information on the global spacetime geometry [100] as well as on local inhomogeneities [10, 59, 82] and gravitational waves [9];
(7) *image distortion effects* (strong and weak lensing) are caused by the Weyl tensor, which in turn is generated by local matter inhomogeneities through the 'div E' equation (3.48).

Finally, to fully determine the spacetime geometry [44, 86] we should also measure

(8) *transverse velocities*, corresponding to proper motions in the sky. However, these are so small as to be undetectable and so measurements only give weak upper limits in this case.

To evaluate the limits put on inhomogenity and anisotropy by observations, one must calculate observational relations predicted in anisotropic and inhomogenous models.

3.8.4.1 Bianchi observations

One can examine observational relations in the spatially homogeneous class of models, for example determining predicted Hubble expansion anisotropy, CBR anisotropy patterns, and nucleosynthesis results in Bianchi universes. These enable one to put strong limits on the anisotropy of these universe models since decoupling, and limits on the deviation from FL expansion rates during nucleosynthesis. However although these analyses put strong limits on the shear and vorticity in such models today, nevertheless they could have been very anisotropic at very early times—in particular, before nucleosynthesis—without violating the observational limits, and they could become anisotropic again at very late times. Also these limits are derived for specific spatially homogeneous models of particular Bianchi type, and there are others where they do not apply. For example, there exist Bianchi models in which rapid oscillations take place in the shear at late times, and these oscillations prevent a build up of CBR anisotropy, even though the universe is quite anisotropic at many times.

3.8.4.2 *Inhomogeneity and observations*

Similarly, one can examine observational relations in specific inhomogeneous models, for example the Tolman–Bondi spherically symmetric models and hierarchical Swiss-cheese models. We can then use these models to investigate the spatial homegenity of the universe (cf the next subsection).

The observational relations in linearly perturbed FL models, particularly (a) gravitational lensing properties and (b) CBR anisotropies have been the subject of intense theoretical study as well as observational exploration. A crucial issue that arises is on what scale we are representing the universe, for both its dynamic and observational properties may be quite different on small and large scales, and then the issue arises of how averaging over the small-scale behaviour can lead to the correct observational behaviour on large scales [32]. It seems that this will work out correctly, but really clear and compelling arguments that this is so are still lacking.

3.8.4.3 *Perturbed FL models and FL parameters*

As explained in detail in other chapters, the CBR anisotropies in perturbed FL models, in conjunction with studies of large-scale structure and models of the growth of inhomogeneities in such models, also using large-scale structure and supernovae observations, enables us to tie down the parameters of viable FL background models to a striking degree [8, 75].

3.8.5 Proof of almost-FL geometry

On a cosmological scale, observations appear almost isotropic about us (in particular number counts of many kinds of objects on the one hand, and the CBR temperature on the other). From this we may deduce that the observable region of the universe is, to a good approximation, also isotropic about us. A particular substantial issue, then, is how we can additionally prove the universe is spatially homogeneous, and so has an RW geometry, as is assumed in the standards models of cosmology.

3.8.5.1 *Direct proof*

Direct proof of spatial homogeneity would follow if we could show that the universe has precisely the relation between both area distance $r_0(z)$ and number counts $N(z)$ with redshift z that is predicted by the FL family of models. However, proving this observationally is not easily possible. Current supernova-based observations are indicate a non-zero cosmological constant rather than the relation predicted by the FL models with zero λ, and we are not able to test the $r_0(z)$ relationship accurately enough to show it takes a FL form with non-zero λ [95]. Furthermore number counts are only compatible with the FL models if we assume just the right source evolution takes place to make the observations compatible with spatial homogeneity; but once we take evolution into account, we can fit almost any observational relations by almost any spherically symmetric model (see [98] for exact theorems making this statement precise). Recent statistical observations of distant sources support spatial homogeneity on intermediate scales (between 30 and 400 Mpc [102]), but do not extend to larger scales because of sample limits.

3.8.5.2 Uniform thermal histories

A strong indication of spatial homogeneity is the fact that we see the same kinds of object, more or less, at high z as nearby. This suggests that they must have experienced more or less the same thermal history as nearby objects, as otherwise their structure would have come out different; and this, in turn, suggests that the spacetime geometry must have been rather similar near those objects as near to us, else (through the field equations) the thermal history would have come out different. This idea can be formulated in the *Postulate of Uniform Thermal Histories* (PUTH), stating that uniform thermal histories can occur only if the geometry is spatially homogeneous. Unfortunately, counter-examples to this conjecture have been found [7]. These are, however, probably exceptional cases and this remains a strong observationally-based argument for spatial homogeneity, indeed probably the most compelling at an intuitive level. However, relating the idea to observations also involves untangling the effects of time evolution, and it cannot be considered a formal proof of homogeneity.

