Class. Quantum Grav. **16** (1999) A37–A75. Printed in the UK PII: S0264-9381(99)00323-8

83 years of general relativity and cosmology: progress and problems

George F R Ellis

Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch 7701, Cape Town, South Africa

E-mail: ellis@maths.uct.ac.za

Received 22 December 1998

Abstract. This paper considers the evolution of the relation between gravitational theory and cosmology from the development of the first simple quantitative cosmological models in 1917 to the sophistication of our cosmological models at the turn of the millenium. It is structured around a series of major ideas that have been fundamental in developing today's models, namely: 1, the idea of a cosmological model; 2, the idea of an evolving universe; 3, the idea of astronomical observational tests; 4, the idea of physical structure development; 5, the idea of causal and visual horizons; 6, the idea of an explanation of spacetime geometry; and 7, the idea of a beginning to the universe. A final section considers relating our simplified models to the real universe, and a series of related unresolved issues that need investigation.

PACS numbers: 0420, 9880

Contents

1. Introduction

The application of general relativity theory to the study of cosmology gave rise to the first quantitative cosmological models in 1917—Einstein pioneered the way with the Einstein static universe, followed by de Sitter with his stationary but empty world†. Since then developments have been characterized by a series of major ideas that have been introduced and shaped theory. This has evolved from a simple application of spacetime geometry with a perfect fluid matter source, to a sophisticated physical theory for complex matter sources with highly developed observational implications that enable detailed testing of the realism of cosmological models, thus making them an important part of present-day astronomy. Furthermore, it has become clear that major features of present-day astronomical structures owe their form to the mode of evolution of the universe in general and probably to some high-energy processes of fundamental physics occurring in the early universe, in particular; so cosmological models now also form the broad framework for astrophysics, as well as providing tests of aspects of fundamental physics.

This paper reviews this series of developments from the viewpoint of a relativist. The underlying basic theme is a vindication of Einstein's vision that gravitational effects are embodied in spacetime curvature, determining the motion of matter and radiation as well as the evolution of spacetime itself: in this case, both the evolution of the universe on the large scale, and of the structure embodied in the universe on smaller scales. Because of the embodiment of the gravitational field in curved spacetimes (possibly with a non-trivial global topology), without the benefit of any flat background spacetime as a reference frame, understanding these effects properly implies a continual attention to the role and nature of the coordinate systems used. This has led to the development of covariant and gauge-invariant methods of analysis that greatly clarify our understanding; this is an important theme that runs through the developments discussed below.

What follows is selected highlights, in a more or less historical order, with an eclectic list of references. If followed up fully, these should give access to the main body of literature, which is immense (the literature on inflation alone runs into many thousands of references).

2. The idea of a cosmological model

The initial achievement was the implementation of quantitative self-consistent cosmological models (Einstein 1917, de Sitter 1917a), representing the universe as a whole, with its local structure governed by known and tested laws of physics. In particular, the large-scale structure described by the models is determined by the law of gravitational attraction.

Such models were only achieved once general relativity‡ had been adopted as the (classical) theory of gravitation, despite the complexity of that theory compared with Newtonian theory. The first self-consistent Newtonian models were obtained by McCrea and Milne (1934a, b), 17 years after the first self-consistent general relativity models. The problem preventing earlier Newtonian cosmological models was how to sum the gravitational field due to an infinite set of particles, leading to divergences and indeterminate results when Newtonian theory is applied in a cosmological context§. In general relativity, this is avoided by the local field approach underlying that theory, the gravitational equations being differential equations for the gravitational field (unlike Newtonian theory which is expressed as an inverse

† For references to the original papers, discussed in context, see North (1965) and Ellis (1989, 1990).
‡ See Misner *et al* (1973), Hawking and Ellis (1973) and Wald (1984) for in-depth presentations.
§ See chapter 2 of North (1965).

square force law†). The issue of boundary conditions for these fields is circumvented, in the simple cosmological case, by symmetry considerations (the assumed homogeneity of the universe, justified by *a priori* adoption of a cosmological principle, see, for example, Bondi (1960)), or by assuming closed spatial sections (as in the Einstein static universe).

2.1. The nature of cosmological modelling

The first cosmological models set the basic method that has been followed ever since in most approaches to cosmological modelling‡: (a) a matter source is prescribed (in the Einstein static universe, pressure-free matter, also called 'dust'); (b) a spacetime geometry is assumed for the large-scale structure of the universe (in the Einstein static universe, a static spatially homogeneous and isotropic geometry), described by giving a spacetime metric in suitable coordinates; (c) the field equations express the way in which the matter in the universe locally determines the Ricci curvature, and hence the spacetime geometry; these equations are solved to determine any remaining unknowns (parameters or functions) in the spacetime metric, and then (d) the remaining parameters are compared with astronomical observations in order to fit the model to the real universe (in the case of the de Sitter universe, parallax measurements and paradoxically§ estimates of the density of matter in the universe were used to estimate its size, see de Sitter (1917b)—the first observational paper). Both the initial models included a cosmological constant Λ, required in order to obtain unchanging solutions of the field equations.

The first models also set the framework for representing the geometry of the universe: a high degree of symmetry was assumed, specifically spatial homogeneity and isotropy of both the matter distribution and the spacetime geometry‖. Making this assumption at that time was a bold step indeed, as it was not then known whether the observed 'nebulae' were dust clouds in our own Galaxy, or similar size systems at a much greater distance; as far as the large-scale distribution of matter was known, it was highly anisotropic (concentrated in the Galactic disc) and inhomogeneous. Later measurements have of course shown that these symmetry assumptions are fulfilled to a very high accuracy (on large enough scales), so the gamble of making this assumption, contrary to the observational evidence available at the time, paid off handsomely.

2.2. Unchanging spacetimes

The major error—in hindsight—was the assumption of an unchanging (static) spacetime and matter content. The inbuilt prejudice leading to this assumption—shared by all the major figures in the field for 14 years—rested on a particular philosophical view that was taken for granted at the time. It made modelling geometry of these spacetimes easier, but not greatly so; the dynamics of course was much simpler—being trivial. That philosophical view continues to hold an appeal for many; it was revived in the expanding steady-state universe (Bondi and Gold 1948, Hoyle 1948), supported by the perfect cosmological principle (Bondi 1960), and again in a modified form in such models as chaotic inflationary universes that are statistically

† A potential formalism can be used for Newtonian cosmology (McCrea and Milne 1934a, b)—provided the usual boundary conditions are abandoned, allowing divergent potentials, cf Einstein (1917), Heckmann and Schücking (1956, 1959).

‡ For a discussion of alternative approaches, see Matravers *et al* (1995).

§ Because it is an empty universe!

‖ Mathematically expressed in a seven-dimensional group of isometries, the link between the symmetries of the matter and the spacetime following from the Einstein field equations.

in a steady state, even though local regions are rapidly evolving (Linde 1987, 1990), and the quasi-steady-state universe (Hoyle *et al* 1993, 1995).

The surprising omission was the lack of a consideration of the underlying assumption that one could consistently have non-equilibrium local processes continually taking place in a static universe. Reflection will show that this is deeply inconsistent—one cannot have continual entropy generation in an unchanging universe†. One indication of this feature that was known at the time was Olbers' paradox (see Bondi 1960, Harrison 1990). Precisely because of the link between matter and spacetime geometry embodied in the Einstein field equations (EFE), the continual evolution of the state of matter implied by the second law of thermodynamics must imply an evolution of spacetime geometry.

3. The idea of the evolving universe

Progress now lay in acceptance of the idea of an evolving universe, particularly as embodied in the Robertson–Walker family of evolving spacetimes, and determination of their dynamical evolution. The matter present controls the dynamic evolution of spacetime, the expansion of which in turn controls the evolution of matter.

Friedmann's evolving models (1922, 1924), discovered independently by Lemaître (1927), with the spatially homogeneous and isotropic time-dependent geometries later examined in more detail by Robertson and Walker‡, met with initial resistance or indifference (see, for example, Ellis 1990). However, they received widespread acceptance in 1930, on the one hand because of Eddington's proof of the instability of the Einstein static universe (Eddington 1930) and on the other hand because of the realization of the natural way in which they explained the linear redshift–distance relation for galaxies, which had been placed on a sound footing by Hubble (1929), as summarized in Hubble (1936)§.

3.1. Robertson–Walker geometry and time evolution

The first issue is what it means for the universe to evolve, rather than just for the matter to evolve in an unchanging spacetime. Confusion arises because in high-symmetry universes (specifically, flat spacetime and the de Sitter universe) the matter flow lines are not unique and so there are multiple Robertson–Walker (RW) metric forms possible. In particular, the de Sitter spacetime can be represented in a static form, and by expanding RW metrics of positive, negative, or zero spatial curvature (see Schrödinger 1956); but this possibility arises only because the matter tensor is degenerate (it is Lorentz-invariant, representing a cosmological constant, or equivalently a perfect fluid with $(\mu + p) = 0$).

We start by assuming large-scale spatial homogeneity and isotropy about a particular family of worldlines. The RW models used to describe the large-scale structure of the universe embody those symmetries exactly in their geometry. It follows‖ that comoving coordinates can be chosen so that the metric form takes the form:

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + S^2(t)\,\mathrm{d}\sigma^2, \qquad u^a = \delta^a{}_0 \tag{1}$$

where the surfaces $\{t = \text{constant}\}$ are the surfaces of homogeneity, $\mathrm{d}\sigma^2$ is the metric of a 3-space of constant curvature $k = \pm 1$ or 0, $S(t)$ is the scale factor and the worldlines with

† Unless there is either a modification of the energy conservation laws, as in the steady-state universe, or a timelike singularity present that can continually act as an entropy sink (cf Ellis *et al* 1978), which cannot happen in a spatially homogeneous model. Neither possibility was contemplated in the Einstein static universe.

‡ See Robertson (1933, 1935) and Walker (1936, 1944).

§ However, Hubble never fully embraced the idea of the expanding universe (see Hubble 1953).

‖ See, for example, Robertson (1935), Walker (1936), Ehlers (1961), Weinberg (1972) and Ellis (1987b).

tangent vector u^a represent the motion of fundamental observers. The matter tensor necessarily takes a perfect fluid form relative to these worldlines, and defines the energy density μ and pressure p. Provided

$$(\mu + p) > 0, \tag{2}$$

which we assume in what follows†, u^a is the unique timelike eigenvector of the Ricci tensor, hence the total energy density and total pressure are uniquely defined scalar invariants and the fundamental worldlines and surfaces of symmetry are uniquely determined by the spacetime geometry‡. The space sections have 3-curvature $K = k/S^2(t)$; as $S(t)$ increases, the distance between comoving matter increases, and (assuming standard matter behaviour), the matter density decreases because of the energy conservation equation

$$\dot{\mu} + 3(\mu + p)\dot{S}/S = 0. \tag{3}$$

When $k \neq 0$, the curvature of the 3-space sections is changing, so they are not isometric to each other—as $S(t)$ increases, they become flatter. However, when $k = 0$ these 3-spaces are flat, and hence are isometric to each other and their geometry is unchanging. The spacetime Ricci scalar R is given by $R = \kappa(\mu - 3p) + 4\Lambda$, and so will change except in the case of a cosmological constant ($(\mu+p) = 0 \Rightarrow \dot{\mu} = 0$) or pure radiation ($(p = \mu/3) \Rightarrow \dot{R} = \kappa\dot{\mu}(1-3\mathrm{d}p/\mathrm{d}\mu) = 0$).

Consequently, from these features, in a RW universe obeying the condition (2), as $S(t)$ increases:

- the fundamental worldlines are uniquely defined by the spacetime geometry and the distances between all pairs of these worldlines increases,
- the uniquely determined 3-spaces of homogeneity have a decreasing magnitude of the scalar curvature K if $k \neq 0$;
- if these space sections are compact, either because $k = +1$, or because $k = 0$ or -1 but with a compact topology (see Ellis 1971b), they have a finite volume which increases (the space sections become larger);
- the spacetime invariant R changes, except in the case of a pure radiation universe (with exactly $p = \mu/3$), or a cosmological constant with no other matter present;
- the scalar invariant μ decreases in all cases, reflecting the expansion of the matter in the spacetime.

Thus there can be no timelike Killing vector in these spaces: the variation is intrinsic to the nature of the spacetime rather than being an artefact of the coordinate choice. Putting this together, we can justifiably refer to an *expanding universe*, noting that this is a concept depending on a time slicing by a preferred time parameter. Spacetime itself is changing in that its matter density, distance between uniquely defined worldlines and (provided $k \neq 0$) space sections change as this time increases—we do not just have matter expanding into a spacetime that is itself static, as in the Milne universe (where $(\mu + p) = 0$).

3.2. Friedmann–Lemaître dynamics

The second issue is the dynamic determination of the evolution of a universe with RW geometry by the matter and fields present, when (following Friedmann and Lemaître) the EFE are

† This is an energy condition obeyed by all physically plausible matter in a non-empty spacetime—which we assume to be the case when dealing with large-scale description of a cosmology, because there *is* matter in the universe. The vacuum condition $(\mu + p) = 0$ may, however, be true in many regions on small scales (cf the discussion of the averaging problem in section 9, see point (f)). We always assume $(\mu + p) \geqslant 0$, whatever scale of description we use.

‡ This is not the case in Minkowski and de Sitter spacetimes, where $(\mu + p) = 0$—when this relation is satisfied exactly, only a cosmological constant is present, so these are classed as empty (vacuum) spacetimes.

assumed to determine that time evolution. We call universes of this kind, Friedmann–Lemaître (FL) models (as opposed, for example, to Milne universes, which have the same geometric structure but no gravitational equations). The key equation is the Friedmann equation†:

$$3\dot{S}^2/S^2 - \kappa\mu - \Lambda = -3K, \tag{4}$$

controlling the expansion of the universe, and the conservation equation (3) controlling the density of matter as the universe expands. Equation (4) is just the Gauss equation relating the 3-space curvature to the 4-space curvature, showing how matter directly causes a curvature of 3-spaces in cosmology (Ehlers 1961), and also is a first integral of the Raychaudhuri equation (5) and the conservation equation (3) in any expanding FL universe. Given a determinate matter description (determining the equation of state $p = p(\mu, t)$ explicitly or implicitly) for each matter component, existence and uniqueness of solutions follows both for a single matter component and for a combination of different kinds of matter. Initial data for such solutions at a time t_0 consist of the Hubble constant H_0, density parameter $\Omega_0 = \kappa\mu_0/3H_0^2$ for each type of matter present, and either the corresponding quantity Ω_Λ for the cosmological constant Λ, or the deceleration parameter $q_0 = -(\ddot{S}/S)_0 H_0^{-2}$ if $\Lambda \neq 0$; given the equations of state for the matter, this then determines a unique solution, i.e. a unique corresponding universe history‡.

The dynamical behaviour of these models has been investigated in depth: first for dust plus a cosmological constant, followed by perfect fluids, kinetic theory solutions and fluids with bulk viscosity, and scalar field solutions§, the latter introducing the important idea of the effect on the expansion of the universe of the broken symmetries of particle physics (Guth 1981). Current models employ a realistic mixture of matter components (baryons, radiation, neutrinos, scalar field, cold dark matter and perhaps a cosmological constant‖). Informative phase planes show clearly the way higher-symmetry (self-similar) models act as attractors and saddle points for the other models¶.

These models are the standard models of modern cosmology, and are surprisingly effective in view of their extreme geometrical simplicity. One of their great strengths is their explanatory role in terms of making explicit in a clear way the idea of the local gravitational effect of matter and radiation determining the evolution of the universe as a whole, this in turn forming the dynamic background for local physics (including the evolution of the matter and radiation).

3.3. The hot big bang and ages

The third issue is that important features of the dynamical evolution hold generically for a large class of realistic matter models. The central equation here is the specialization of the Raychaudhuri equation below (8) to the FL case, expressing how the motion of the fundamental

† Originally obtained by Friedmann (1922) for $k = +1$ and positive Λ; and then by Friedmann (1924) for $k = -1$, but in a slightly confused way (see the translation of Friedmann's papers by Ellis and van Elst, to appear in *Gen. Rel. Grav.*); and finally by Robertson (1929) for $k = 0$—the simplest case.

‡ These quantities then determine the spatial curvature $K_0 = k/S_0^2$ and hence the present value of the scale function S_0 if $k \neq 0$. A given solution will have different values of these parameters at different stages of its history, so the evolution of a universe model corresponds to a curve in this parameter space, which can be represented as a phase plane; t_0 characterizes at what instant in that history the observations are being made.