3.8.5.3 Almost-EGS theorem

The most compelling precisely formulated argument is a based on our observations of the high degree of CBR anisotropy around us. If we assume we are not special observers, others will see the same high degree of anisotropy; and then that shows spatial homogeneity: exactly, in the case of exact isotropy (the Ehlers–Geren–Sachs (EGS) theorem [22]) and approximately in the case of almost-isotropy:

Almost-EGS-theorem. *[119]. If the Einstein–Liouville equations are satisfied in an expanding universe, where there is pressure-free matter with 4-velocity vector field u^a ($u_a u^a = -1$) such that (freely-propagating) background radiation is everywhere almost-isotropic relative to u^a in some domain U, then spacetime is almost-FLRW in U.*

This description is intended to represent the situation since decoupling to the present day. The pressure-free matter represents the galaxies on which fundamental observers live, who measure the radiation to be almost isotropic. This deduction is very plausible, particularly because of the argument just mentioned in the last subsection: conditions there *look* more or less the same, so there is no reason to think they are very different. Nevertheless, in the end this argument rests on an unproved philosophical assumption (that other observers see more or less what we do), and so is highly suggestive rather than a full observational proof. In addition, there is a technical issue of substance, namely what derivatives of the CBR temperature should be included in this formulation (remembering here that there are Bianchi models where the matter shear remains small but its time derivative can be large; these can have a large Weyl tensor but small CBR anisotropy [99]).

3.8.5.4 Theoretical arguments

Given the observational difficulties, one can propose theoretical rather than observational arguments for spatial homogeneity. Traditionally this was done by appeal to a *cosmological principle* [6,131]; however, this is no longer fashionable. Still some kinds of theoretical argument remain in vogue.

One can try to argue for spatial homogeneity on the basis of *probability*: this is more likely than the case of a spherically symmetric inhomogeneous universe, where we are near the centre (see [43] for detailed development of such a model). However, that argument is flawed [30], because spatially homogeneous universe models are intrinsically less likely than spherically symmetric inhomogeneous ones (as the latter have more degrees of freedom, and so are more general). In additionally, it is unclear that any probability arguments at all can be applied to the universe, because of its uniqueness [37].

Alternatively, one can argue that *inflation guarantees that the universe must be spatially homogeneous*. If we accept that argument, then the implication is that we are giving preference to a theoretically based analysis over what can, in fact, be established from observational data. In addition, it provides a partial rather than complete solution to the issues it addresses (see the discussion in the next section). Nevertheless it is an important argument that many find fully convincing.

Perhaps the most important argument in the end is that from *cumulative evidence*: none of these approaches by themselves proves spatial homogeneity, but taken together they give a sound cumulative argument that this is indeed the case—within the domain previously specified above.

3.8.5.5 Domains of plausibility

Accepting that argument, to what spacetime regions does it apply? We may take it as applying to the observable region of the universe $\mathcal{V}$, that is, *the region both inside our visual horizon, and lying between the epoch of decoupling and the present day*. It will then also hold in some larger neigbourhood of this region, but there is no reason to believe it will hold elsewhere; specifically, it need not hold (i) very far out from us (say, 1000 Hubble radii away), hence chaotic inflation is a possibility; nor (ii) at very early times (say, before nucleosynthesis), so Bianchi anisotropic modes are possible at these early times; nor (iii) at very late times (say in another 50 Hubble times), so late-time anisotropic modes which are presently negligble could come to dominate (cf the discussion in the section on evolution of Bianchi models above). Thus we can observationally support the supposition of spatial homegeneity and isotropy within the domain $\mathcal{V}$, but not too far outside of it.

3.8.6 Importance of consistency checks

Because we have no knock-out observational proof of spatial homogeneity, it is important to consider all the possible observationally based consistency checks on the standard model geometry. The most important are as follows:

(1) *Ages*. This has been one of the oldest worries for expanding universe models: the requirement that the age of the universe must be greater than the ages of all objects in it. However with present estimates of the ages of stars on the one hand, and of the value of the Hubble constant on the other, this no longer seems problematic, particularly if current evidence for a positive cosmological constant turn out to be correct.

(2) *Anisotropic number counts*. If our interpretation of the CBR dipole as due to our motion relative to the FL model is correct, then this must also be accompanied by a dipole in all cosmological number counts at the 2% level [38]. Observationally verifying that this is so is a difficult task, but it is a crucial check on the validity of the standard model of cosmology.

(3) *High-z observations*. The best check on spatial homogeneity is to try to check the physical state of the universe at high redshifts and hence at great distances from us, and to compare the observations with theory. This can be done in particular (a) for the CBR, whose temperature can be measured via excited states of particular molecules; this can then be compared with the predicted temperature $T = T_0(1+z)$, where T_0 is the present day temperature of 2.75 K. It can also be done (b) for element abundances in distant objects, specifically helium abundances. This is particularly useful as it tests the thermal history of the universe at very early times of regions that are far out from us [34].

3.9 Explaining homogeneity and structure

This is the unique core business of physical cosmology: explaining both why the universe has the very improbable high-symmetry FL geometry on the largest scales, and how structures come into existence on all smaller scales. Clearly only cosmology itself can ask the first question; and it uniquely sets the initial conditions underlying the astrophysical and physical processes that are the key to the second, underlying all studies of origins.There is a creative tension between two aims: smoothing processes, on the one hand, and structure growth, on the other. Present day cosmology handles this tension by suggesting a change of equation of state: at early enough times, the equation of state was such as to cause smoothing on all scales; but at later times, it was such as to cause structure growth on particular scales. The inflationary scenario, and the models that build on it, are remarkably successful in this regard, particularly through predicting the CBR anisotropy patterns (the 'Doppler peaks') which seem to have been found now (but significant problems remain, particularly as regards compatibility with the well-established nucleosynthesis arguments).

Given these astrophysical and physical processes, explanation of the large-scale isotropy and homogeneity of the universe together with the creation of smaller-scale structures means determining the dynamical evolutionary trajectories relating initial to final conditions, and then essentially either (a) explaining initial conditions or (b) showing they are irrelevant.

3.9.1 Showing initial conditions are irrelevant

This can be attempted in a number of different ways.

3.9.1.1 Initial conditions are irrelevant because they are forgotten

Demonstrating minimal dependence of the large-scale final state on the initial conditions has been the aim of

- the *chaotic cosmology* programme of Misner, where physical processes such a viscosity wipe out memories of previous conditions [97]; and
- the *inflationary family of theories*, where the rapid exponential expansion driven by a scalar field smooths out the universe and so results in similar memory loss [79].

The (effective) scalar field is slow-rolling, so the energy condition (3.36) is violated and a period of accelerating expansion can take place through many e-foldings, until the scalar field decays into radiation at the end of inflation. This drives the universe model towards flatness, and is commonly believed to predict that the universe must be very close indeed to flatness today, even though this is an unstable situation, see the phase planes of Ω against S [94]. It can also damp out both anisotropy, as previously explained and inhomogeneity, if the initial situation is close enough to a FL model of that inflation can in fact start. In a chaotic inflationary scenario, with random initial conditions occurring at some initial time, inflation will not succeed in starting in most places, but those domains where it does start will expand so much that they will soon be the dominant feature of the universe: there will be many vast FL-like domains, each with different parameter values and perhaps even different physics, separated from each other by highly inhomogeneous transition regions (where physics may be very strange). In the almost-FL domains, quantum fluctuations are expanded to a very large scale in the inflationary era, and form the seeds for structure formation at later times. Inflation then goes on to provide a causal theory of initial structure formation from an essentially homogeneous early state (via amplification of initial quantum fluctuations)—a major success if the all the details can be sorted out.

This is an attractive scenario, particularly because it ties in the large-scale structure of the universe with high-energy physics. It works fine for those regions that start off close enough to FL models, and, as noted earlier this suffices to explain the existence of large FL-like domains, such as the one we inhabit. It does not necessarily rule out the early and late anisotropic modes that were discussed in the section on Bianchi models. It fits the observations provided one has enough auxiliary functions and parameters available to mediate between the basic theory and the observations (specifically, evolution functions, a bias parameter or function, a dark matter component, a cosmological constant or 'quintessence' component at late times). However, it is not at present a specific physical model, rather it is a family of models (see e.g. [78]), with many competing explanations for the origin of the inflaton, which is not yet identified with any specific matter component or field. It will become a well-defined physical theory when one or other of these competing possibilities is identified as the real physical driver of an inflationary early epoch.

There are three other issues to note here. First, the *issue of probability*: inflation is intended as a means of showing the observed region of the universe is in fact probable. But we have no proper measure of probability on the family of universe models, so this has not been demonstrated in a convincing way. Second, the *Trans-Planckian problem* [96]: inflation is generally very successful in generating a vast expansion of the universe. The consequence is that the spacetime region that has been expanded to macroscopic scales today is deep in the Planck (quantum-gravity) era, so the nature of what is predicted depends crucially on our assumptions about that era; but we do not know what conditions were like there, and indeed even lack proper tools to describe that epoch, which may have been of the nature of a spacetime foam, for example. Thus the results of inflation for large-scale structure depend on specific assumptions about the nature of spacetime in the strong quantum gravity regime, and we do not know what that nature is. Penrose suggests it was very inhomogeneous at that time, in which case inflation will amplify that inhomogeneous nature rather than creating spatial homgeneity. As in the previous case, whether or not the process succeeds will depend on the initial conditions for the expansion of the universe as it emerges from the Planck (quantum gravity) era. Thirdly, there are still unsolved problems regarding *the end of inflation*. These relate to the fact that if one has a very slow rolling field as is often claimed, then the inertial mass density is very close to zero so velocities are unstable.

It must be emphasized that in order to investigate this issue of isotropisation properly, *one must examine the dynamical behaviour of very anisotropic and inhomogeneous cosmologies*. This is seldom done—for example, almost all of the literature on inflation examines only its effects in RW geometries, which is precisely when there is no need for inflation take place in order to explain the smooth geometry—for then a smooth geometry has been assumed *a priori*.

When the full range of inhomogeneities and anisotropies is taken into account (e.g. [128]), it appears that *both approaches are partially successful*: with or without inflation one can explain a considerable degree of isotropization and homogenization of the physical universe (see e.g. [127]), but this will not work in all circumstances [105,106]. It can only be guaranteed to work if initial conditions are somewhat restricted—so in order for the programme to succeed, we have to go back to the former issue of somehow explaining why it is probable for a restricted set of initial data to occur.

3.9.1.2 Initial conditions are irrelevant because they never happened

Some attempts involve avoiding a true beginning by going back to some form of eternal or cyclic state, so that the universe existed forever. Initial conditions are pushed back into the infinite past, and thus were never set. Examples are as follows.