§ See, respectively, Robertson (1933); Kramer *et al* (1980, section 10.2); Walker (1936), Ehlers *et al* (1968) and Treciokas and Ellis (1971); Kolb and Turner (1990) and Ellis and Madsen (1991), for these cases and references.

‖ As is well known, Einstein abandoned the cosmological constant in view of the expansion of the universe, but others such as Eddington insisted on keeping it in the equations. Present-day astronomical observations suggest it may be non-zero, see section 7.2. The idea of a 'varying cosmological constant' contradicts the very clear reason for its inclusion in the EFE, namely its spacetime constancy; one should rather refer to a scalar field term, a varying gravitational constant, or similar.

¶ See Stabell and Refsdall (1966), Madsen and Ellis (1988), Ehlers and Rindler (1989), Wainwright and Ellis (1997), Goliath and Ellis (1999) and references therein.

observers feels the spacetime curvature. This is also the geodesic deviation equation for the preferred timelike worldlines†:

$$3\ddot{S}/S = -\tfrac{1}{2}\kappa(\mu + 3p) + \Lambda. \tag{5}$$

It shows that the active gravitational mass density of the matter and fields present is $(\mu + 3p)$. For ordinary matter this will be positive:

$$(\mu + 3p) > 0. \tag{6}$$

When this inequality is satisfied, one obtains the

FL singularity theorem.‡ *In a FL universe where* $(\mu + 3p) > 0$ *at all times and* $\Lambda \leqslant 0$*, at any instant when* $H_0 = \frac{1}{3}\Theta_0 > 0$ *there is a time* $t_0 < 1/H_0$ *ago such that* $S(t) \to 0$ *as* $t \to t_0$*; a spacetime singularity occurs there, where* $\mu \to \infty$ *and* $T \to \infty$ *for ordinary matter (with* $(\mu + p) > 0$*).*

The underlying physical feature is the nonlinear nature of the EFE: going back into the past, the more the universe contracts, the higher the active gravitational density causing it to contract even more. The pressure p that one might have hoped would help stave off the collapse makes it even worse because (as a consequence of the form of the EFE) p enters algebraically into the Raychaudhuri equation (5) with the same sign as the energy density μ.

This conclusion can, in principle, be avoided by a cosmological constant, but in practice this cannot work because we know the universe has expanded by at least a factor of 6, because we have seen objects at a redshift of 5; the cosmological constant would have to have an effective magnitude at least $6^3 = 216$ times the present matter density to dominate and cause a turn around at any earlier time, and could not have remained undetected. However, energy-violating matter components such as a scalar field can avoid this conclusion, if they dominate at early enough times; but this will only be the case when quantum fields are significant.

Thus the major conclusion is that a *hot big bang* (HBB) must have occurred; densities and temperatures must have risen at least to high enough energies that quantum fields were significant—at something like the GUT energy. At very early times and high temperatures, only elementary particles can survive and even neutrinos and photons will have a very small mean free path; as the universe cools down and complex structures start to form, neutrinos and then photons will decouple from the matter and then stream freely through space. The study of these processes is the subject of physical cosmology (see section 5).

Furthermore, from the Raychaudhuri equation (5), in any expanding FL universe with vanishing cosmological constant and satisfying the energy condition (6), ages are strictly constrained by the Hubble expansion rate $H_0 = (\dot{S}/S)_0$: namely,

FL age theorem. *If* $\Lambda = 0$ *and* $(\mu + 3p) > 0$ *holds at all times, then at every instant, the age* t_0 *of the universe satisfies* $t_0 < 1/H_0$.

More precise ages $t_0(H_0, \Omega_0)$ can be determined for any specific cosmological model from the Friedmann equation (4) in terms of the Hubble constant H_0 and density parameter Ω_0; in particular in a matter-dominated early universe the same result will hold with a factor of $\frac{2}{3}$ on the right-hand side, while in a radiation-dominated universe the factor will be $\frac{1}{2}$. Note that this relation also applies in the early universe when the expansion rate was much higher, and hence shows that the hot early epoch ended shortly after the initial singularity; indeed, it

† For a derivation from this point of view, see Ellis and van Elst (1999a).

‡ See Tolman and Ward (1932), Raychaudhuri (1955), Ehlers (1961), Ellis (1971a). Closely related to this are two other important results: (a) a static universe model containing ordinary matter requires $\Lambda > 0$ (Einstein's discovery of 1917), and (b) the Einstein static universe is unstable (Eddington's discovery of 1930).

is the reason why the major physical evolution took place in the first three minutes (Weinberg 1977)—another consequence of the nonlinearity of the equations.

These age limits can be violated by a cosmological constant that dominates the recent expansion of the universe†; indeed, this is where a positive cosmological constant might play an important role, because age limits are one of the central issues in modern cosmology‡.

4. The idea of astronomical observational tests

Making cosmological models part of astronomy depends on development of detailed observational tests of their suitability to describe the real universe, and then observationally determining the set of realistic parameters that can characterize viable models. Without such tests, the models are not a serious part of astronomy (or physics); thus determining their observational properties is an essential part of developing a cosmological model.

What we can discover is determined by the range of possible observations (Harwitt 1984). There are two main kinds of observations: astronomical observations, based on evidence coming to us along the geodesics generating our past null cone (the topic of this section), and 'geological' observations, based on evidence available essentially along timelike worldlines (discussed in the next section). Curiously, the very important cosmic background radiation (CBR) observations can be regarded as either, see section 5.2. The ideal aim is to combine these two types of observations, thus resolving the observational uncertainty (see section 4.5), but this has not yet been concluded successfully, see section 5.4.

Developing observational relations to provide cosmological tests in essence involves two main steps: first, determining the basic effect and resulting formulae, and secondly, refining the tests until they become usable in practice. In each case the elegant simplicity apparent in the basic theory becomes overlaid in the second step by a series of astrophysical, statistical and measurement issues. Understanding them involves detailed modelling of astrophysical sources, without which successful interpretation is impossible. Thus cosmological observations become deeply entwined with and dependent on astrophysical understanding, and considerable ambiguity and uncertainty enters. Determining how best to handle this is an essential part of successful cosmological tests. The final step is then determining the limits of what can and cannot be established by such tests.

4.1. FL observations

We consider observations in FL universes here; those in more general universes will be discussed below (sections 7.1 and 7.3). In principle, the observational parameters determining a FL cosmology—as discussed in section 3.2, H_0, Ω_0 for each type of matter present, and the deceleration parameter q_0 if Λ is non-zero—depend only on conditions 'here and now'. However, in fact we have to observe to some depth to determine these parameters, and also to verify that a RW metric is indeed a reasonable description of the observable part of spacetime. The light by which we receive this information travels to us along null geodesics, which 'feel' the spacetime curvature and convey information on that curvature to us. This also implies that as we look further out in distance§ we necessarily look further back in time—we see

† Or in principle by other matter components that violate the energy condition (6); but they presumably are irrelevant to the age of the universe defined from the end of the quantum-dominated era to the present day, because it is implausible that such fields are significant during that epoch.

‡ See, for example, Gott *et al* (1974), Ostriker and Steinhardt (1995), Gottlöber and Börner (1997) and Coles and Ellis (1997).

§ Coordinate distance but not necessarily proper distance (see Ellis and Rothman 1993).

the sources as they were a long time ago, giving the idea of 'lookback time' associated with each redshift. We can distinguish observations of distant sources (galaxies, radio sources, quasi-stellar objects (QSOs), x-ray sources, etc), and of background radiation.

4.2. Observations of distant sources

The essential spacetime geometrical elements underlying observations are redshift, area distance and observational volumes†, while the basic observations for each type of source are source apparent size, apparent luminosity and numbers‡. A derived relation between two of these quantities for a given family of sources can be compared with observational data. More detailed observations, e.g. circular velocities in disc galaxies, luminosity fluctuations in ellipticals or the globular cluster luminosity function in galaxies, can be used to give independent estimates of the luminosity or distances of galaxies which can be used in observational relations (see Bothun (1998) for a discussion), but they depend on detailed assumptions about galaxy structure. All the distance relations are normalized through measurements of indicators such as Cepheids and RR–Lyrae stars in nearby galaxies.

Cosmological redshifts in a FL model are determined by the integrated overall expansion of the universe between the time of emission of light t_e and the time of reception t_0, in the form $1+z = S(t_0)/S(t_e)$. Measuring redshifts has become a highly developed and accurate science; their interpretation is not so straightforward because local relative velocities of galaxies lead to Doppler contributions that cannot be distinguished observationally from the cosmological contribution, so redshift observations alone can give a distorted impression of source distances. Thus one has to make some model of source clusters in order to separate out these two components, and the issue then is which objects are to be identified as belonging to a specific cluster and which not. There is a danger of circularity in argument here, or at least of selecting data in a way which favours one's own chosen interpretation; hence the controversy over the interpretation of redshifts (Field *et al* 1973).

Observer area distance is the same as the angular diameter distance in a FL universe, being determined from the matter content of the universe by the geodesic deviation equation, which clearly shows how the matter content determines the variation of distances between neighbouring geodesics (Ellis and van Elst 1999a). The conceptual problem arises here because even in a galaxy or star cluster most of space is empty, so unless intergalactic space is filled uniformly with a high density of undetected dark matter, most of the light rays in the universe travel in empty space between matter. Consequently, see Bertotti (1966), the real focusing effect is caused by the Weyl tensor in the vacuum regions, present through the tidal effects of the matter (i.e. Ricci tensor terms) in those small regions where it is concentrated—mainly the stars in galaxies—rather than through a spatially homogeneous Ricci tensor, as in a RW geometry. Thus on small angular scales angular diameter distances for rays passing between matter will be different than in FL models (Dyer and Roeder 1973, 1981). The question then is why the large-angle averaged effect of the convergence produced by this Weyl-induced distortion (causing weak and strong gravitational lensing) should be the same as would be produced by the Ricci tensor if the universe were exactly uniform (as is assumed when one calculates the area distance for a FL model). This does in fact seem to work out correctly (see, for example, Holz and Wald 1998) but there are conceptual traps here and simple arguments for this result from energy conservation are wrong. This problem is alleviated in a universe dominated by uniformly distributed dark matter, but remains if that matter is clumped.

† Kristian and Sachs (1966), Ellis (1971a), Weinberg (1972) and Ellis (1987b).
‡ Heckmann (1942), Hoyle (1960), Sandage (1961), Gunn (1978) and Sandage (1993).

The first observational problem is the lack of sharp edges of most large objects, which fade off into the background; hence it is difficult to determine a fixed scale distance in images of an elliptical galaxy, for example. Rather one has to measure apparent galaxy size up to some chosen isophote and then interpret the result to determine area distances; this depends on a detailed model of the galaxy profile. The equivalent problem arises if one measures luminosity distances instead, which are the same as area distances (up to redshift factors) because of the reciprocity theorem†. The problem is that one has to decide up to which contour one will collect radiation; again one will usually chose an isophote. Supposing this is solved, one then has to interpret the results; and here one faces a formidable battery of problems, essentially because galaxies and other macroscopic sources are poor standard candles. As well as the intrinsic spread of source properties, we have to contend with source evolution, which is now known to be significant; but we have no good theory of source evolution in most cases.

The bright light within the horizon is that supernovae have the potential to act as excellent standard candles, because their properties are mainly determined by local physical features rather than by astronomical history. Consequently, major programmes are now aiming to detect supernovae in distant galaxies and determine their intrinsic luminosities from their light curves. This has the potential to at last turn the theoretical simplicity of the magnitude versus redshift test of cosmologies (Sandage 1961) into a reality. The problem here again is determining cluster membership: does the galaxy in which the supernova is observed have a distance typical of the galaxy cluster? Nevertheless, as statistics build up, we should be able to determine H_0 and q_0 to good accuracy from supernovae observations.

Number counts depend on determining volumes in physical space corresponding to an increase in some observable parameter related to distance. As distance is not directly measurable, we have to use some surrogate, often source luminosity‡. This can be easily worked out in the FL case to give formulae for expected numbers of sources in a given volume, which can be used to test for spatial homogeneity and in principle to determine cosmological parameters; indeed, the first investigations of the geometry of the universe by Hubble were based on number counts. However, in order to utilize number counts to test spacetime geometry, we must understand selection effects, determining which sources are detected and included in any systematic catalogue and which are not. This will depend at least on source luminosity (or magnitude), apparent size and spectrum, as well as on cosmological distance and intervening matter. A series of observational problems arise.

Firstly, there is the unknown evolution of source numbers and luminosities, strongly affecting which are detected and which not; so we have to model this. This is essential: observations cannot consistently fit the data—particularly for radio sources (Ryle and Scheuer 1955, Longair 1978) and QSOs—in a FL universe, unless there is substantial source evolution§. Secondly, there is a considerable variation in source intrinsic luminosity, so we have to model this (via a luminosity function). Thirdly, there may be intervening matter that causes absorption and reddening, which also has to be modelled. This in principle allows detection of matter (as in the Lyman–Forest observations in QSO spectra), failing which one can put limits on the amount of matter present (as in the Gunn–Petersen (1965) limits on neutral hydrogen, determined through lack of a Lyman-alpha absorption trough in distant QSO spectra). Putting these together, there is a substantial modelling input needed before we can determine the

† See Kristian and Sachs (1966) and Ellis (1971a). This results from a first integral of the geodesic deviation equation, and remains true in anisotropic models. It is responsible for the important fact that radiation intensity (and hence CBR temperature) is independent of area distance—depending only on redshift in both FL and anisotropic models.

‡ Redshift would be better, but even with present-day fibre optics technology it is time consuming obtaining a large enough number of redshift measurements to give good statistics.

§ This also is a major reason the steady-state universe cannot be reconciled with observations; but see Hoyle *et al* (1993, 1995).

detection function required to adequately interpret the number count observations—or indeed the statistics of any of the other tests, such as an angular diameter versus redshift or a magnitude versus redshift tests; and the nature of the problem will change with observational wavelength (optical, infrared, radio, x-ray, etc).

Two specific issues are worth mentioning here. First, it now seems highly likely that there is a dominant amount of dark matter in the universe† whose detection is extremely difficult because it interacts very little with ordinary matter, and so neither emits nor absorbs much radiation. Hence the detection function for this matter through direct observations is very small indeed; its main effects are via gravitational interaction, allowing us to detect it by its effect on galaxy rotation curves and galaxy velocities, as well as by its contribution to area distances and ages. Secondly, the detection of distant galaxies depends both on the apparent luminosity of the galaxy as well as its apparent size, because detector response relates to surface brightness levels. Thus low surface brightness galaxies can be missed, see Disney (1976). Furthermore, distinguishing galaxies from other objects demands an adequate degree of resolution relative to their distance. Consequently, any adequate theory of detection limits must be based on at least two parameters characterizing the sources observed (e.g. their intrinsic size and intrinsic luminosity) and their images (e.g. their apparent size and apparent luminosity), these quantities being related to each other by a well defined observational map for each class of objects, see Ellis *et al* (1984). However, many analyses do not adequately take this into account, for example, referring to 'magnitude-limited samples', implying only one parameter can be used to characterize detection limits. One consequence is that in order to properly analyse the effects of evolution on source detection, one must also model the change of source size with epoch, as this has a stronger effect on surface brightness than does luminosity evolution. The importance of these considerations is underlined by the growing evidence for faint populations of galaxies that are difficult to detect unless one is specifically looking for them.

4.3. Background radiation

As well as observations of distinguishable sources, one can observe background radiation (i.e. integrated radiation from sources that cannot be individually distinguished) in general, and the 2.75 K CBR, in particular. The integrated radiation from all sources at all wavelengths provides useful cosmological constraints on the amount of matter present in various forms and its temperature history (e.g. restricting the amount of hot intergalactic matter present from limits on the observed x-ray background). In effect, this is the modern version of the Olber's paradox calculation (Bondi 1960). The CBR‡ is identified by its isotropy and its black-body spectrum, and is interpreted as remnant radiation from the HBB. Its existence provides strong confirmation of the physical models of early eras, discussed in the next section. Detailed testing of the CBR spectrum and anisotropy pattern is possible (Partridge and Wilkinson 1967, Smoot *et al* 1992), and has become one of the central concerns of observational cosmology at the turn of the century. This will be discussed in what follows.