- The original *steady state universe* proposal of Bondi [6], and its updated form as the *quasi-steady state universe* of Hoyle, Burbidge and Narlikar [71,72].
- Linde's *eternal chaotic inflation*, where ever-forming new bubbles of expansion arising within old ones exist forever; this can prevent the universe from ever entering the quantum gravity regime [90].
- The Hartle–Hawking *'no-boundary' proposal* (cf [67]) avoids the initial singularity through a change of spacetime signature at very early times, thereby entering a positive-definite ('space–space') regime where the singularity theorems do not apply (the physical singularity of the big bang gets replaced by the coordinate singularity at the south pole of a sphere). There is no singularity and no boundary, and so there are no boundary conditions. This gets round the issue of a creation event in an ingenious way: there is no unique start to the universe, but there is a beginning of time.
- The Hawking–Turok *initial instanton proposal* is a variant of this scenario, where there is a weak singularity to start with, and one is then able to enter a low-density inflationary phase.
- Gott and Liu's *causality violation in the early universe* does the same kind of thing in a different way: causality violation takes place in the early universe, enabling the universe to 'create itself' [63]. Like the chaotic inflation picture, new expanding universe bubbles are forming all the time; but one of them is the universe region where the bubble was formed, this being possible because closed timelike lines are allowed, so 'the universe is its own mother'. This region of closed timelike lines is separated from the later causally regular regions by a Cauchy horizon.

There are thus a variety of ingenious and intriguing options which, in a sense, allow avoidance of setting initial conditions. But this is really a technicality: the issue still arises as to why in each case one particular initial state 'existed' or came into being rather than any of the other options. Some particular solutions of the equations have been implemented rather than the other possibilities; boundary conditions choosing one set of solutions over others have still been set, even if they are not technically initial conditions set at a finite time in the past.

3.9.1.3 Initial conditions are irrelevant because they all happened

The idea of an ensemble of universes, mentioned earlier, is one approach that sidesteps the problem of choice of specific initial data, because by hypothesis *all that can occur has then occurred.* Anthropic arguments select the particular universe in which we live from all those in this vast family (see e.g. [57,70]). This is again an intriguing and ingenious idea, extending to a vast scale the Feynman approach to quantum theory. However, there are several problems.

First, it is not clear that the selection of universes from this vast family by anthropic arguments will necessarily result in as large and as isotropic a universe as we see today; here one runs up against the unsolved problem of justifying a choice of probabilities in this family of universes. Second, this proposal suffers from complete lack of verifiability. In my view, this means this is a metaphysical rather than scientific proposal, because it is completely untestable. And in the end, this suggestion does not solve the basic issue in any case, because then one can ask: *Why does this particular ensemble exist, rather than a different ensemble with different properties?*; and the whole series of fundamental questions arises all over again, in an even more unverifiable form than before.

3.9.2 The explanation of initial conditions

The explanation of initial conditions has been the aim of the family of theories one can label collectively as *quantum cosmology* and the more recent studies of *string cosmology*.

3.9.2.1 Explanation of initial conditions from a previous state of a different nature

One option has been explaining the universe as we see it as arising from some completely different initial state, for example:

- proposals for *creation of the universe as a bubble formed in a flat spacetime or de Sitter spacetime*, for example Tryon's vacuum fluctuations and Gott's open bubble universes; or
- Vilenkin's *tunnelling universe* which arises from a state with no classical analogue (described as 'creation of the universe from nothing', but this is inaccurate).

These proposals (like the proposals by Hartle and Hawking, Hawking and Turok, and Gott and Liu previously mentioned; for a comparative discussion and references, see Gott and Liu [63]) are based on the quantum cosmology idea of *the wavefunction of the universe*, taken to obey the Wheeler–de Witt equation (a generalization to the cosmological context of the Schrödinger equation) (see e.g. [67]). This approach faces considerable technical problems, related to

- the meaning of time, because vanishing of the Hamiltonian of general relativity means that the wavefunction appears to be explicitly independent of time;
- divergences in the path-integrals often used to formulate the solutions to the Wheeler–de-Witt equation;
- the meaning of the wavefunction of the universe, in a context where probabilities are ill defined [56];
- the fundamentally important issue of the meaning of measurement in quantum theory (when does 'collapse of the wavefunction' take place, in a context where a classical 'observer' does not exist);
- the conditions which will lead to these quantum equations having classical-like behaviour at some stage in the history of the universe [65]; and
- the way in which this reduced set of equations, taken to be valid irrespective of the nature of the full quantum theory of gravity, relates to that as yet unknown theory.

The alternative is to work with the best current proposal for such a theory, taken by many to be *M-theory*, which aims to unite the previously disparate superstring theories into a single theory, with the previously separate theories related to each other by a series of symmetries called dualities. There is a rapidly growing literature on *superstring cosmology,* relating this theory to cosmology [89]. In particular, much work is taking place on two approaches:

- The *pre big-bang proposal*, where a 'pre big-bang' branch of the universe is related to a 'post big-bang' era by a duality: $a(t) \to 1/a(t)$, $t \to -t$, and dimensional reduction results in a scalar field (a 'dilaton') occurring in the field equations (see Gasperini [58] for updated references).

This approach has major technical difficulties to solve, particularly related to the transition from the 'pre big-bang' phase to the 'post big-bang' phase, and to the transition from that phase to a standard cosmological expansion. In additionally it faces fine-tuning problems related to its initial conditions. So this too is very much a theory in the course of development, rather than a fully viable proposal.

- The *brane cosmology proposal*, where the physical universe is confined to a four-dimensional 'brane' in a five-dimensional universe. The physics of this proposal are very speculative, and issues arise as to why the initial conditions in the 5D space had the precise nature so as to confine matter to this lower-dimensional subspace; and then the confinement problem is why they remain there.

Supposing these technical difficulties can be overcome in each case, it is still unclear that these proposals avoid the real problem of origins. It can be claimed they simply postpone facing it, for one now has to ask all the same questions of origins and uniqueness about the supposed prior state to the present hot big bang expansion phase: Why did this previous state have the properties it had? (whether or not it had a classical analogue)? This 'pre-state' should be added to one's cosmology, and then the same basic questions as before now arise regarding this completed model.

3.9.2.2 Explanation of initial conditions from 'nothing'

Attempts at an 'explanation' of a true origin, i.e. not arising from some pre-existing state (whether it has a classical analogue or not), are difficult even to formulate.

They may depend on *assuming a pre-existing set of physical laws* that are similar to those that exist once spacetime exists, for they rely on an array of properties of quantum field theory and of fields (existence of Hilbert spaces and operators, validity of variational principles and symmetry principles, and so on) that seem to hold sway independently of the existence of the universe and of space and time (for the universe itself, and so space and time, is to arise out of their validity). This issue arises, for example, in the case of Vilenkin's tunnelling universes: not only do they come from a pre-existent state, as remarked previously, but they also take the whole apparatus of quantum theory for granted. This is far from 'nothing'—it is a very complex structure; but there is no clear locus for those laws to exist in or material for them to act on. The manner of their existence or other grounds for their validity in this context are unclear—and we run into the problems noted before: there are problems with the concepts of 'occurred', 'circumstances' and even 'when'—for we are talking *inter alia* about the existence of spacetime. Our language can hardly deal with this. Given the feature that no spacetime exists before such a beginning, brave attempts to define a 'physics of creation' stretch the meaning of 'physics'. There cannot be a prior physical explanation, precisely because physics and the causality associated with physics does not exist there/then.

Perhaps the most radical proposal is that

> *order arises out of nothing: all order, including the laws of physics, somehow arises out of chaos*,

in the true sense of that word—namely a total lack of order and structure of any kind (e.g. [1]). However, this does not seem fully coherent as a proposal. If the pre-ordered state is truly chaotic and without form, I do not see how order can arise therefrom when physical action is as yet unable to take place, or even how we can meaningfully contemplate that situation. We cannot assume any statistical properties would hold in that regime, for example; even formulating a description of states seems well nigh impossible, for that can only be done in terms of concepts that have a meaning only in a situation of some stability and underlying order such as is characterized by physical laws.

3.9.3 The irremovable problem

Thus a great variety of possibilities is being investigated. However, the same problem arises in every approach: even if a literal creation does not take place, as is the case in various of the present proposals, this does not resolve the underlying issue. Apart from all the technical difficulties, and the lack of experimental support for these proposals, none of these can get around the basic problem: given any specific proposal,

> *How was it decided that this particular kind of universe would be the one that was actually instantiated and what fixed its parameters?*

A choice between different contingent possibilities has somehow occurred; the fundamental issue is what underlies this choice. Why does the universe have one specific form rather than another, when other forms seem perfectly possible? Why should any one of these approaches have occurred if all the others are possibilities? This issue arises even if we assume an ensemble of universes exists: for then we can ask why this particular ensemble, and not another one?

All approaches face major problems of verifiability, for the underlying dynamics relevant to these times can never be tested. Here we inevitably reach the limits to what the scientific study of the cosmos can ever say—if we assume that such studies must of necessity involve an ability to observationally or experimentally check the relevant physical theories. However we can attain some checks on these theories by examining their predictions for the present state of the universe—its large-scale structure, smaller scale structure and observable features such as gravitational waves emitted at very early times. These are important restrictions, and are very much under investigation at the present time; we need to push our observations as far as we can, and this is indeed happening at present (particularly through deep galactic observations; much improved CBR observations; and the prospect of new generation gravitational wave detectors coming on line).

If it could be shown that only one of all these options was compatible with observations of the present day universe, this would be a major step forward: it would select one dynamical evolution from all the possibilities. However, this does not seem likely, particularly because of the proliferation of auxiliary functions that can be used to fit the data to the models, as noted before. In addition, even if this was achieved, it would not show why that one had occurred rather than any of the others. This would be achieved if it could be eventually shown that only one of these possibilities is self-consistent: that, in fact, fatal flaws in all the others reduce the field of possibilities to one. We are nowhere near this situation at present, indeed possibilities are proliferating rather than reducing.

Given these problems, any progress is of necessity based on specific philosophical positions, which decide which of the many possible physical and metaphysical approaches is to be preferred. These philosophical positions should be identified as such and made explicit [37, 88]. As explained earlier, no experimental test can determine the nature of any mechanisms that may be in operation in circumstances where even the concepts of cause and effect are suspect. Initial conditions cannot be determined by the laws of physics alone—for if they were so determined they would no longer be contingent conditions, the essential feature of initial data, but rather would be necessary. A purely scientific approach cannot succeed in explaining this specific nature of the universe.

Consequent on this situation, it follows that unavoidably, whatever approach one may take to issues of cosmological origins, metaphysical issues inevitably arise in cosmology: philosophical choices are needed in order to shape the theory. That feature should be explicitly recognized, and then sensibly developed in the optimal way by carefully examining the best way to make such choices.