4.4. RW topology

It is possible that the universe has compact spatial topology, even if the curvature is zero or negative (see Ellis 1971b), the point being§ that the EFE (local equations for spacetime

† See, for example, Bahcall *et al* (1987), Coles and Ellis (1997), Audouze and Tran Thanh Van (1988) and Bothun (1998) for a summary of evidence.
‡ See, for example, Partridge (1995) for a comprehensive review.
§ As realized already by Friedmann in 1922.

curvature) cannot by themselves determine the topology of spacetime, or of space sections in spacetime. If the space sections were compact with small enough size (and a non-trivial topology), we would be able to see multiple images of the same objects in different directions in the sky. Such a situation (a 'small universe') is testable in principle through detecting multiple images of the same object; but in practice this is very difficult to establish observationally (Ellis and Schreiber 1986), although some interesting proposals have been made (see, for example, Roukema 1996, Roukema and Edge 1997). An exciting recent development is that this possibility can be tested observationally by searching for different (metric) circles in the sky on which the same pattern of CBR temperature fluctuations occur (Cornish *et al* 1998). Such observations could, in principle, determine the spatial topology (which could be extremely complicated).

4.5. The observational predicament

The experience gained in developing and implementing these observational tests† is fundamentally important, as it is what makes these models a serious part of astronomy. The aim is an integration of all available data to determine the basic cosmological parameters‡. General relativity plays a central role in the formulation of the observational tests, but astrophysical issues become central in implementing them. This essential partnership means that one has to include in a complete cosmological model, a description of the relevant source population and their astrophysical behaviour (their time evolution properties, for example), which are then tested along with the cosmology itself.

One then runs into the fundamental modelling problem: to be astronomically useful, one must include all this further information in a cosmological model; but then, even if one adds further observational tests to try to determine the new parameters and functions entering, the description one uses may contain as many arbitrary functions and/or parameters as there are observational tests. Any set of observations can then be described by these models, irrespective of the geometry, because of the plethora of arbitrary functions and parameters§. The hope is to reduce this arbitrariness by giving some of these functions a proper physical grounding.

5. The idea of physical structure development

The expanding universe provides the context for local physics to determine what structures originate from an initially highly uniform big bang (this uniformity, at least at the time of decoupling of matter and radiation, being indicated by the CBR near-isotropy, see section 7.1). These structures develop at all scales, from the formation of baryons out of quarks and of nuclei out of protons and neutrons in the hot early universe, to the formation of galaxies and clusters of galaxies after the decoupling of matter and radiation.

Physical cosmology studies the development of this variety of structures. As they are all made of massive particles, they move at speeds less than the speed of light; indeed, almost without exception at very low speeds relative to the cosmic rest frame. Thus when we observe the nature of the resulting structures in our vicinity, they give us information on conditions very close to our past worldline back to very early times, and hence can be called timelike observations (in contrast to the null observations discussed in the previous section)‖.

† See Bothun (1998) for a good survey.

‡ See, for example, Gott *et al* (1974), Gunn (1978), Ostriker and Steinhardt (1995), Coles and Ellis (1997).

§ Cf Peebles (1998); and see Mustapha *et al* (1998) for a specific example.

‖ Cf Hoyle (1960) and Ellis (1971a). They are sometimes characterized as observations of a *geological* nature, as geology is a familiar example of such observational tests of very early conditions in our history.

This physical evolution depends critically on the physical behaviour of the matter present, usually expressed in terms of equations of state for the different matter components, together with equations describing their interactions with each other. The simplifying factor is that most of the time before decoupling, the mixture of particles and radiation is close to equilibrium (because the relevant reaction rates are much more rapid than the expansion rate). After decoupling, however, the situation is essentially non-equilibrium†.

5.1. Before decoupling: particles and radiation

Local interactions before decoupling cause a physical evolution that is now well understood from a temperature of about 10^{10} K and below, including pair (re)combination, neutrino decoupling, nucleosynthesis, through recombination and decoupling of matter and radiation, and also (less well understood) the quark–hadron transition and baryosynthesis. The relics from each of these epochs (particles, nuclei, atoms, radiation) provide evidence of the interactions that took place at those early times. The CBR (the black-body relic radiation from the HBB era, freely propagating to us since the time of decoupling) is convincing evidence both of the existence of this hot early era, and of the close to equilibrium nature of interactions then (the measured spectrum of this radiation being a perfect black-body spectrum within the error limits). Interactions at this time—particularly pair production and annihilation at very early times and Thomson scattering at later times—keep the matter and radiation tightly coupled. The primary role of gravity is in regulating the rate of change of temperature with respect to time (determining the expansion rate of the universe, via the Friedmann equation). This plays a crucial role, for example, in nucleosynthesis (see, for example, Peebles 1966), which is a successful theory giving results that accord with observations (see, for example, Schramm and Turner 1998) and provides one of the main pillars for our belief in the HBB picture (Peebles *et al* 1991).

The possibility of analysis of this kind was already known to Lemaître, who tried to analyse cosmic rays as relics from the early universe (Lemaître 1931b). The major breakthrough came after the 1939–1945 war‡, when the theory of nuclear reactions was on a sound footing. After the pioneering work by Gamow (1946), developed in a series of papers by Alpher, Herman and Hayashi§, the discovery of the CBR‖ and its relation to nucleosynthesis¶ brought the study of this era into mainstream physics[+].

The study of the interaction of matter and radiation during this era is essentially carried out by considering interactions taking place in a typical small cell at each epoch (Harrison 1977), for example, absorption and emission of radiation can be modelled in such a cell. Apart from the interactions mentioned above, a key feature is that density perturbations will cause acoustic oscillations on scales less than the Jeans' length (where radiation pressure succeeds in counteracting gravitational attraction) and growth of inhomogeneities on scales greater than the Jeans' length (where gravitational attraction wins). Thus any initial spectrum of density fluctuations (left over, for example, from an initial quantum gravity era or an early

† 'Decoupling' means that the interactions which used to keep the different components in close contact with each other are no longer strong enough to do so.

‡ One of the puzzles is how Tolman failed to develop the subject further, after making a promising start on examining thermodynamic relations in the early universe (Tolman 1934). He should at least have been able to predict the existence of the CBR.

§ And the magnificent paper by Burbidge *et al* (1957) setting nucleosynthesis in stars on a sound footing.

‖ Penzias and Wilson (1965) and Dicke *et al* (1965).

¶ Peebles (1966, 1971), Wagoner *et al* (1968) and Weinberg (1972).

[+] Signified by the inclusion of cosmology in the biennial particle physics summary in *Reviews of Modern Physics*, see, e.g., Barnett *et al* (1996), particularly: *Big Bang Cosmology*, 708–9; *Big Bang Nucleosynthesis*, 710–12; *The Hubble Constant*, 713–16; *Dark Matter*, 717–18; and *Cosmic Background Radiation*, 719–22.

era of inflationary expansion) will become modified by these processes and provide the seed fluctuations at the time of matter–radiation equilibrium†, from which the growth of structures by gravitational instability will take place‡.

5.2. After decoupling: astronomical structure formation

After decoupling, gravitational attraction forms local structures out of initial fluctuations in cold dark matter (CDM) and baryonic matter, these inhomogeneities then generate relative motion of matter by their gravitational attraction. Although radiation interacts with matter in many ways as structures form§, only at a late stage of collapse is sufficient heat generated to become important in the local dynamical evolution of the matter. Free-streaming radiation conveys information on perturbations at last scattering to the observation event (here and now), also feeling alterations in the gravitational potential that occur along the way. Thus the temperature anisotropies measured in the CBR reflect the potential inhomogeneities at last scattering, modified by small-scale spacetime curvature encountered along the way‖.

Herein lies a problem. Because we can only solve the EFE in very special (high-symmetry) nonlinear cases, we of necessity have to carry out analytic studies of the growth of structures in 'almost-FL' universe models, which can be regarded as linearized perturbations of FL solutions (Lifshitz 1946). In this situation the representation of the perturbations, in particular, the velocities and the gravitational potential, is gauge dependent, where 'gauge' refers to the choice of background spacetime in the lumpy universe; so the answer obtained can depend on the gauge used. The issue¶ is that the perturbation δT in any quantity T is defined at each point by

$$\delta T = T - \overline{T} \tag{7}$$

where $\overline{T}$ is the background value at that point. But, precisely because there is no fixed background spacetime in general relativity, we can choose any correspondence we like[+] between the background spacetime $\overline{M}$ and the lumpy universe model M, and so can give any value to the perturbations; for example, we can always chose a gauge that will set the density perturbation $\delta\mu$ to zero (by choosing the background surfaces of constant density to be the same as the real surfaces of constant density).

There are two ways to handle this problem. One is to very carefully keep track of the gauge used and the remaining gauge freedom[*], thereby assigning a (gauge-dependent) meaning to the variables used, and identifying which of the perturbation modes found are pure gauge modes and which are physical. This is the approach adopted by many; however, the history of the subject is not encouraging, particularly when some published papers that claim to fully sort out the problem contain errors that result in even greater confusion[#], and some influential

† The Jeans' length drops drastically at this time, where a transition takes place from the earlier radiation-dominated era to the later matter-dominated era. When this takes place depends on how much cold dark matter is present.

‡ Rees (1978, 1987, 1995); Peebles (1971, 1980, 1993).

§ Field (1969), Sciama (1971) and Longair (1993)

‖ Sachs and Wolfe (1967), Peebles (1980), Padmanabhan (1993), Hu and Sugiyama (1995a, b), Hu and White (1996) and Jones and Lasenby (1998).

¶ Sachs and Wolfe (1967), Stewart and Walker (1974), Bardeen (1980) and Ellis and Bruni (1989).

[+] Defined by a map Φ from $\overline{M}$ into M, often represented by giving coordinates for the background model in the more realistic one.

[*] See, for example, Sachs and Wolfe (1967), Bertschinger (1992) and Ma and Bertschinger (1995).

[#] For example, in analysing density perturbations of pressure-free matter, finding solutions that have as characteristics the lightcone—when the true characteristics are in fact timelike lines (Ehlers *et al* 1987, van Elst and Ellis 1999). These perturbation modes are unaware of either the speed of sound or the speed of light!

presentations mix gauge choices in one calculation—which will almost always lead to errors—without commenting on this fact.

The other way to handle this is to use gauge-independent variables, such as for example in the important work by Bardeen (1980, 1988) that carefully examined the changes in variables as the gauge is varied, and hence identified combinations of these variables that are gauge independent. Many important analyses of the growth of fluctuations† and of CBR anisotropies‡ use this approach. An alternative gauge-invariant approach (Ellis and Bruni 1989) uses a covariant 1 + 3 splitting of spacetime to define covariant and gauge-invariant tensorial variables with a clear physical meaning; this has also been developed in depth to study growth of inhomogeneities and CBR anisotropies§. In the author's opinion, it is far better to use one of these gauge-invariant approaches than a gauge-dependent one. Whether this is done or not, one of the crucial aspects of application of general relativity to cosmology is being fully aware of this problem, and tackling it in a clear and unambiguous way.

For any serious calculations, the final development will of necessity be numerical because of the number of variables involved and the complexity of their interactions. Furthermore, a numerical approach will also be called for in the nonlinear stages of structure development because we cannot attain analytic solutions for these stages. Numerical cosmology is a large subject in its own right; for a description of what has been done and references, see Anninos (1998). Perhaps most important is determination by numerical means of the transfer function, describing the change of the perturbations spectrum from an initial time t_1 (perhaps the time of decoupling of matter and radiation) to a final time t_2 (perhaps the present day).

Putting this together, we obtain statistical predictions of the distribution of matter (in particular, number densities, luminosity functions, two-point correlation functions) and the expected motions generated by gravitational attractions, which can be tested both directly and by means of gravitational lensing observations (weak lensing distortions, and multiple images and arcs due to strong lensing‖). Furthermore, we also obtain statistical predictions for the accompanying CBR anisotropies to be expected, and specification of the way observations can be used to extract the major cosmological and matter parameters from the data (see, for example, Jungman *et al* 1996). Thus there is presently great activity extending the observations and putting the theory and observations together, finding out what parameter space of the background model is compatible with the overall picture as well as what structure formation scenarios are consistent with all the data. This work is not yet complete—the large and smaller angular scale pictures do not yet fit together in a fully consistent way—but a CDM picture seems broadly consistent, perhaps with an admixture of hot dark matter as well, provided one does not insist that $\Omega_{CDM} = 1$, as the real enthusiasts require.

Two further points are of interest here. Many astrophysicists wish to study structure formation in an essentially Newtonian way. There are particular gauges that allow the relativistic equations to be formulated in a way that is close to Newtonian theory (Bardeen 1980); however, there are consistency problems associated with such gauges that must be examined very carefully if they are to be used in a reliable way. There is still interesting work to be done in attaining a clear nonlinear Newtonian limit of general relativity theory in a cosmological context (van Elst and Ellis 1998).

Secondly, it is clear that after decoupling the CBR reaches us on null geodesics. It is thus

† For example, Bardeen *et al* (1983) and Mukhanov *et al* (1992).

‡ For example, see Panek (1986) for a photon approach and Hu and Sugiyama (1995a, b) for a kinetic theory approach.

§ See, for example, Bruni *et al* (1992), Dunsby (1997), Challinor and Lasenby (1998), Ellis and van Elst (1999b), and references therein for the linear case, and Ellis and Bruni (1989), Maartens *et al* (1999) for nonlinear equations.

‖ See Zwicky (1937a, b), Blandford *et al* (1991), Schneider *et al* (1992), Blandford and Narayan (1992), Sasaki (1993), Fort and Mellier (1994) and Wambsganss (1998).

somewhat paradoxical that many kinetic theory CBR calculations in fact essentially proceed by timelike integrations¶, rather than by null integration (as in the original Sachs–Wolfe paper of 1967). This proceeds essentially by Harrison's method of considering radiation in a typical cell (Harrison 1977), as mentioned in section 5.1, even though the radiation is freely flowing; the implicit assumption that allows this to succeed is that, because of spatial homogeneity, the radiation entering the cell is balanced by that leaving. Interesting issues arise in seeing how the null and timelike integrations give the same results for the CBR anisotropies.

5.3. The arrow of time

An important issue arising in these processes is their non-equilibrium nature, a consequence of the expansion of the universe. It is a consequence of this feature, together with the long-range attractive nature of gravitation, that allows structure formation to take place spontaneously on many different scales in the expanding universe (Dyson 1971, Reeves 1987). Two fundamental issues arise here.

One is the still unresolved issue of the arrow of time—how all macroscopic physics has a strong unidirectional arrow of time governing everything that happens, despite the lack of such an arrow in the underlying microphysical laws. This is widely presumed to relate to the difference between the initial and final boundary conditions of cosmology, but the details of how this happens (and, in particular, how the various physical and biological arrows of time are tied in to each other) remains elusive†.

The second point is that the issue of gravitational entropy is unresolved, except for the case of the entropy of a black hole; but this is fundamentally important in terms of the spontaneous generation of large-scale structure that has happened in the universe, and enabled all the smaller-scale structure growth to take place. Does gravitational entropy exist? If so, what is a satisfactory definition, and what laws does it obey? (Does it always increase, or only under restricted conditions?‡) Does black hole entropy include cosmological horizons (as in de Sitter space)? Despite a number of interesting analyses§, this is one of the outstanding unsolved problems of classical gravitation theory, with fundamentally important implications for physical cosmology. As in the rest of physics, we impose the arrow of time by hand‖ without really resolving the fundamental underlying issues.

5.4. Physical cosmology

In principle, physical cosmology provides the information necessary to close the observational gap, by giving us evolutionary scenarios for the evolving structures in the universe (see, for example, Peebles 1993). In practice this hope has not yet succeeded, although a lot of progress has been made. Part of the problem is the sheer complexity of interactions that arise as structure formation proceeds—specifically, the problems that arise in handling the nonlinear phase of structure formation and the complex interactions as radiation processes start to seriously influence their evolution. Thus we do not yet have a theory of the evolution of galaxies, radio sources or QSOs that is sufficiently developed and tested to adequately determine the

¶ Of a hierarchy of divergence relations for moments of the distribution function.
† See Ellis and Sciama (1972), Davies (1974), Penrose (1979, 1989a) and Zeh (1992).
‡ Gravitation is always an attractive force, causing inhomogeneity to spontaneously increase—but this is true in both directions of time!
§ See, for example, Penrose (1989b) and Hawking and Hunter (1999).
‖ Studies of structure growth, for example, routinely accept the growing solutions (in the forward direction of time) and reject the decaying solutions—but if the latter are included, the relation between density perturbations and CBR anisotropies is fundamentally altered.

evolutionary properties affecting observational tests of cosmological models. However, the theory is approaching that state as regards the CBR anisotropies which are directly related to the seed perturbations leading to large-scale structure, and can be related to measurements of that structure in the context of specific models of the history of the universe (inflation with CDM, for example, see, e.g., Kolb and Turner (1990), Steinhardt (1995), Peacock (1999)). The major problem is that here too, we need extra functions to make the theory and observations fit—in this case, a bias parameter or function, relating the fraction of visible objects formed to the overall matter inhomogeneities (and perhaps an admixture of hot dark matter, a second inflationary potential, or similar). We do not yet have an adequate theory for the bias parameter†, so at present it plays the role of an arbitrary function that can be used to fit theory to observations.

Two related problems arise. The first is that at early enough times, it is difficult if not impossible to obtain laboratory tests of the physical interactions that are dominant at those times. For example, this already becomes a problem in relation to the quark–hadron transition and to baryosynthesis. This becomes much more pronounced when one comes to the times when quantum effects dominate (see section 8.3), and of course applies in particular to the theory of gravity relevant at those times. Indeed, as has been noted by Zel'dovich and others (Zel'dovich 1970, Yoshimura 1978), we have to run the argument the other way round: as the early universe is the only place where such high energies are attained and we can also obtain some observational evidence of their results, we can try to use the early universe as a laboratory for testing the relevant physics. Perhaps the outstanding success of this project so far is the celebrated determination from nucleosynthesis arguments that the number of families of neutrinos should not exceed three (Steigmann *et al* 1977), since verified by measurements at CERN. However, it should be noted that such arguments presume the correctness firstly of our models of the early universe at that time, and secondly of the gravitational equations governing the expansion then. Both are features we would like to test. Hence one would like the arguments to be stated in ways that at least acknowledge the need for openness in both of these issues, and hopefully even try to include them as variables to be tested from the observational data.

The second is the series of problems that arise, with the arrow of time issue being symptomatic, because we do not know what influence the form of the universe has on the physical laws operational in the universe. Many speculations have occurred about such possible effects, particularly under the name of Mach's principle‡, and, for example, made specific in various theories about a possible time variation in the 'fundamental constants' of nature, and specifically the gravitational constant (Dirac 1938). These proposals are to some extent open to test (Cowie and Songaila 1995), as in the case of the Dirac–Jordan–Brans–Dicke theories of a time-varying gravitational constant. Nevertheless, in the end the foundations of these speculations are untestable because we live in one universe whose boundary conditions are given to us and are not amenable to alteration, so we cannot experiment to see what the result is if they are different. The uniqueness of the universe is an essential ultimate limit on our ability to test our cosmological theories experimentally, particularly with regard to the interaction between local physics and the boundary conditions in the universe (Ellis 1999b). This therefore also applies to our ability to use cosmological data to test the theory of gravitation under the dynamic conditions of the early universe.

† See, e.g., Coles and Ellis (1997), Bothun (1998, section 5.2.2). The usual assumption of linear biasing is almost certainly incorrect.

‡ The issue of the origin of inertia and its relation to cosmology, see, e.g., Bondi (1960), Wheeler (1968) and Barbour and Pfister (1995).

6. The idea of causal and visual horizons

A fundamental feature affecting the formation of structure and our observational situation is the limits arising because causal influences cannot propagate at speeds greater than the speed of light. Thus the region that can causally influence us is bounded by our past null cone; combined with the finite age of the universe, this leads to the existence of particle horizons limiting the part of the universe with which we can have had causal connection.

We can only observe on our past null cone, and the *particle horizon* is by definition comprised of the worldlines characterizing the limit of matter that has intersected this null cone (Rindler 1956). This is the limit of matter that we can have had any kind of causal or observational contact with. Such horizons will exist in FL cosmologies for all ordinary matter, unless we live in a small universe with compact space sections, as discussed in section 4.4. Their nature was the subject of considerable confusion initially, specifically because it was difficult to separate out coordinate horizons from causal horizons; however, a classic paper by Rindler (1956) clarified their nature in analytic terms, and Penrose' powerful use of conformal methods (Penrose 1963, see also Hawking and Ellis 1973, Tipler *et al* 1980) gave a very clear geometrical picture of their nature. They may not exist in non-FL universes, cf section 7.4. One may note here some paradoxes associated with the physical size of the horizons in FL models† and the existence and location of the sphere around us beyond which matter is moving away at a speed greater than the speed of light c, that are interesting to examine and clarify (Ellis and Rothman 1993)‡. The horizon always grows: despite many contrary statements in the literature, it is not possible that matter leaves the horizon once it has entered. In a (perturbed) FL model, once causal contact has taken place, it remains until the end of the universe.

The importance of horizons is twofold: they underlie causal limitations relevant in the origin of structure and uniformity (Misner 1969, Guth 1981), and they represent absolute limits on what is testable in the universe (Ellis 1975, 1980).

6.1. Causal limitations

As to causal limitations, horizons are important with regard both to the smoothness of the universe on large scales, and the lumpiness of the universe on small scales. The issue of smoothness is encapsulated in the *horizon problem*: if we measure the temperature of the CBR arriving here from opposite directions in the sky, it came from regions of the surface of last scattering that can have had no causal contact of any kind with each other. Why then are conditions so similar in these widely separated regions§? As to the lumpiness, the issue here is that if we believe there was a state of the universe that was very smooth—as indicated at the time of decoupling, by the low degree of anisotropy of the CBR, and represented by the RW geometry of the FL models—then there are limits to the sizes of structures that can have grown since then by causal physical processes, and to the relative velocities of motion that can have been caused by gravitational attraction. If there are larger-scale structures or higher velocities, these must have been imprinted in the perturbations at the time of last scattering, as they cannot have been generated in a causal way since that time. They are set

† For example, the particle horizon is at a distance $3ct_0$ in an Einstein–de Sitter universe, when the age of the universe is t_0.

‡ For example, the matter that emitted the CBR was moving away from us at a speed of about $61c$ when it did so.

§ Misner (1968) and Guth (1981). Note that this question is of a philosophical rather than a physical nature, i.e. there is no contradiction here with any experiment, but rather an unease with an apparent fine tuning in initial conditions.

into the initial conditions, rather than having arisen by physical causation from a more uniform situation†.

Actually the domain of causal influence is even more tightly constricted than indicated here: the limits coming from the horizon size are limits on what can be influenced by particles and forces acting at the speed of light. However, only freely travelling photons, massless neutrinos and gravitons can move at that speed. Any massive particles, or massless particles that are interacting with matter, will travel slower (for example, before decoupling light has a very small mean free path and information will travel only by diffusion in the tightly coupled matter–radiation fluid). Furthermore, the characteristics for pressure-free scalar and vector perturbations are timelike curves, moving at zero velocity relative to the matter; while density perturbations with pressure can move at the speed of sound, only tensor perturbations can travel at the speed of light‡. Thus the true domain of causal influence is much smaller than indicated by the horizon size.

Associated with the existence of horizons is the prediction that physical fields in different regions in the universe should be uncorrelated after symmetry breaking takes place, because they cannot have interacted causally. Consequently, topological defects such as monopoles and cosmic strings may be expected as relics of the expansion of the very early universe (Kibble 1976, 1980, Kolb and Turner 1990). In a standard cosmology, far too many monopoles are predicted. However, inflationary models solve this problem (Guth 1981).

A major discovery of the inflationary models (Guth 1981) is that there exist physically possible equations of state for fields in the early universe for which exponential expansion will take place; then particle horizons in FL models will be much larger than in the standard models with ordinary matter. Indeed, this will occur if there is a period of accelerated expansion of the universe, possible if $(\mu + 3p) < 0$, which can happen if a scalar field dominates the dynamics, for in the slow-rolling case this leads to $(\mu + p) = 2\dot{\phi}^2 \simeq 0 \Rightarrow (\mu + 3p) \simeq -2\mu < 0$. This then enables a resolution of the horizon problem: if sufficient inflation took place in the early universe, then all the regions from which we receive CBR were causally connected (indeed, if the universe began in an inflationary state, or was inflationary with compact spatial sections, there may be no causal horizons at all). The inflationary models also cause perturbations to die away, including velocity perturbations (hence explaining the observed smoothness of the universe on large scales, see sections 7.2 and 7.4). Assuming no new velocities are generated during reheating at the end of inflation§, this means that the second problem above remains: inflation by itself does not explain large-scale correlations such as cosmic voids, walls, etc, nor systematic velocities, that are larger than can have been generated by gravitational attraction since radiation domination ended.

6.2. Observational limitations

Horizons also imply the important feature of limits to observational possibilities. We cannot see matter beyond the particle horizon (Rindler 1956). Actually we cannot even see beyond the *visual horizon*, comprised of the furthest matter we can observe—namely, the matter that emitted the CBR at the time of last scattering (Ellis and Stoeger 1988, Ellis and Rothman 1993). Attempts to see further fail because the universe was opaque at those times.

† The relevant initial conditions are perfectly possible, but we prefer other conditions we regard as more probable. However, it is highly debatable whether the idea of 'probable' can be applied in any meaningful sense to the initial conditions for the entire universe, see Ellis (1999b).

‡ The Weyl tensor has characteristic modes with associated speeds of travel of $v = 0$ (the Coulomb part), $v = c$ (the transverse part), and $v = c/2$ (the longitudinal part) relative to the fluid velocity in a space with a perfect fluid matter source (see van Elst and Ellis 1999).

§ Which is a far from trivial assumption (see, e.g., Anninos *et al* 1991).

The position of the visual horizon is determined by the geometry since decoupling. The matter present at these times presumably has an ordinary equation of state, and the geometry is plausibly almost-RW then (see the next section). Thus visual horizons do indeed exist, unless we live in a small (spatially closed) universe we can see around. There is no change in these visual horizons if there was an early inflationary period, as inflation does not affect the expansion or null geodesics during this later period. The major consequence of their existence is that many present-day speculations about the super-horizon structure of the universe (e.g. the chaotic inflationary theory) are simply untestable, because one can obtain no definite information whatever about what lies beyond the visual horizon (Ellis 1975, 1980). This is one of the major limits that must be taken into account in our attempts to test the veracity of cosmological models.

7. The idea of the explanation of geometry

If any sensible measure is used on the space of spacetimes, it is clear that the RW geometries are infinitely improbable, because of their very high symmetry (perfect spatial homogeneity and isotropy). Generic cosmologies are anisotropic and inhomogeneous; so the question arises as to why the real universe should conform to such an unlikely model. On reflection, it is clear there are actually two issues: first, is the universe indeed like a FL model, as is usually supposed, or is this a wishful supposition we make because the resulting models are so easy to handle? What is the real evidence that the universe is well represented by a FL model? Secondly, if this is indeed so, then what physical processes could have led to this highly improbable result?

7.1. Testing deviation from RW geometries

The real universe is anisotropic and inhomogeneous on all scales except the largest observable scales. Hubble had already used number counts to provide evidence towards isotropy and homogeneity on large scales. However, what is striking is that every time we obtain detailed evidence on the distribution and motion of matter on larger scales, we find structure at those scales: voids, walls, large-scale motions due to great attractors (see, e.g., Rubin and Coyne 1988, Bothun 1998). The well developed correlation function measurements (Peebles 1980) are not sensitive to such structures, which are now being mapped in detail.

However, on the largest scales things seem to look homogeneous. We mean two rather different things by this. First, we do not see any specific signs of inhomogeneity or anisotropy of the matter distribution on these scales that would signal a non-RW geometry; we do not, for example, see a highly anisotropic Hubble flow, or greatly different number counts in one direction than another. The matter we see looks pretty much the same in each direction and we do not see any class of sources concentrated in one region in the sky. However, it should be noted here that some of the primary data we need to determine the geometry of the universe† is difficult to measure; specifically, we have only very poor limits on the transverse velocities of matter, so we could be missing some important signals of anisotropy. Nevertheless, when we examine what we can measure, we observe things to be isotropic about us, when averaged on a large enough scale; we do not see one particular preferred direction in the sky that is a candidate as the centre of the universe.

Secondly, we can make specific FL models that fit the data. However, as has already been remarked, we can only do so if we include evolution functions of the right form to allow this fit—taken at their face value, the number count data do not support spatial homogeneity.

† What those data are has been characterized in detail (see Kristian and Sachs 1966 and Ellis *et al* 1985).

Indeed, the step from the isotropy we observe about us to spatial homogeneity is not easy to justify on purely observational grounds—those data we take as showing evolution in time of sources we observe could just as easily, in fact, show spatial inhomogeneity in a spherically symmetric universe†. The best argument for spatial homogeneity is via a weak Copernican assumption plus the observed very low anisotropy of the CBR radiation. Specifically, if freely propagating CBR were exactly isotropic everywhere (i.e. at all times and places) in a spacetime region U since decoupling in an expanding universe, then the universe must have exactly a RW geometry in that region (Ehlers *et al* 1968); if it is approximately isotropic everywhere in such a region, then the universe is almost FL in that region (Stoeger *et al* 1995)‡. Thus if such almost-isotropy of the CBR in the observable region since decoupling is true, then we live in an almost-RW geometry. The Copernican assumption comes in because we cannot§ verify observationally that the CBR is as isotropic when measured in other parts of the observable universe and at earlier times, as it is here and now. Note that the deduction does not extend either to very early times (long before decoupling) or to very large distances (outside our visual horizon).

Because 'geological' observations test conditions way back in the past, a very desirable confirmation of spatial homogeneity would be via observing geological type data far out, i.e. at high redshift, hence combining past null cone and timelike observations (Ellis 1987a, 1995). The most promising development of this kind is the observation of helium abundances at high redshift. If those work out at the FL values, then that strongly suggests that the universe was like a FL model far out (at spatial distances corresponding to the observed redshift values) back to the time of nucleosynthesis‖.

7.2. Explaining RW geometry

The question, then, is are there physical processes that will lead to isotropization and homogenization of a generic initial cosmology? Misner's chaotic cosmology programme (Misner 1968, 1969) sought to show that this was so—specifically, that chaotic anisotropic in a Bianchi IX universe would remove horizons that restrict causal effectiveness of such processes, and that viscous effects would indeed isotropize the universe. This was developed in a very interesting examination of the dynamics of Bianchi (spatially homogeneous but anisotropic) cosmologies, which in the end showed that such processes could work to some degree but not totally—that is, they would isotropize a significant class of universe models but not generic ones (see, e.g., Stewart 1969).

In effect, this programme was picked up some decades later by the inflationary universe project (Guth 1981, Blau and Guth 1987), which extended physical cosmology to include the ideas of particle physics that were imported from solid-state physics, specifically broken symmetries and associated phase changes. Exciting progress was made, relating the evolution of the very early universe to particle physics (Kolb and Turner 1990), specifically showing that because of the existence of particle horizons, topological defects (monopoles, strings, domain

† See Ellis (1975), Ellis *et al* (1978) and Mustapha *et al* (1998).

‡ It is assumed here that the space and time derivatives of the CBR anisotropy are also small. This is physically highly plausible. As pointed out by Nilsson *et al* (1999), without such conditions, the Weyl tensor may not be highly restricted. The way such derivatives relate to CBR anisotropies locally is discussed by Maartens *et al* (1995a, b)

§ Except weakly via the Sunyaev–Zel'dovich (1970) effect, see Goodman (1995).

‖ This is in effect part of a larger programme to show that if physical conditions look similar at high redshift—the same mixture of galaxy types, for example—then the thermal history and so past early expansion rates there must have been the same as nearby, hence giving evidence of spatial homogeneity. This deduction, however, is not straightforward—there are counterexamples to the simplest version of this hypothesis (see Bonnor and Ellis 1986).

walls) could be expected as relics of the expansion of the very early universe†, and that under a wide variety of conditions, an effective scalar field could dominate the expansion of the early universe and lead to a period of exponential expansion (inflation) which would smooth out fluctuations in the universe (accounting for its large-scale similarity to a RW geometry) and simultaneously vastly extend the scale of the causal horizons, thus removing the horizon problem in relation to the isotropy of the CBR. Furthermore, quantum fluctuations in the very early universe would then provide seeds for a nearly scale-free spectrum of density fluctuations, thus explaining the growth of inhomogeneities from an almost homogeneous beginning‡. Some varieties of inflation (chaotic inflation) predict major inhomogeneities on super-horizon scales, with many expanding FL-like regions with different parameters and properties growing out of earlier expanding regions, like a multi-headed hydra (Linde 1987, 1990).

Overall this has been a very exciting proposal, extending the ideas of physical cosmology to the limits of particle physics, and showing a possible influence of the very small (microphysics) on the very large (cosmology) in a way that exemplifies the underlying physics project of unifying quite different areas by means of a single explanatory scheme. It has not yet fully succeeded for a number of reasons: particularly because first, although the theory is fully framed in accord with present-day particle physics ideas, the link is incomplete because there is no specific proposal for physical identification of the inflationary field (the 'inflaton'); no specific scalar field has been found in the laboratory that has the properties needed to give an inflationary universe with the desired early-universe behaviour§. Hence it is at present an 'in principle' proposal, developed in a great variety of speculative ways, rather than a development of the consequences of existence of an identified physical field.