3.10 Conclusion

There is a tension between theory and observation in cosmology. The issue we have considered here is, Which family of models is consistent with observations? To answer this demands an equal sophistication of geometry and physics, whereas in the usual approaches there is a major imbalance: very sophisticated physics and very simple geometry. We have looked here at tools to deal with the geometry in a resaonably sophisticated way, and summarized some of the results that are obtained by using them. This remains an interesting area of study, particularly in terms of relating realistic inhomogeneous models to the smoothed out standard FL models of cosmology.

Further problems arise in considering the physics of the extremely early universe, and any pre-physics determining initial conditions for the universe. We will need to develop approaches to these topics that explicitly recognizes the limitations of the scientific method—assuming that this method implies the possibility of verification of our theories.

References

[1] Anandan J 1998 *Preprint* quant-phy/9808045

[2] Bardeen J 1980 *Phys. Rev.* D **22** 1882

[3] Barrow J and Tipler F J 1986 *The Anthropic Cosmological Principle* (Oxford: Oxford University Press)

[4] Boerner G and Gottlober S (ed) 1997 *The Evolution of the Universe* (New York: Wiley)

[5] Bondi H 1947 *Mon. Not. R. Astron. Soc.* **107** 410

[6] Bondi H 1960 *Cosmology* 1960 (Cambridge: Cambridge University Press)

[7] Bonnor W B and Ellis G F R 1986 *Mon. Not. R. Astron. Soc.* **218** 605

[8] Bridle S L, Zehavi I, Dekel A, Lahav O, Hobson M P and Lasenby A N 2000 *Mon. Not. R. Astron. Soc.* **321** 333

[9] Challinor A 2000 *Class. Quantum Grav.* **17** 871 (astro-ph/9906474)

[10] Challinor A and Lasenby A 1998 *Phys. Rev.* D **58** 023001

[11] Cohn P M 1961 *Lie Algebras* (Cambridge: Cambridge University Press).

[12] Collins C B 1971 *Commun. Math. Phys.* **23** 137

[13] Collins C B 1977 *J. Math. Phys.* **18** 2116

[14] Collins C B 1985 *J. Math. Phys.* **26** 2009

[15] Collins C B and Ellis G F R 1979 *Phys. Rep.* **56** 63

[16] Collins C B and Hawking S W 1973 *Astrophys. J.* **180** 317

[17] Cornish N J and Spergel D N 1999 *Phys. Rev.* D **92** 087304

[18] Cornish N J, Spergel D N and Starkman G 1966 *Phys. Rev. Lett.* **77** 215

[19] d'Inverno R 1992 *Introducing Einstein's Relativity* (Oxford: Oxford Univerity Press)

[20] Dunsby P K S, Bassett B A C and Ellis G F R 1996 *Class. Quantum Grav.* **14** 1215

[21] Ehlers J 1961 *Akad. Wiss. Lit. Mainz, Abhandl. Math.-Nat. Kl.* **11** 793 (Engl. transl. 1993 *Gen. Rel. Grav.* **25** 1225)

[22] Ehlers J, Geren P and Sachs R K 1968 *J. Math. Phys.* **9** 1344

[23] Einstein A and Straus E G 1945 *Rev. Mod. Phys.* **17** 120

[24] Eisenhart L P 1933 *Continuous Groups of Transformations* (Princeton, NJ: Princeton University Press) reprinted: 1961 (New York: Dover)

[25] Ellis G F R 1967 *J. Math. Phys.* **8** 1171

[26] Ellis G F R 1971 General relativity and cosmology *Proc. XLVII Enrico Fermi Summer School* ed R K Sachs (New York: Academic Press)

[27] Ellis G F R 1971 *Gen. Rel. Grav.* **2** 7

[28] Ellis G F R 1973 *Cargèse Lectures in Physics* vol 6, ed E Schatzman (New York: Gordon and Breach)

[29] Ellis G F R 1975 *Q. J. R. Astron. Soc.* **16** 245

[30] Ellis G F R 1979 *Gen. Rel. Grav.* **11** 281

[31] Ellis G F R 1980 *Ann. New York Acad. Sci.* **336** 130

[32] Ellis G F R 1984 *General Relativity and Gravitation* ed B Bertotti *et al* (Dordrecht: Reidel) p 215

[33] Ellis G F R 1987 *Vth Brazilian School on Cosmology and Gravitation* ed M Novello (Singapore: World Scientific)

[34] Ellis G F R 1987 *Theory and Observational Limits in Cosmology* ed W Stoeger (Vatican Observatory) pp 43–72

Ellis G F R 1995 *Galaxies and the Young Universe* ed H von Hippelein, K Meisenheimer and J H Roser (Berlin: Springer) p 51

[35] Ellis G F R 1990 *Modern Cosmology in Retrospect* ed B Bertotti *et al* (Cambridge: Cambridge University Press) p 97

[36] Ellis G F R 1991 *Mem. Ital. Ast. Soc.* **62** 553–605

[37] Ellis G F R 1999 *Astron. Geophys.* **40** 4.20

Ellis G F R 2000 *Toward a New Millenium in Galaxy Morphology* ed D Block *et al* (Dordrecht: Kluwer)