Secondly, inflation is usually taken to predict a critical density universe today; but that does not seem to accord with current observations (see Coles and Ellis (1997) for a summary). To save inflation one either has to move to inflationary models with lower present-day matter densities‖ or introduce a cosmological constant. The latter introduces a new fine-tuning problem that is presently unresolved—why is this constant so close to zero and yet non-zero? (Weinberg 1989), but may be indicated independently by supernovae-based observations of the distance–redshift relation for distant galaxies. Either way, this evidence is awkward for standard inflationary theory.

Thirdly, almost all of the inflationary universe models discussed in the literature have a RW geometry. However, in that case there is no need for inflation to solve the horizon problem or to smooth out the universe, because spatial homogeneity and isotropy has then been assumed *a priori*. Such models give no evidence on whether inflation can succeed or not in homogenizing and isotropizing the universe. Inflation may succeed in solving the horizon problem in FL models, but can it do so in generic geometries? Further, it is not clear whether inflation can start in very anisotropic or inhomogeneous models, and whether it can, if it will indeed isotropize them effectively¶. Thus paradoxically the 'success' of inflation in explaining the homogeneity of the universe has mainly been considered in precisely those cases where it is not needed. The true test is whether it can succeed in more general models.

† Kibble (1976, 1980), Turok (1988), Kolb and Turner (1990) and Vilenkin and Shellard (1994).

‡ See Gibbons *et al* (1983), Barrow (1983), Kolb and Turner (1990) and Peacock (1999).

§ Indeed, papers in the area often treat the inflationary potential as arbitrarily adjustable to suit astrophysical need (see, e.g., Lidsey *et al* 1997). Of course, should such properties be identified from the cosmology side, and then laboratory experiments verify that a field exists with precisely the characteristics thus determined, this would be one of the great achievements of physics.

‖ See, e.g., Ellis *et al* (1991), Ratra and Peebles (1995) and Hawking and Turok (1998).

¶ See, e.g., Rothman and Ellis (1986), Penrose (1989a) and Raychaudhuri and Modak (1998).

7.3. Inhomogeneous and anisotropic models

In order to explore alternatives to RW geometries, one needs to develop a sound understanding of anisotropic and inhomogeneous cosmological models†. Given the complexity of the EFE, great progress has been made in this regard, based firstly on an increasing understanding of the role and nature of coordinates and symmetries in particular cosmological models‡. Secondly, on a covariant 1 + 3 decomposition applicable to general cosmologies, leading to important general relations such as the Raychaudhuri equation, vorticity conservation relations, and the relation of the matter flow to the Weyl tensor§. Thirdly, on development of tetrad methods (or equivalent 1-form methods) both for generic spacetimes‖ and in examining particular classes of exact solutions¶.

We obtain a series of parametrized alternative models (particularly, Bianchi and Kantowski–Sachs spatially homogeneous but anisotropic models; Lemaître–Tolman–Bondi spherically symmetric models; and Swiss-cheese 'cut and paste' inhomogeneous models) whose observational predictions for discrete sources, background radiation and nucleosynthesis can be determined and compared with observational data. This begins to give some quantitative meaning to the claim that the real universe is 'close to FL', as we can limit the anisotropy and inhomogeneity parameters in such models that are compatible with present-day observations+. However, as remarked in sections 4.5 and 5.4, these comparisons are plagued by the unknown evolutionary functions (see Mustapha *et al* 1998), allowing considerable freedom in fitting cosmological models to the observations. To resolve this needs tying down the source evolution histories by a combination of theory and observation. Nevertheless, the models are important in terms of the alternative behaviours they offer at early and late times.

7.4. Evolutionary histories: the space of spacetimes

We can also examine the dynamic evolution of these models, for example whether chaotic types of behaviour occur*, whether horizons occur (Thorne 1967, Misner 1969), and whether they isotropize or not. An important paper in this regard is that by Wald (1983), where he showed that a cosmological constant would tend to isotropize Bianchi models, making them like a de Sitter solution. However, he did not show that the matter flow velocity relative to the chosen surfaces of homogeneity would tend to zero (Raychaudhuri and Modak 1998), and in general this will not happen (see Goliath and Ellis 1999). Furthermore, the investigation needs to be extended to the full dynamic scalar field behaviour, which is often replaced by simplifying assumptions ('slow rolling' in particular).

To investigate such issues properly, one needs to look not at the evolution of individual models but of families of models represented in suitable phase spaces, which in turn are

† See, for example, MacCallum (1979) and Krasiński (1997).

‡ For example, Gödel's universe (see Gödel 1949, Hawking and Ellis 1973), and the Lemaître–Tolman–Bondi spherically symmetric models (see Bondi 1947), and their generalizations (see Krasiński 1997), including in particular the non-analytic Swiss-cheese models (Einstein and Strauss 1945a b, Schücking, 1954).

§ See Ehlers (1961), Hawking (1966), Ellis (1971a), Maartens (1997) and Ellis and van Elst (1999b).

‖ MacCallum (1973) and van Elst and Uggla (1997).

¶ See, for example, Ellis (1967), Ellis and MacCallum (1969), Misner (1968, 1969), Ryan and Shepley (1975) and Wainwright and Ellis (1997).

+ See Thorne (1967), Barrow (1976, 1984), Rothman and Matzner (1984), Matzner *et al* (1986) concerning nucleosynthesis; Collins and Hawking (1973a, b), Barrow *et al* (1983, 1985), Bajtlik *et al* (1986), Bunn *et al* (1996) concerning CBR anisotropies.

* Hobill *et al* (1994), Wainwright and Ellis (1997) and Cornish and Levin (1997).

representations of the space of cosmological spacetimes†. Using the theory of dynamical systems enables one to examine the evolutionary behaviour of generic models, and to relate this to specific high-symmetry (self-similar) models that act as attractors and saddle points in the phase spaces (Wainwright and Ellis 1997), and, in particular, one can examine when isotropization occurs and when it does not (Wainwright *et al* 1998, Goliath and Ellis 1999). Interesting results emerge: isotropization can take place even without inflation; but in many cases it is a temporary state, with highly anisotropic phases occurring before in the very early universe and after in the very late universe. This intermediate isotropization allows anisotropic models to be indistinguishably close to a FL model for a long time, yet to be quite different at very early and very late stages‡. If we believe that the real universe is in some sense generic, then we must expect these unstable anisotropic modes to be present as perturbation modes of the presently almost-FL universe.

Underlying such investigations of the evolution of generic models is an unsolved issue, related to the question of gravitational entropy mentioned in section 5.3. We have at present no fully satisfactory measure of the distance between two spacetime models, or of the generality of a model (although proposals to use function and parameter counting are promising), or of the probability of any particular model occurring in the space of all cosmologies. Without such a solid base, intuitive measures are often used (for example, in the discussion of inflation and associated probabilities of different behaviours occurring); the results obtained are dependent on the variables chosen, and could be misleading—one can change them by changing the variables used or the associated assumptions§. So if one wishes to talk about the probability of the universe or of specific cosmological models, as physicists wish to do, the proper foundation for those concepts is not yet in place‖.

Nevertheless, we have sufficient indication that the idea of isotropization holds in an interestingly diverse class of models even without inflation, and a much larger class of models with inflation. However, even in the latter case what can be achieved in terms of isotropization by physical processes is restricted: probably inflation will succeed in isotropizing the universe in a large class of cases, but not in many others¶. It may succeed in some regions but not others. Nevertheless, provided physical fields exist that do indeed cause inflation, those cases where it does succeed will soon dominate in terms of volume over the rest (precisely because they have inflated to a very large volume). Many suggestive arguments of this situation have been given, particularly by Linde (1987, 1990). It would be useful to have them formalized in terms of dynamical-systems-type arguments, with proper measures of probability on the phase space.

† The space of metrics has been investigated extensively by Fischer and Marsden (1979); however, they used a quotient space in which all copies of the same spacetime are identified, and this leads to nasty differential properties of this space-of-spacetimes. It may be suggested that it is better not to make that identification (Ellis and van Elst 1999b).

‡ See Collins and Hawking (1973a), Wainwright *et al* (1998) and Ellis and van Elst (1999b).

§ Just as important as the variables chosen is what is held constant (see, e.g., Ellis 1991a).

‖ The most sustained proposal is that of Gibbons *et al* (1987), who used a symplectic measure for FL models with a scalar field matter source; but the total measure diverges, and the ratios of the total measures of inflating and non-inflating models are ill-defined (Page 1987). The Wheeler–De Witt metric could be the foundation needed, but has not been implemented successfully in the generic context we have in mind here. Use of Bayesian methods suggests there is no 'flatness problem' as ordinarily understood, see Evrard and Coles (1995).

¶ See Penrose (1989a) and Rothman and Ellis (1986), for example.

8. The idea of a beginning of the universe

One of the oldest questions in cosmology is whether there was a beginning to the universe or not. As has been mentioned in section 3.3, the classical theory applied to FL models states that there must have been a singularity in the past, where the energy density diverges and spacetime structure breaks down, if (a) general relativity theory is the correct theory of gravity, and (b) matter satisfies the energy condition (6)—which ordinary matter does.

It must be emphasized that what happens here is very radical: it is not just that matter starts there at a state of infinite temperature and density; rather the laws of physics themselves begin then, as does spacetime itself. And that formulation is already very misleading: the old question, 'What happened before?' has no meaning as there was no before!—but the word 'begins' suggests there was a time before, when there was nothing. However, it must be emphasized that the initiation implied here was *ex nihilo*, where this 'nothing' is not a vacuum state at an earlier time—it is no spacetime at all.

The idea of a beginning to the universe has become familiar to us by repetition, and that removes some of the shocking nature of what is implied†. Once this is understood, it is tempting to try to avoid it by various means (Ellis 1984b). Indeed initially Friedmann, Eddington and Lemaître considered evolving models where a cosmological constant avoided the singularity; but as has been mentioned in section 3.3, that option is not open to us in realistic universe models today‡. Nevertheless, there has been a continual desire to avoid it in one way or another, and many such proposals are with us today.

8.1. Singularity avoidance by geometry?

Many have suggested the possibility that the singularities in FL models are a consequence of their high symmetry, and would go away if one adopted more realistic inhomogeneous and anisotropic models. Initial attempts to investigate this were plagued by the problem of distinguishing coordinate singularities from physical singularities. The introduction of covariant methods was essential in facing this issue.

A first step forward was provided by Raychaudhuri's derivation (1955) of his fundamentally important equation for generic dust spacetimes, extended by Ehlers (1961) to general fluids. This equation is the basic equation of gravitational attraction, and takes the general form

$$3\ddot{S}/S = -2\sigma^2 + 2\omega^2 + \dot{u}^a{}_{;a} - \tfrac{1}{2}(\mu + 3p) + \Lambda, \tag{8}$$

generalizing (5) to the case of anisotropic and inhomogeneous expansion, where $\dot{u}^a = u^a{}_{;b}u^b$ is the fluid acceleration, ω the magnitude of the vorticity, and σ the magnitude of the rate of distortion§. This shows immediately that for irrotational pressure-free matter, the same singularity theorem holds as before, irrespective of the degree of anisotropy or inhomogeneity in the spacetime‖. However, acceleration (due to pressure gradients) or vorticity could in principle upset this prediction. Examination of specific classes of models failed to find specific realistic cosmological models where the singularity was avoided¶, but analytic examination of

† However, emphasized particularly by John Wheeler, who called the existence of spacetime singularities the greatest crisis facing physics (see Misner, Thorne and Wheeler, p 1196).

‡ Particularly if we agree that the CBR is convincing evidence of a hot big bang—since at the time of decoupling, the matter density was 10^9 higher than today. A cosmological constant large enough to cause a turn around at that time would scarcely be ignorable today!

§ See Ehlers (1961), Ellis (1971a) and Ellis and van Elst (1999b).

‖ See Raychaudhuri (1955), Ehlers (1961) and Ellis (1971a).

¶ An apparent example is given by pressure-free Newtonian models that are shear-free, expanding, and rotating; but these have no general relativity counterparts (Ellis 1967, 1971a).

the full set of covariant or tetrad equations failed to provide a proof that all realistic (anisotropic and inhomogeneous) cosmologies are singular.

This situation was dramatically changed by Roger Penrose' pioneering work on black hole singularities (1965), giving a theorem predicting the existence of singularities in realistic gravitational collapse cases. This was extended to the case of cosmology by Stephen Hawking and others, proving a series of theorems culminating in the major Hawking–Penrose singularity theorem of 1970. The key elements were† (a) use of the timelike and null versions of the Raychaudhuri equation for families of irrotational geodesics with suitable energy conditions implying intersection of these geodesics after a finite distance or time, and (b) a very careful analysis of the causal properties of spacetime and the domains of dependence of initial data on spacelike surfaces (these domains being bounded by null horizons generated by null geodesics). Under very general circumstances characterizing both a black hole geometry and the situation arising after refocusing of our past lightcone in a realistic cosmology, and providing causal violations are avoided, these theorems showed the existence of an edge to spacetime, implied by the existence of incomplete geodesics. They did not, however, determine the nature of the singularity—that is, they did not necessarily imply that an infinite matter density would arise.

The further important point is that it was shown‡ that the existence of the CBR was by itself adequate proof of the refocusing of our past lightcone which is the central geometrical feature§ leading to the prediction of existence of a singularity (provided the energy conditions are satisfied). Hence in this way, general relativity implies the existence of an edge to spacetime associated with a singularity. Many examples show the variety of singularities that might exist at the beginning of the universe (they can be spacelike or timelike; velocity dominated or not; scalar singularities or non-scalar; chaotically oscillating in character; isotropic, cigarlike or pancakelike‖). However, singularity avoidance is only possible—given suitable causality and energy conditions—if matter is concentrated in such small isolated regions that reconvergence of our past lightcone is avoided; and that would imply either major CBR anisotropies that are not observed, or lack of enough matter to cause the observed black-body spectrum of that radiation.

8.2. Singularity avoidance by physics

It is clear that if the early universe is dominated by matter that does not obey the energy conditions, the singularity can be avoided; and indeed eternal inflation is possible for this reason (the scalar field postulated in this model does indeed violate these conditions). However, as pointed out earlier, this situation will only arise when quantum fields dominate; so the singularity theorems can be taken as predicting at a minimum that conditions in the early universe will become so severe that quantum fields will dominate. They might then succeed in avoiding a singularity; but then classical physics will have broken down at the macroscopic level.

The alternative is that the EFE might be wrong: other theories of gravity might hold that avoid the singularity, as, for example, in the steady-state universe and its variants (Hoyle 1948, Bondi 1960, Hoyle *et al* 1993, 1995) where in effect the energy conditions are violated and also energy conservation is no longer true. However, in a certain sense we know this situation will arise in the very early universe: in extreme enough situations not only will quantum fields

† Hawking and Penrose (1970). See Hawking and Ellis (1973) for a technical description of these methods and results, and Tipler *et al* (1980) for an overview set in historical context.

‡ Hawking and Ellis (1968, 1973) and Hawking and Penrose (1970).

§ Implying the existence of closed trapped surfaces, as in the black hole case.

‖ See Thorne (1967), Misner (1969), Lifshitz and Khalatnikov (1973), Belinskiĭ *et al* (1970), Eardley *et al* (1971) and Ellis and King (1974).

be important, but quantum gravity will dominate. General relativity will break down, and all sorts of new possibilities then arise.

8.3. The origin of the universe

The ultimate efforts to explain the beginning of the universe, and the particular initial conditions that have shaped its evolution, of necessity rest in some approach or other to applying quantum theory to the creation of the universe (Lemaître 1931a), and so inevitably in considering the implications of quantum gravity for cosmology. Many innovative attempts have been made here; as this paper focuses on general relativity and its application to cosmology, and it would be impossible to do justice to the various approaches to quantum cosmology without a very much longer paper, I will just make a few comments on the relation between these approaches and the various themes that have been outlined so far.

The attempt to develop a fully adequate quantum gravity approach to cosmology is of course hampered by the lack of a fully adequate theory of quantum gravity, as well as by the problems at the foundation of quantum theory (the measurement problem, collapse of the wavefunction, etc—see Isham (1997)), which can be ignored in many laboratory situations but have to be faced in the cosmological context. Given this context, the various attempts each develop in depth some specific aspect of quantum theory applied to the universe as a whole that may be expected to emerge from any successful theory of quantum gravity. In effect they either attempt (a) a true theory of creation *ex nihilo*, or (b) to describe a self-sustaining or self-referential universe which in some way bypasses the issue of creation, either by (b1) being or originating from an eternally existing state, for example, via the recurring idea of a phoenix universe (Dicke and Peebles 1979), the chaotic inflationary models of Linde, or the evolving universe idea of Smolin (1992)†, or (b2) by starting from a state with different properties of time than usual (or an emergent notion of time), as in the Hartle–Hawking no-boundary proposal, see Hawking (1987, 1993), and the Gott causal violation proposal‡. These may be combined with one or other proposals for (c) an effective ensemble of universes, realized either (c1) in time or space or in spacetime regions that are part of a larger spacetime but effectively disconnected from each other, or (c2) in truly disconnected form (Tegmark 1998).

From the viewpoint of this paper, these are all attempts to bring the physical cosmology approach to a final conclusion, by relating cosmology either to a 'theory of everything' (presumably some form of M-theory, see, e.g., Gibbons (1998)), or at least to the most unified view of physics we can successfully develop. We therefore run full tilt into the problem highlighted in section 5.4, namely the major difficulty—indeed probable impossibility—of testing the physics involved. This comes particularly to the fore in 'theories of initial conditions for the universe'—for here we are apparently proposing a theory with only one object§. Additionally most of these theories propose models of the universe that involve essentially untestable descriptions of spacetime geometry, either because they propose major models of conditions far outside our horizons which are therefore completely inaccessible to observation,

† One of the most intriguing projects in that it unites two of the great discoveries of modern science—Darwinian evolution and the evolving universe idea.

‡ See Gott and Li (1998) for a description of this proposal for the universe 'to create itself', together with a useful summary of the other approaches mentioned here.

§ Unless we seriously propose an infinite ensemble of completely disconnected universes—but that is completely untestable, despite hopeful remarks sometimes made, see, e.g., Tegmark (1998). The 'tests' proposed there are quite different than in the rest of physics and astronomy, and there is total lack of confirmation in any serious sense—anyone can claim any properties they like for such an ensemble, provided it includes at least one universe something like ours. If the different universe regions are in fact connected, as in chaotic inflation, we are dealing with one universe, as in the usual understanding of cosmology.

or of conditions in the very early universe which are inaccessible to testing, because all memory of that state has been wiped out by a period of inflation and later equilibrium processes.

Thus choices have to be made on general philosophical grounds† rather than in terms of testable physical theories. Whichever approach is adopted, there remain even in these intriguing approaches irremovable problems related to the creation and existence of a unique universe. The specific problem for models of creation *ex nihilo* is that they apparently rely on the existence of major physically effective structures (the apparatus of quantum field theory, for example) that are in some sense existent prior to or independent of the formation of the universe. The issue there is how does this physics come to apply—how can it precede the universe and the existence of space and time? Does it live in some Platonic super-space that is independent of the existence of spacetime? If the proposal is evolution from a previous eternal state (Minkowski space, for example) then why did that come into existence, and why did the universe expansion as a bubble from that vacuum start when it did, rather than at some previous time in the pre-existent eternity? Whenever it started, it should have started before!

One can try to avoid these problems by one or other of the self-referential or self-repeating universes, but they cannot overcome the ultimate existential question: why has one specific state occurred rather than any of the other possibilities?—why this self-referential scheme and not another one? This question cannot be solved by physics alone, unless one can show that only one form of physics is self-consistent; and the variety of proposals presently being made is evidence against that suggestion. We are considering here what might have been: trying to distinguish what is logically possible from the physically possible, comparing these with what actually occurred. It is difficult to make these distinctions in the context of cosmology, where the separation between physical laws and boundary conditions becomes obscured, so the usual separation of contingent events from necessary situations may also become unclear.

9. Relating models to the real universe

This survey sketches the developing relation of general relativity to cosmology from 1917 to 1999 in a broad-brush way. However, much of it has been taken up with issues such as the astrophysical development of objects, observational methods and limits, and tests of fundamental physical ideas, rather than detailed descriptions of topics in general relativity. I suggest that this is the proper perspective to use in looking at this relation; and indeed the necessity to take these subjects seriously is a mark of the great unification that has already been achieved, and the consequent move of cosmology and general relativity jointly from being fringe subjects only of esoteric interest (as in the 1920s to 1950s) to significant parts of both physics and astronomy today, the key turning point having been the discovery of the 3 K CBR in 1965 by Penzias and Wilson. This year may also be regarded as the year that theory moved to an altogether higher level of sophistication, marked by Penrose' 1965 paper on gravitational collapse.

Different approaches to cosmology, to some degree apparent in different historical stages of development of the subject (cf Ellis 1993), have focused on the various main ideas discussed in this paper. The ongoing tension in cosmological modelling is between *observation* and *explanation*—between a detailed description of what exists (examining its history and geography), and an analysis of its overall dynamics (a physical explanation of the origin of geometry and structure). To be useful, a model must cover both aspects‡, but in the end it must either be based on significant simplification (allowing an understandable causal model

† See Ellis (1991b) for a discussion of such issues.
‡ See Matravers *et al* (1995) for ways of tackling this tension.

of what is seen, based in universal physical processes), or be very detailed in its representation (thereby implicitly including many causal threads which are difficult to disentangle from each other) and containing a major contingent element—the specific details of what happens to be there.

Our models have reached a high degree of sophistication, particularly in the past decade, and are now playing a significant role in both the description and explanation of the cosmos. Spacetime curvature and its effects on matter and radiation are key features in this developing synthesis; the relation to astronomical observations and to high-energy physics brings the whole into fruitful interaction with cutting edge research in both areas, with general relativity a key element linking them. However, it is important to critically review the success of the models from time to time†, or else we may assume that our models are better representations of reality than they are. There remain important issues that are unresolved‡, in particular, related to the degree of realism of the models used, which may still be over-idealized in significant ways. These lead to themes that need careful attention in the coming decades. The following is one list of such themes.

(a) *Best-fit FL model with well determined astrophysical evolution.* The major present effort is in terms of determining observationally the parameters characterizing a best-fit FL model to the real universe, with a statistical description of the matter content related in a plausible way to a theory of structure formation. It is usually assumed there was an early inflationary phase, allowing physical prediction of the spectrum of inhomogeneities at late times if the inflationary potential is known. Apart from identification of the inflationary field and any dark matter components present, thus confirming the physical picture proposed, the key issue here (section 5.4) is the need to obtain a fit where the auxiliary astronomical parameters and variables (the bias parameter and source evolution functions, for example) are themselves determined in terms of plausible astrophysical models, rather than being arbitrary quantities one determines only via astronomical observations. Many issues arise here that have been widely discussed (see the references given above), such as determining the size and age of the universe from distant supernovae, tracing the distribution of dark matter in galaxies and clusters by gravitational lensing, using detailed surveys to map the distribution in depth and velocity fields of matter, and above all, using element abundances and CBR anisotropies to probe the early dynamical evolution of the universe, as discussed in the references above.

(b) *Appropriate handling of limits to verification.* As emphasized in section 6, there are limits to testability first due to visual and causal horizons, limiting our ability to reliably describe most of the real universe, and secondly due to the energy limits on Earth-based experiments, limiting our ability to apply well based physical theory to the evolution both of the universe itself and of structures in the universe. In particular, this uncertainty applies to the physics that may be related to the creation of the universe. There is a need for development of cosmological theory that takes these limits seriously, extending both theory and observational testing as far as they can sensibly go, but making quite clear the limits of what is well tested and what is not. A set of parametrized alternative models with clearly demarcated levels of probability for their various historical stages and physical elements would be desirable.

(c) *The inhomogeneities and anisotropies of the universe.* This parametrized set of models should include anisotropic and inhomogeneous models (section 7.3) so that we can properly place the universe in this broader context and determine limits on deviations

† See, e.g., Stoeger (1987), Gottlöber and Börner (1997) and Peebles (1998).
‡ See, e.g., Bahcall and Ostriker (1997), Turok (1997) and Münch *et al* (1997).

from a RW geometry. We need to push to the limit observational determination of quantities that characterize non-RW aspects of the geometry of the universe (section 7.1): i.e. inhomogeneities in the large-scale matter distribution, large-scale velocity fields, overall anisotropies in observations, systematic distortions of distant images and transverse velocities of matter at cosmological distances. Additionally, we should pursue all the consistency checks one can carry out on FL models: CBR temperature and element abundance measurements as a function of redshift, for example, as well as checking the predicted number-count dipole anisotropy that should accompany the CBR dipole (Ellis and Baldwin 1984) and the limits on CBR anisotropy at high z one may obtain from the Sunyaev–Zel'dovich (1970) effect (Goodman 1995).

(d) *Evolution trajectories of inhomogeneous cosmologies.* As regards understanding the way the universe got to be as it is, we need to characterize more clearly how the evolution of families of inhomogeneous models (section 7.4) relates to that of families of higher-symmetry models. It appears that a skeleton of higher-symmetry models may guide the evolution of lower-symmetry models in the space of spacetimes (see Wainwright and Ellis 1997); this needs further elucidation. Also the evolutionary paths in this state space of general (anisotropic and inhomogeneous) inflationary models are relatively little explored, but underlie plausibility arguments for inflation as an explanation of the present geometry and structure of the universe.

(e) *Probabilities of models and measures on the space of cosmologies.* We need to find a suitable measure of probability in the full space of cosmological spacetimes (see section 7.4). The requirement is a natural measure that is plausible and gives unique results for the relative probabilities of families of models with specific properties. One of the problems here is the relation between measures on 3-space geometries (as in the Hamiltonian approach) and spacetime geometries. Closely related to this theme is the issue of determining the stability or fragility of the results we derive from cosmological modelling†.

(f) *The averaging scales underlying cosmological models.* An important issue is examining the implications of the averaging scales assumed in any model used; since when one describes the same physical situation at different scales, both the resulting dynamics and observational properties may be expected to differ‡. The averaging scale assumed in a model is not often discussed, let alone the relation between representations at different scales, which bring in interesting observational and dynamical effects. In particular, averaging may be expected to result in effective (polarization) contributions to the energy–momentum tensor, arising because averaging does not commute with calculating the field equations for a given metric. Of importance is the issue (see section 4.2) of how the almost-everywhere empty real universe can have dynamical and observational relations that average out to high precision to the FLRW relations on a large scale. Conversely, we need more consideration of the way in which best-fitting of a background model to the real lumpy universe should be done (this underlies such issues as the modelling of velocities caused by inhomogeneities in the universe), and how this relates to averaging procedures (Ellis and Stoeger 1987).

(g) *Gravitational entropy and the arrow of time.* Related to this is the question (see section 5.3) of a definition of entropy for gravitating systems in cosmological models (Penrose 1989b, Hawking and Hunter 1999): we need a good generic definition, and a determination of its properties—or a proof that it does not exist. A successful approach may be expected to

† Tavakol and Ellis, (1988) and Rebouças *et al* (1998).

‡ See, e.g., Ellis (1984a), Zotov and Stoeger (1995), Futamase (1996) and Boersma (1998).

result from (or at least imply) a coarse-graining, and so is strongly related to averaging. It is an important issue in terms of its relation to the spontaneous formation of structure in the early universe, and also relates to the still unresolved but fundamentally important arrow of time problem—what relation does it have to cosmology?

(h) *The nature of physical laws in the expanding universe.* This in turn relates to the deep series of further issues concerned with the effect on local physics of boundary conditions at the beginning of the universe (Mach's principle, for example, and the arrow of time—see section 5.3), and how local physics may be determined to some extent by the structure of the universe. Can we somehow delimit the kinds of local physics that may be expected to underlie regularities in various conceivable cosmologies? Can we plausibly argue about how the nature of physical laws can arise as the universe comes into being? (See section 8.3.)

(i) *Appropriate handling of the uniqueness of the universe.* Underlying all these issues is the series of problems arising because of the uniqueness of the universe, which is what gives cosmology its particular character, underlying the special problems in cosmological modelling and the application of probability theory to cosmology (Ellis 1999b). Proposals to deal with this by considering an ensemble of universes realized in one way or another are in fact untestable and, hence, of a metaphysical rather than physical nature; but this needs further exploration. Can this be made plausible? Alternatively, how can the scientific method properly handle a theory which has only one unique object of application?

Apart from these strictly cosmological issues, what is still unresolved is the historical development of the many further levels of complexity after stars and planets form, and the relation of this development to gravitation. We have not considered here the evolution of truly complex structures—specifically, the origin of life—in the expanding universe. This is not presently considered a necessary or indeed legitimate part of most present-day cosmological investigation. It may be speculated, however, that this subject will come to be a major part of cosmology in the next millennium, when the basic parameters of the observable region of the universe have been determined, its major physical history is understood as well as possible given our experimental limitations, and the historico-geography of the observable region of the universe is being mapped in ever greater detail.

Gravitation is an important part of this story. It played a key role in enabling life—given cosmological initial conditions, it established a fruitful environment in which structures could grow, and then played a crucial role in development of the first round of self-organizing macro-structures (stars and galaxies). Present initial speculations related to the anthropic principle (see, e.g., Barrow and Tipler 1986) and the origin of life may be developed in the next millennium into a new degree of sophistication and integration with biology and evolutionary theory, taking the whole process of *unification* and *the studies of origins*—the two central themes emerging in this story—one stage further. The new focus will bring into play new domains of description and explanation (for example, relating to the nature of hierarchical structuring and the origin of self-organizing systems, see, e.g., Kauffman (1995)) as theory attempts to bring biological events too into a greater synthesis with cosmology and gravitational theory in a meaningful way.

Acknowledgments

I thank Henk van Elst, Gary Gibbons, Roy Maartens, Malcolm MacCallum, John Wainwright and David Matravers for helpful comments on a previous draft of this paper, and Henk van Elst for assistance with the references.

References

Anninos P 1998 Computational cosmology: from the early Universe to the large-scale structure *Max–Planck–Gesellschaft Living Reviews Series* no 1998–9 URL: `http://www.livingreviews.org/Articles/Volume1/1998-9anninos/`

Anninos P, Matzner R A, Rothman T and Ryan M P 1991 How does inflation isotropise the Universe? *Phys. Rev.* D **43** 3821

Audouze J and Tran Thanh Van J 1988 Dark matter *Proc. 23rd Roncontres de Moriond* (Gif-sur-Yvette: Editions Frontières)

Bahcall J N and Ostriker J P 1997 *Unsolved Problems in Astrophysics* (Princeton, NJ: Princeton University Press)

Bahcall J, Piran T and Weinberg S 1987 *Dark Matter in the Universe* (Singapore: World Scientific)

Bajtlik S, Juszkiewicz R, Prózszyński M and Amsterdamski P 1986 2.7 K radiation and the isotropy of the Universe *Astrophys. J.* **300** 463

Barbour J and Pfister H 1995 *Mach's Principle: from Newton's Bucket to Quantum Gravity* (Basel: Birkhäuser)

Bardeen J M 1980 Gauge-invariant cosmological perturbations *Phys. Rev.* D **22** 1882

——1988 Cosmological perturbations. From quantum fluctuations to large-scale structure *Cosmology and Particle Physics* ed Fang Li-Zhi and A Zee (New York: Gordon and Breach) p 1

Bardeen J M, Steinhardt P J and Turner M S 1983 Spontaneous creation of almost scale-free density perturbations in an inflationary universe *Phys. Rev.* D **28** 679

Barnett R M *et al* (Particle Data Group) 1996 Particle physics summary *Rev. Mod. Phys.* **68** 611

Barrow J D 1976 Light elements and the isotropy of the Universe *Mon. Not. R. Astron. Soc.* **175** 359

——1983 Cosmology, elementary particles and the regularity of the Universe *Relativistic Astrophysics and Cosmology* ed X Fustero and E Verdagauer (Singapore: World Scientific) p 137

——1984 Helium formation in cosmologies with anisotropic curvature *Mon. Not. R. Astron. Soc.* **211** 221

Barrow J D, Juszkiewicz R and Sonoda D H 1983 Structure of the cosmic microwave background *Nature* **305** 397

——1985 Universal rotation: how large can it be? *Mon. Not. R. Astron. Soc.* **213** 917

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

Belinskiĭ V A, Khalatnikov I M and Lifshitz E M 1970 Oscillatory approach to a singular point in the relativistic cosmology *Adv. Phys.* **19** 525

Bertotti B 1966 The luminosity of distant galaxies *Proc. R. Soc.* **294** 195

Bertschinger E 1992 Large scale structures and motions: linear theory and statistics *New Insights into the Universe* ed V J Matinez, M Portilla and D Saez (New York: Springer) p 62