Ellis G F R 1999 *Astrophysics and Space Science* **269–279** 693

[38] Ellis G F R and Baldwin J 1984 *Mon. Not. R. Astron. Soc.* **206** 377–81

[39] Ellis G F R and Bruni M 1989 *Phys. Rev.* D **40** 1804

[40] Ellis G F R, Bruni M and Hwang J C 1990 *Phys. Rev.* D **42** 1035

[41] Ellis G F R and van Elst H 1999 *Theoretical and Observational Cosmology (Nato Science Series C, 541)* ed M Lachieze-Rey (Dordrecht: Kluwer) p 1
[42] Ellis G F R and King A R 1974 *Commun. Math. Phys.* **38** 119
[43] Ellis G F R, Maartens R and Nel S D 1978 *Mon. Not. R. Astron. Soc.* **184** 439–65
[44] Ellis G F R, Nel S D, Stoeger W, Maartens R and Whitman A P 1985 *Phys. Rep.* **124** 315
[45] Ellis G F R and MacCallum M A H 1969 *Commun. Math. Phys.* **12** 108
[46] Ellis G F R and Madsen M S 1991 *Class. Quantum Grav.* **8** 667
[47] Ellis G F R, Perry J J and Sievers A 1984 *Astron. J.* **89** 1124
[48] Ellis G F R and Schreiber G 1986 *Phys. Lett.* A **115** 97–107
[49] Ellis G F R and Sciama D W 1972 *General Relativity (Synge Festschrift)* ed L O'Raifeartaigh (Oxford: Oxford University Press)
[50] Ellis G F R and Stoeger W R 1988 *Class. Quantum Grav.* **5** 207
[51] van Elst H and Ellis G F R 1996 *Class. Quantum Grav.* **13** 1099
[52] van Elst H and Ellis G F R 1998 *Class. Quantum Grav.* **15** 3545
[53] van Elst H and Ellis G F R 1999 *Phys. Rev.* D **59** 024013
[54] Ellis G F R and van Elst H 1999 *On Einstein's Path: Essays in Honour of Englebert Schucking* ed A Harvey (Berlin: Springer) pp 203–26
[55] van Elst H and Uggla C 1997 *Class. Quantum Grav.* **14** 2673
[56] Fink H and Lesche H 2000 *Found. Phys. Lett.* **13** 345
[57] Garriga J and Vilenkin A 2000 *Phys. Rev.* D **61** 083502
[58] Gasperini M 1999 *String Cosmology* http://www.to.infin.it/∼gasperini
[59] Gebbie T and Ellis G F R 2000 *Ann. Phys.* **282** 285
Gebbie T, Dunsby P K S and Ellis G F R 2000 *Ann. Phys.* **282** 321
[60] Gödel K 1949 *Rev. Mod. Phys.* **21** 447
[61] Gödel K 1952 *Proc. Int. Cong. Math. (Am. Math. Soc.)* **175**
[62] Goliath M and Ellis G F R 1999 *Phys. Rev.* D **60** 023502 (gr-qc/9811068)
[63] Gott J R and Liu L 1999 *Phys. Rev.* D **58** 023501 (astro-ph/9712344)
[64] Harrison E R 1981 *Cosmology: The Science of the Universe* (Cambridge: Cambridge University Press)
[65] Hartle J B 1996 Quantum mechanics at the Planck scale *Physics at the Planck Scale* ed J Maharana, A Khare and M Kumar (Singapore: World Scientific)
[66] Hawking S W 1966 *Astrophys. J.* **145** 544
[67] Hawking S W 1993 *Hawking on the Big Bang and Black Holes* (Singapore: World Scientific)
[68] Hawking S W and Ellis G F R 1973 *The Large Scale Structure of Spacetime* (Cambridge: Cambridge University Press)
[69] Hawking S W and Penrose R 1970 *Proc. R. Soc.* A **314** 529
[70] Hogan C J 1999 *Preprint* astro-ph/9909295
[71] Hoyle F, Burbidge G and Narlikar J 1993 *Astrophys. J.* **410** 437
[72] Hoyle F, Burbidge G and Narlikar J 1995 *Proc. R. Soc.* A **448** 191
[73] Hsu L and Wainwright J 1986 *Class. Quantum Grav.* **3** 1105
[74] Isham C J 1997 *Lectures on Quantum Theory: Mathematical and Structural Foundations* (London: Imperial College Press, Singapore: World Scientific)
[75] Jaffe A H *et al* 2001 *Phys. Rev. Lett.* **86** 3475
[76] Kantowski R and Sachs R K 1966 *J. Math. Phys.* **7** 443
[77] King A R and Ellis G F R 1973 *Commun. Math. Phys.* **31** 209
[78] Kinney W H, Melchiorri A and Riotto A 2001 *Phys. Rev.* D **63** 023505