Blandford R D and Narayan R 1992 Cosmological applications of gravitational lensing *Ann. Rev. Astron. Astrosci.* **30** 311

Blandford R D, Saust A, Brainerd T G and Villumsen J V 1991 The distortion of distant galaxy images by large-scale structure *Mon. Not. R. Astron. Soc.* **251** 600

Blau S K and Guth A H 1987 Inflationary cosmology *300 Years of Gravitation* ed S W Hawking and W Israel (Cambridge: Cambridge University Press) p 524

Boersma J P 1998 Averaging in cosmology *Phys. Rev.* D **57** 798

Bondi H 1947 Spherically symmetric models in general relativity *Mon. Not. R. Astron. Soc.* **107** 410

——1960 *Cosmology* (Cambridge: Cambridge University Press)

Bondi H and Gold T 1948 The steady state theory of the expanding Universe *Mon. Not. R. Astron. Soc.* **74** 36

Bonnor W B and Ellis G F R 1986 Observational homogeneity of the Universe *Mon. Not. R. Astron. Soc.* **218** 605

Bothun G 1998 *Modern Cosmological Observations and Problems* (London: Taylor and Francis)

Bruni M, Dunsby P K S and Ellis G F R 1992 Cosmological perturbations and the physical meaning of gauge-invariant variables *Astrophys. J.* **395** 34

Bunn E F, Ferreira P and Silk J 1996 How anisotropic is our Universe? *Phys. Rev. Lett.* **77** 2883

Burbidge E M, Burbidge G, Fowler W A and Hoyle F 1957 Synthesis of the elements in stars *Rev. Mod. Phys.* **29** 547

Challinor A D and Lasenby A N 1998 Cosmic microwave background anisotropies in the CDM model: a covariant and gauge-invariant approach *Preprint* astro-ph/9804301 (*Astrophys. J.* to be published)

Coles P and Ellis G F R 1997 *Is the Universe Open or Closed: the Density of Matter in the Universe* (Cambridge: Cambridge University Press)

Collins C B and Hawking S W 1973a The rotation and distortion of the Universe *Mon. Not. R. Astron. Soc.* **162** 307

——1973b Why is the Universe isotropic? *Astrophys. J.* **180** 317

Cornish N J and Levin J J 1997 The mixmaster universe is chaotic *Phys. Rev. Lett.* **78** 998

Cornish N J, Spergel D N and Starkman G D 1998 Circles in the sky: finding topology with the microwave background radiation *Class. Quantum Grav.* **15** 2657

Cowie L and Songaila A 1995 Astrophysical limits on the evolution of dimensionless physical constants over cosmological time *Astrophys. J.* **453** 596

Davies P C W 1974 *The Physics of Time Asymmetry* (London: Surrey University Press)
de Sitter 1917a On the relativity of inertia: remarks concerning Einstein's latest hypothesis *Koninklijke Nederlandsche Akademie van Wetenschappen (Amsterdam)* **19** 1217
——1917b On Einstein's theory of gravitation and its astronomical consequences *Mon. Not. R. Astron. Soc.* **78** 3
Dicke R H and Peebles P J E 1979 The big bang cosmology—enigmas and nostrums *General Relativity: an Einstein Centenary Survey* ed S W Hawking and W Israel (Cambridge: Cambridge University Press) p 504
Dicke R H, Peebles P J E, Roll P G and Wilkinson D T 1965 Cosmic blackbody radiation *Astrophys. J.* **142** 414
Dirac P A M 1938 New basis for cosmology *Proc. R. Soc.* **165** 199
Disney M J 1976 Visibility of galaxies *Nature* **263** 573
Dunsby P K S 1997 A fully covariant description of cosmic microwave background anisotropies *Class. Quantum Grav.* **14** 3391
Dyer C C and Roeder R C 1973 Distance–redshift relation for universes with some intergalactic medium *Astrophys. J.* **180** L31
——1981 On the transition from Weyl to Ricci focusing *Gen. Rel. Grav.* **13** 1157
Dyson F J 1971 Energy in the universe *Sci. Am.* Sept. 1971; also in *Energy and Power* (New York: Freeman) (A Scientific American book)
Eardley D, Liang E and Sachs R K 1971 Velocity dominated singularities in irrotational dust cosmologies *J. Math. Phys.* **13** 99
Eddington A S 1930 On the instability of Einstein's spherical world *Mon. Not. R. Astron. Soc.* **90** 668
Ehlers J 1961 Beiträge zur relativistischen Mechanik kontinuierlicher Medien *Akad. Wiss. Lit. Mainz, Abhandl. Math.-Nat. Kl.* **11** 793 (Engl. transl. Ehlers J 1993 Contributions to the relativistic mechanics of continuous media *Gen. Rel. Grav.* **25** 1225)
Ehlers J, Geren P and Sachs R K 1968 Isotropic solutions of the Einstein–Liouville equations *J. Math. Phys.* **9** 1344
Ehlers J, Prasanna A R and Breuer R A 1987 Propagation of gravitational waves through pressureless matter *Class. Quantum Grav.* **4** 253
Ehlers J and Rindler W 1989 A phase-space representation of Friedmann–Lemaître universes containing both dust and radiation and the inevitability of a big bang *Mon. Not. R. Astron. Soc.* **238** 503
Einstein A 1917 Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie *Sitz.-Ber. Preuß. Akad. Wiss., Berlin* 142. (Engl. transl. Einstein A 1923 Cosmological considerations on general relativity theory *The Principle of Relativity* ed H A Lorentz, A Einstein, H Minkowski and H Weyl (New York: Dover))
Einstein A and Strauss E G 1945a On the influence of the expansion of space on the gravitational field surrounding individual stars *Rev. Mod. Phys.* **17** 120
——1945b *Rev. Mod. Phys.* **18** 148
Ellis G F R 1967 Dynamics of pressure-free matter in general relativity *J. Math. Phys.* **8** 1171
——1971a Relativistic cosmology *General Relativity and Cosmology* (*Proc. 47th Enrico Fermi Summer School*) ed R K Sachs (New York: Academic) p 104
——1971b Topology and cosmology *Gen. Rel. Grav.* **2** 7
——1975 Cosmology and verifiability *Q. J. R. Astron. Soc.* **16** 245
——1980 Limits to verification in cosmology *Ann. NY Acad. Sci.* **336** 130
——1984a Relativistic cosmology: its nature, aims and problems *General Relativity and Gravitation* (*Invited Papers and Discussion Reports of the 10th Int. Conf.*) ed B Bertotti, F de Felice and A Pascolini (Dordrecht: Reidel) p 215
——1984b Alternatives to the big bang *Ann. Rev. Astron. Astrophys.* **22** 157
——1987a Observational cosmology after Kristian and Sachs *Theory and Observational Limits in Cosmology* ed W Stoeger (Rome: Vatican Observatory) p 43
——1987b Standard cosmology *5th Brazilian School on Cosmology and Gravitation* ed M Novello (Singapore: World Scientific) p 83
——1989 A history of cosmology 1917–1955 *Einstein and the History of General Relativity* (*Einstein Study Series* vol 1) ed D Howard and J Stachel (Boston, MA: Birkhäuser) p 367
——1990 Innovation resistance and change: the transition to the expanding universe *Modern Cosmology in Retrospect* ed B Bertotti, R Balbinto, S Bergia and A Messina (Cambridge: Cambridge University Press) p 97
——1991a Standard and inflationary cosmologies *Gravitation (Proc. Banff Summer Research Institute on Gravitation)* ed R Mann and P Wesson (Singapore: World Scientific) p 3
——1991b Major themes in the relation between philosophy and cosmology *Mem. Italian Astron. Soc.* **62** 553
——1993 The physics and geometry of the early Universe: changing viewpoints *Q. J. Roy. Astron. Soc.* **34** 315
——1995 Observations and cosmological models *Galaxies and the Young Universe* ed H von Hippelein, K Meisenheimer and J H Roser (Berlin: Springer) p 51
——1999a Relativistic cosmology 1999: issues and problems *Gen. Rel. Grav.* to appear

——1999b The unique nature of cosmology *Nature* to be published (*Stachel Festschrift* ed J Renn)
Ellis G F R and Baldwin J 1984 On the expected anisotropy of radio source counts *Mon. Not. R. Astron. Soc.* **206** 377
Ellis G F R and Bruni M 1989 Covariant and gauge-invariant approach to cosmological density fluctuations *Phys. Rev.* D **40** 1804
Ellis G F R and King A R 1974 Was the big bang a whimper? *Commun. Math. Phys.* **38** 119
Ellis G F R, Lyth D H and Mijić M B 1991 Inflationary models with Ω not equal to 1 *Phys. Lett.* B **271** 52
Ellis G F R, Maartens R and Nel S D 1978 The expansion of the universe *Mon. Not. R. Astron. Soc.* **184** 439
Ellis G F R and MacCallum M A H 1969 A class of homogeneous cosmological models *Commun. Math. Phys.* **12** 108
Ellis G F R and Madsen M 1991 Exact scalar field cosmologies *Class. Quantum Grav.* **8** 667
Ellis G F R, Nel S D, Stoeger W, Maartens R and Whitman A P 1985 Ideal observational cosmology *Phys. Rep.* **124** 315
Ellis G F R, Perry J J and Sievers A 1984 Cosmological observations of galaxies: the observational map *Astronom. J.* **89** 1124
Ellis G F R and Rothman T 1993 Lost horizons *Am. J. Phys.* **61** 93
Ellis G F R and Schreiber G 1986 Observational and dynamic properties of small universes *Phys. Lett.* A **115** 97
Ellis G F R and Sciama D W 1972 Global and non-global problems in cosmology *General Relativity (Synge Festschrift)* ed L O'Raifeartaigh (Oxford: Oxford University Press) p 35
Ellis G F R and Stoeger W R 1987 The fitting problem in cosmology *Class. Quantum Grav.* **4** 1679
——1988 Horizons in inflationary universes *Class. Quantum Grav.* **5** 207
Ellis G F R and van Elst H 1999a Deviation of geodesics in FLRW spacetime geometries *On Einstein's Path: Essays in Honour of Engelbert Schücking* ed A Harvey (New York: Springer) p 203
(——1997 *Preprint* gr-qc/9709060)
——1999b Cosmological models *Cargèse Lecture Notes 1998* ed M Lachièze-Rey to appear
(——1998 *Preprint* gr-qc/9812046)
Evrard G and Coles P 1995 Getting the measure of the flatness problem *Class. Quantum Grav.* **12** L93
Field G B 1969 Cosmic background radiation and its interaction with cosmic matter *Nuovo Cimento* **1** 87
Field G B, Arp H and Bahcall J N 1973 *The Redshift Controversy* (Reading, MA: Benjamin)
Fischer A E and Marsden E J 1979 The initial value problem and the dynamical formulation of general relativity *General Relativity: an Einstein Centenary Survey* ed S W Hawking and W Israel (Cambridge: Cambridge University Press) p 138
Fort B and Mellier Y 1994 Arc(let)s in clusters of galaxies *Astron. Astrophys. Rev.* **5** 239
Friedmann A 1922 Über die Krümmung des Raumes *Z. Phys.* **10** 377 (Engl. transl. Friedmann A 1999 On the curvature of space *Gen. Rel. Grav.* to appear)
——1924 Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes *Z. Phys.* **21** 326 (Engl. transl. Friedmann A 1999 On the possibility of a world with constant negative curvature of space *Gen. Rel. Grav.* to appear)
Futamase T 1996 Averaging of a locally inhomgeneous realistic universe *Phys. Rev.* D **53** 681
Gamow G 1946 Expanding universe and the origin of the elements *Phys. Rev.* **70** 572
Gibbons G W 1998 Quantum gravity/strings/M-theory as we approach the 3rd millenium *Gravitation and Relativity: at the Turn of the Millenium* ed N Dadhich and J Narlikar (Pune: IUCAA)
Gibbons G W, Hawking S W and Siklos S T C 1983 *The Very Early Universe* (Cambridge: Cambridge University Press)
Gibbons G W, Hawking S W and Stewart J M 1987 A natural measure on the set of all universes *Nucl. Phys.* B **281** 736
Gödel K 1949 An example of a new type of cosmological solution of Einstein's field equations of gravitation *Rev. Mod. Phys.* **21** 447
Goliath M and Ellis G F R 1999 Homogeneous cosmologies with cosmological constant *Phys. Rev.* D **60** 023502
(——1998 *Preprint* gr-qc/9811068)
Goodman J 1995 Geocentrism reexamined *Phys. Rev.* D **52** 1821
Gott J R III, Gunn J E, Schramm D N and Tinsley B M 1974 An unbound Universe? *Astrophys. J.* **194** 543
Gott J R III and Li L-X 1998 Can the Universe create itself? *Phys. Rev.* D **58** 023501
Gottlöber S and Börner G 1996 *The Evolution of the Universe* (New York: Wiley)
Gunn J E 1978 The Friedmann model and optical observations in cosmology *Observational Cosmology* ed J E Gunn, M S Longair and M J Rees (Saas-Fee: Geneva Observatory) p 1
Gunn J E and Peterson B A 1965 On the density of neutral hydrogen in intergalactic space *Astrophys. J.* **142** 1633
Guth A H 1981 Inflationary universe: a possible solution to the horizon and flatness problems *Phys. Rev.* D **23** 347
Harrison E R 1977 Radiation in homogeneous and isotropic models of the Universe *Vistas in Astronomy* **20** 341

——1990 *Cosmology: the Science of the Universe* (Cambridge: Cambridge University Press)
Harwitt M 1984 *Cosmic Discovery: the Search Scope and Heritage of Astronomy* (Boston, MA: MIT Press)
Hawking S W 1966 Perturbations of an expanding universe *Astrophys. J.* **145** 544
——1987 Quantum cosmology *300 Years of Gravitation* ed S W Hawking and W Israel (Cambridge: Cambridge University Press) p 631
——1993 *Hawking on the Big Bang and Black Holes (Advanced Series in Astrophysics and Cosmology)* (Singapore: World Scientific)
Hawking S W and Ellis G F R 1968 The cosmic black-body radiation and the existence of singularities in our Universe *Astrophys. J.* **152** 25
——1973 *The Large Scale Structure of Space-Time* (Cambridge: Cambridge University Press)
Hawking S W and Hunter C J 1999 Gravitational entropy and global structure *Phys. Rev.* D **59** 044025
Hawking S W and Penrose R 1970 The singularities of gravitational collapse and cosmology *Proc. R. Soc.* **314** 529
Hawking S W and Turok N 1998 Open inflation without false vacua *Phys. Lett.* B **425** 25
Heckmann O 1942 *Theorien der Kosmologie* (Berlin: Springer) (reprinted 1968)
Heckmann O and Schücking E 1956 Remarks on Newtonian cosmology *Z. Astrophys.* **40** 81
——1959 Relativistic and Newtonian cosmology *Handbuch der Physik* vol LIII (Berlin: Springer)
Hobill D W, Burd A and Coley A A (eds) 1994 *Deterministic Chaos in General Relativity* (New York: Plenum)
Hoyle F 1948 A new model of the expanding Universe *Mon. Not. R. Astron. Soc.* **108** 372
——1960 Cosmological tests of gravitational theories *Rendiconti Scuola Enrico Fermi. XX Corso* (New York: Academic Press) p 141
Hoyle F, Burbidge G and Narlikar J 1993 A quasi-steady state cosmological model with creation of matter *Astrophys. J.* **410** 437
——1995 The basic theory underlying the quasi-steady state cosmological model *Proc. R. Soc.* A **448** 191
Holz D E and Wald R M 1998 New method for determining cumulative gravitational lensing effects in inhomogeneous universes *Phys. Rev.* D **58** 063501
Hu W and Sugiyama N 1995a Anisotropies in the cosmic microwave background: an analytic approach *Astrophys. J.* **444** 489
——1995b Toward understanding CMB anisotropies and their implications *Phys. Rev.* D **51** 2599
Hu W and White M 1996 Acoustic signatures in the cosmic microwave background *Astrophys. J.* **471** 30
Hubble E 1929 A relation between distance and radial velocity among extragalactic nebulae *Proc. Natl Acad. Sci., USA* **15** 169
——1936 *The Realm of the Nebulae* (New Haven, CT: Yale University Press)
——1953 The law of redshifts *Mon. Not. R. Astron. Soc.* **113** 658
Isham C J 1997 *Lectures on Quantum Theory: Mathematical and Structural Foundations* (London: Imperial College Press, Singapore: World Scientific)
Jones A W and Lasenby A N 1998 The cosmic microwave background *Max–Planck–Gesellschaft Living Reviews Series* no 1998–11 URL: `http://www.livingreviews.org/Articles/Volume1/1998-11jones/`
Jungman G, Kamionkowski M, Kosowsky A and Spergel D N 1996 Cosmological-parameter determination with microwave background maps *Phys. Rev.* D **54** 1332
Kauffman S 1995 *At Home in the Universe: the Search for the Laws of Complexity* (London: Penguin)
Kibble T W B 1976 Topology of cosmic domains and strings *J. Phys. A: Math. Gen.* **9** 1387
——1980 Some implications of a cosmological phase transition *Phys. Rep.* **67** 183
Kolb E W and Turner M S 1990 *The Early Universe* (New York: Wiley)
Kramer D, Stephani H, MacCallum M A H and Herlt E 1980 *Exact Solutions of Einstein's Field Equations* (Berlin: VEB, Cambridge: Cambridge University Press)
Krasiński A 1996 *Inhomogeneous Cosmological Models* (Cambridge: Cambridge University Press)
Kristian J and Sachs R K 1966 Observations in cosmology *Astrophys. J.* **143** 379
Lemaître E 1927 A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae *Ann. Soc. Sci. Bruxelles I* A **47** 49 (Engl. transl. Lemaître E 1931 *Mon. Not. R. Astron. Soc.* **91** 483)
——1931a The beginning of the world from the viewpoint of quantum theory *Nature* **127** 706
——1931b The evolution of the universe: discussion *Nature* **128** 704
Lidsey J E, Liddle A R, Kolb E W, Copeland E J, Barriero T and Abney M 1997 Reconstructing the inflation potential—an overview *Rev. Mod. Phys.* **69** 373
Lifschitz E M 1946 On the gravitational instability of the expanding Universe *J. Phys. USSR* **10** 116 (Engl. transl. Lifschitz E M 1946 *Sov. Phys.–JETP* **16** 587)
Lifschitz E M and Khalatnikov I M 1963 Investigations in relativistic cosmology *Adv. Phys.* **12** 185