[79] Kolb E W and Turner M S 1990 *The Early Universe* (New York: Wiley)
[80] Kompaneets A S and Chernov A S 1965 *Sov. Phys.–JETP* **20** 1303
[81] Lachieze R M and Luminet J P 1995 *Phys. Rep.* **254** 136
[82] Lewis A, Challinor A and Lasenby A 2000 *Astrophys. J.* **538** 3273 (astro-ph/9911177)
[83] Kramer D, Stephani H, MacCallum M A H and Herlt E 1980 *Exact Solutions of Einstein's Field Equations* (Cambridge: Cambridge University Press)
[84] Krasiński A 1983 *Gen. Rel. Grav.* **15** 673
[85] Krasiński A 1996 *Physics in an Inhomogeneous Universe* (Cambridge: Cambridge University Press)
[86] Kristian J and Sachs R K 1966 *Astrophys. J.* **143** 379
[87] Lemaître G 1933 *Ann. Soc. Sci. Bruxelles I* A **53** 51 (Engl. transl. 1997 *Gen. Rel. Grav.* **29** 641)
[88] Leslie J (ed) 1998 *Modern Cosmology and Philosophy* (Amherst, NY: Prometheus Books)
[89] Lidsey J E, Wands D and Copeland E J 2000 Superstring cosmology *Phys. Rep.* **337** 343–492
[90] Linde A D 1990 *Particle Physics and Inflationary Cosmology* (Chur, Switzerland: Harwood Academic)
[91] Maartens R 1997 *Phys. Rev.* D **55** 463
[92] MacCallum M A H 1973 *Cargèse Lectures in Physics* vol 6, ed E Schatzman (New York: Gordon and Breach)
[93] MacCallum M A H 1979 *General Relativity, An Einstein Centenary Survey* ed S W Hawking and W Israel (Cambridge: Cambridge University Press)
[94] Madsen M and Ellis G F R 1988 *Mon. Not. R. Astron. Soc.* **234** 67
[95] Maor I, Brustein R and Steinhardt P J 2001 *Phys. Rev. Lett.* **86** 6
[96] Martin J and Brandenberger R H 2001 *Phys. Rev.* D **63** 123501
[97] Misner C W 1968 *Astrophys. J.* **151** 431
[98] Mustapha N, Hellaby C and Ellis G F R 1999 *Mon. Not. R. Astron. Soc.* **292** 817
[99] Nilsson U S, Uggla C and Wainwright J 1999 *Astrophys. J. Lett.* **522** L1 (gr-qc/9904252)
[100] Nilsson U S, Uggla C and Wainwright J 2000 *Gen. Rel. Grav.* **32** 1319 (gr-qc/9908062)
[101] Oszvath I and Schücking E 1962 *Nature* **193** 1168
[102] Pan J and Coles P 2000 *Mon. Not. R. Astron. Soc.* **318** L51
[103] Peacocke J R 1999 *Cosmological Physics* (Cambridge: Cambridge University Press)
[104] Peebles P J E, Schramm D N, Turner E L and Kron R G 1991 *Nature* **352** 769
[105] Penrose R 1989 *Proc. 14th Texas Symposium on Relativistic Astrophysics (Ann. New York Acad. Sci.)* ed E Fenves
[106] Penrose R 1989 *The Emperor's New Mind* (Oxford: Oxford University Press) ch 7
[107] Pirani F A E 1956 *Acta Phys. Polon.* **15** 389
[108] Pirani F A E 1957 *Phys. Rev.* **105** 1089
[109] Raychaudhuri A and Modak B 1988 *Class. Quantum Grav.* **5** 225
[110] Rindler W 1956 *Mon. Not. R. Astron. Soc.* **116** 662
[111] Rothman A and Ellis G F R 1986 *Phys. Lett.* B **180** 19
[112] Schucking E 1954 *Z. Phys.* **137** 595
[113] Senovilla J M, Sopuerta C and Szekeres P 1998 *Gen. Rel. Grav.* **30** 389

[114] Smolin L 1992 *Class. Quantum Grav.* **9** 173
[115] Stabell R and Refsdal S 1966 *Mon. Not. R. Astron. Soc.* **132** 379
[116] Stephani H 1987 *Class. Quantum Grav.* **4** 125
[117] Stephani H 1990 *General Relativity* (Cambridge: Cambridge University Press)
[118] Stewart J M and Ellis G F R 1968 *J. Math. Phys.* **9** 1072
[119] Stoeger W, Maartens R and Ellis G F R 1995 *Astrophys. J.* **443** 1
[120] Szekeres P 1965 *J. Math. Phys.* **6** 1387
[121] Szekeres P 1975 *Commun. Math. Phys.* **41** 55
[122] Szekeres P 1975 *Phys. Rev.* D **12** 2941
[123] Tegmark M 1998 *Ann. Phys., NY* **270** 1
[124] Tipler F J, Clarke C J S and Ellis G F R 1980 *General Relativity and Gravitation: One Hundred Years after the Birth of Albert Einstein* vol 2, ed A Held (New York: Plenum)
[125] Tolman R C 1934 *Proc. Natl Acad. Sci., USA* **20** 69
[126] Wainwright J 1988 *Relativity Today* ed Z Perjes (Singapore: World Scientific)
[127] Wainright J, Coley A A, Ellis G F R and Hancock M 1998 *Class. Quantum Grav.* **15** 331
[128] Wainwright J and Ellis G F R (ed) 1997 *Dynamical Systems in Cosmology* (Cambridge: Cambridge University Press)
[129] Wald R M 1984 *General Relativity* (Chicago, IL: University of Chicago Press)
[130] Wald R M 1983 *Phys. Rev.* D **28** 2118
[131] Weinberg S W 1972 *Gravitation and Cosmology* (New York: Wiley)
[132] Wheeler J A 1968 *Einstein's Vision* (Berlin: Springer)