Linde A D 1987 Inflation and quantum cosmology *300 Years of Gravitation* ed S W Hawking and W Israel (Cambridge: Cambridge University Press) p 604
——1990 *Particle Physics and Inflationary Cosmology* (Chur: Harwood Academic)
Longair M S 1978 Radio astronomy and cosmology *Observational Cosmology* ed J E Gunn, M S Longair and M J Rees (Saas-Fee: Geneva Observatory) p 127
——1993 The physics of background radiation *The Deep Universe* (*Saas-Fee Advanced Course* vol 23) ed A Sandage, R G Kron and M S Longair (Berlin: Springer) p 317
Ma C-P and Bertschinger E 1995 Cosmological perturbation theory in the synchronous and conformal Newtonian gauges *Astrophys. J.* **455** 7
Maartens R 1997 Linearisation instability of gravity waves? *Phys. Rev.* D **55** 463
Maartens R, Ellis G F R and Stoeger W J 1995a Limits on anisotropy and inhomogeneity from the cosmic background radiation *Phys. Rev.* D **51** 1525
——1995b Improved limits on anisotropy and inhomogeneity from the cosmic background radiation *Phys. Rev.* D **51** 5942
Maartens R, Gebbie T and Ellis G F R 1999 Cosmic microwave background anisotropies: non-linear dynamics *Phys. Rev.* D **59** 083506
MacCallum M A H 1973 Cosmological models from a geometric point of view *Cargèse Lectures in Physics* vol 6, ed E Schatzman (New York: Gordon and Breach) p 61
——1979 Anisotropic and inhomogeneous relativistic cosmologies *General Relativity: an Einstein Centenary Survey* ed S W Hawking and W Israel (Cambridge: Cambridge University Press) p 533
Madsen M S and Ellis G F R 1988 The evolution of Ω in inflationary universes *Mon. Not. R. Astron. Soc.* **234** 67
Matravers D R, Ellis G F R and Stoeger W R 1995 Complementary approaches to cosmology: relating theory and observations *Q. J. R. Astron. Soc.* **36** 29
Matzner M, Rothman A and Ellis G F R 1986 Conjecture on isotope production in the Bianchi cosmologies *Phys. Rev.* D **34** 2926
McCrea W H and Milne E A 1934a Newtonian universes and the curvature of space *Q. J. Math.* **5** 73
——1934b *Q. J. Math.* **6** 81
Misner C W 1968 The isotropy of the Universe *Astrophys. J.* **151** 431
——1969 Mixmaster universe *Phys. Rev. Lett.* **22** 1071
Misner C W, Thorne K S and Wheeler J A 1973 *Gravitation* (New York: Freeman)
Mukhanov V F, Feldman H A and Brandenberger R H 1992 Theory of cosmological perturbations *Phys. Rep.* **215** 203
Münch G, Mampaso A and Sánchez F 1997 *The Universe at Large: Key Issues in Astronomy and Cosmology* (Cambridge: Cambridge University Press)
Mustapha N, Hellaby C and Ellis G F R 1998 Large scale inhomogeneity vs source evolution: can we distinguish them? *Mon. Not. R. Astron. Soc.* **292** 817
Nilsson U S, Uggla C, Wainwright J and Lim W C 1999 An almost isotropic cosmic microwave background temperature does not imply and almost isotropic universe *Preprint* astro-ph/9904252
North J D 1965 *The Measure of the Universe* (Oxford: Oxford University Press)
Ostriker J P and Steinhardt P J 1995 The observational case for a low-density Universe with a non-zero cosmological constant *Nature* **377** 600
Padmanabhan T 1993 *Structure Formation in the Universe* (Cambridge: Cambridge University Press)
Page D N 1987 Probability of R^2 inflation *Phys. Rev.* D **36** 1607
Panek M 1986 Large-scale microwave background fluctuations: gauge-invariant formalism *Phys. Rev.* D **34** 416
Partridge B 1995 *3 K: the Cosmic Microwave Background Radiation* (*Cambridge Astrophysics Series* no 25) (Cambridge: Cambridge University Press)
Partridge R B and Wilkinson D T 1967 Isotropy and homogeneity of the Universe from measurements of the cosmic microwave background *Phys. Rev. Lett.* **18** 557
Peacock J 1999 *Cosmological Physics* (Cambridge: Cambridge University Press)
Peebles P J E 1966 Primordial helium abundance and the primordial fireball *Astrophys. J.* **146** 542
——1971 *Physical Cosmology* (Princeton, NJ: Princeton University Press)
——1980 *The Large Scale Structure of the Universe* (Princeton, NJ: Princeton University Press)
——1993 *Principles of Physical Cosmology* (Princeton, NJ: Princeton University Press)
——1998 Is cosmology solved? An astrophysical cosmologist's viewpoint *Preprint* astro-ph/9810497
Peebles P J E, Schramm D N, Turner E L and Kron R G 1991 The case for the relativistic hot big bang cosmology *Nature* **352** 769
Penrose R 1963 Conformal treatment of infinity *Relativity, Groups and Topology* ed C M DeWitt and B S DeWitt (New York: Gordon and Breach) p 565
——1965 Gravitational collapse and space-time singularities *Phys. Rev. Lett.* **14** 57

——1979 Singularities and time-asymmetry *General Relativity: an Einstein Centenary Survey* ed S W Hawking and W Israel (Cambridge: Cambridge University Press) p 581
——1989a Difficulties with inflationary cosmology *Proc. 14th Texas Symp. on Relativistic Astrophysics* ed E J Fergus (New York: New York Academy of Sciences) p 249
——1989b *The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics* (Oxford: Oxford University Press)
Penzias A A and Wilson R W 1965 A measurement of excess antenna temperature at 4080 Mc/s *Astrophys. J.* **142** 419
Ratra B and Peebles P J E 1995 Inflation in open universes? *Phys. Rev.* D **52** 1837
Raychaudhuri A 1955 Relativistic cosmology *Phys. Rev.* **98** 1123
Raychaudhuri A K and Modak B 1988 Cosmological inflation with arbitrary initial conditions *Class. Quantum Grav.* **5** 225
Rebouças M J, Tavakol R K and Teixeira A F F 1998 Topology and fragility in cosmology *Gen. Rel. Grav.* **30** 535
Rees M J 1978 Growth and fate of inhomogeneities in a big bang cosmology *Observational Cosmology* ed J E Gunn, M S Longair and M J Rees (Saas-Fee: Geneva Observatory) p 261
——1987 The emergence of structure in the Universe: galaxy formation and dark matter *300 Years of Gravitation* ed S W Hawking and W Israel (Cambridge: Cambridge University Press) p 459
——1995 *Perspectives in Astrophysical Cosmology* (Cambridge: Cambridge University Press)
Reeves H 1987 Sources of information and free energy in an expanding universe *Unified View of the Macro and the Micro Cosmos: 1st Int. School on Astro-Particle Physics* ed A de Rujula, D V Nanopolous and P A Shaver (Singapore: World Scientific)
Rindler W 1956 Visual horizons in world models *Mon. Not. R. Astron. Soc.* **116** 662
Robertson H P 1929 Foundations of relativistic cosmology *Proc. Natl Acad. Sci., USA* **15** 822
——1933 Relativistic cosmology *Rev. Mod. Phys.* **5** 62
——1935 Kinematics and world structure *Astrophys. J.* **82** 248
Rothman A and Ellis G F R 1986 Can inflation occur in anisotropic cosmologies? *Phys. Lett.* B **180** 19
Rothman A and Matzner R 1984 Nucleosynthesis in anisotropic cosmologies revisited *Phys. Rev.* D **30** 1649
Roukema B F 1996 On determining the topology of the observable Universe via three-dimensional quasar positions *Mon. Not. R. Astron. Soc.* **283** 1147
Roukema B and Edge A C 1997 Constraining cosmological topology via highly luminous x-ray clusters *Mon. Not. R. Astron. Soc.* **292** 105
Rubin V C and Coyne G V 1988 *Large-Scale Motions in the Universe* (Princeton, NJ: Princeton University Press)
Ryan M and Shepley L 1975 *Homogeneous Relativistic Cosmologies* (Princeton, NJ: Princeton University Press)
Ryle M and Scheuer P A G 1955 The spatial distribution and the nature of radio stars *Proc. R. Soc.* **230** 448
Sachs R K and Wolfe A M 1967 Perturbations of a cosmological model and angular variations of the microwave background *Astrophys. J.* **147** 73
Sandage A R 1961 The ability of the 200-inch telescope to discriminate between selected world-models *Astrophys. J.* **133** 355
——1993 Practical cosmology: inventing the past *The Deep Universe* (*Saas-Fee Advanced Course* vol 23) ed A Sandage, R G Kron and M S Longair (Berlin: Springer) p 1
Sasaki M 1993 Cosmological gravitational lens equation—its validity and limitation *Prog. Theor. Phys.* **90** 753
Schneider P, Ehlers J and Falco E E 1992 *Gravitational Lenses* (Berlin: Springer)
Schramm D N and Turner M S 1998 Big-bang nucleosynthesis enters the precision era *Rev. Mod. Phys.* **70** 303
Schrödinger E 1956 *Expanding Universes* (Cambridge: Cambridge University Press)
Schücking E 1954 The Schwarzschild line element and the expansion of the Universe *Z. Phys.* **137** 595
Sciama D W 1971 Astrophysical cosmology *General Relativity and Cosmology* (*Proc. 47th Enrico Fermi Summer School*) ed R K Sachs (New York: Academic) p 183
Smolin L 1992 Did the Universe evolve? *Class. Quantum Grav.* **9** 173
Smoot G F *et al* 1992 Structure in the COBE differential microwave radiometer first year maps *Astrophys. J.* **396** L1
Stabell R and Refsdal S 1966 Classification of general relativistic world models *Mon. Not. R. Astron. Soc.* **132** 379
Steigmann G, Schramm D N and Gunn J E 1977 Cosmological limits to the number of massive leptons *Phys. Lett.* B **66** 202
Steinhardt P J 1995 Cosmology confronts the cosmic microwave background *Int. J. Mod. Phys.* A **10** 1091
Stewart J M 1969 Non-equilibrium processes in the early universe *Mon. Not. R. Astron. Soc.* **145** 347
Stewart J M and Walker M 1974 Perturbations of space-times in general relativity *Proc. R. Soc.* **341** 49
Stoeger W R (ed) 1987 *Theory and Observational Limits in Cosmology* (Rome: Vatican Observatory, Castel Gandolfo)
Stoeger W, Maartens R and Ellis G F R 1995 Proving almost-homogeneity of the Universe: an almost-Ehlers, Geren and Sachs theorem *Astrophys. J.* **443** 1

Sunyaev R A and Zel'dovich Ya B 1970 Small scale entropy and adiabatic density perturbations—antimatter in the universe *Astrophys. Space Sci.* **9** 368
Tavakol R K and Ellis G F R 1988 On the question of cosmological modelling *Phys. Lett.* A **130** 217
Tegmark M 1998 Is 'the theory of everything' merely the ultimate ensemble theory? *Ann. Phys., NY* **270** 1
Thorne K S 1967 Primordial element formation, primordial magnetic fields and the isotropy of the Universe *Astrophys. J.* **148** 51
Tipler F J, Clarke C J S and Ellis G F R 1980 Singularities and horizons: a review article *General Relativity and Gravitation: One Hundred Years after the Birth of Albert Einstein* vol 2, ed A Held (New York: Plenum) p 97
Tolman R C 1934 *Relativity, Thermodynamics, Cosmology* (Oxford: Clarendon)
Tolman R and Ward M 1932 On the behaviour of non-static models of the universe when the cosmological constant is omitted *Phys. Rev.* **39** 835
Treciokas R and Ellis G F R 1971 Isotropic solutions of the Einstein–Boltzmann equations *Commun. Math. Phys.* **23** 1
Turok N 1988 Cosmic strings *Cosmology and Particle Physics* ed Fang Li-Zhi and A Zee (New York: Gordon and Breach) p 207
——(ed) 1997 *Critical Dialogues in Cosmology* (Princeton, NJ: Princeton University Press)
van Elst H and Ellis G F R 1998 Quasi-Newtonian dust cosmologies *Class. Quantum Grav.* **15** 3545
——1999 Causal propagation of geometrical fields in relativistic cosmology *Phys. Rev.* D **59** 024013
van Elst H and Uggla C 1997 General relativistic 1 + 3 orthonormal frame approach *Class. Quantum Grav.* **14** 2673
Vilenkin A and Shellard P 1994 *Cosmic Strings and other Topological Defects* (Cambridge: Cambridge University Press)
Wagoner R V, Fowler W A and Hoyle F 1968 On the synthesis of elements at very high temperatures *Astrophys. J.* **148** 3
Wainright J, Coley A A, Ellis G F R and Hancock M 1998 On the isotropy of the Universe: do Bianchi VII_h cosmologies isotropize? *Class. Quantum Grav.* **15** 331
Wainwright J and Ellis G F R (eds) 1997 *Dynamical Systems in Cosmology* (Cambridge: Cambridge University Press)
Wald R M 1983 Asymptotic behavior of homogeneous cosmological models in the presence of a positive cosmological constant *Phys. Rev.* D **28** 2118
——1984 *General Relativity* (Chicago, IL: University of Chicago Press)
Walker A G 1936 On Milne's theory of world structure *Proc. Lond. Math. Soc.* **42** 90
——1944 Completely symmetric spaces *J. Lond. Math. Soc.* **19** 219
Wambsganss J 1998 Gravitational lensing in astronomy *Max–Planck–Gesellschaft Living Reviews Series* no 1998–12 URL: `http://www.livingreviews.org/Articles/Volume1/1998-12wamb/`
Weinberg S 1972 *Gravitation and Cosmology* (New York: Wiley)
——1977 *The First Three Minutes* (New York: Basic)
——1989 The cosmological constant problem *Rev. Mod. Phys.* **61** 1
Wheeler J A 1968 *Einstein's Vision* (Berlin: Springer)
Yoshimura M 1978 The universe as a laboratory for high energy physics *Cosmology and Particle Physics* ed Fang Li-Zhi and A Zee (New York: Gordon and Breach) p 293
Zeh H D 1992 *The Physical Basis of the Direction of Time* (Berlin: Springer)
Zel'dovich Ya B 1970 The Universe as a hot laboratory for the nuclear and particle physicist *Comment. Astrophys. Space Phys.* **2** 12
Zotov N and Stoeger W R 1995 Averaging Einstein's equations over a hierarchy of bound and unbound fragments *Astrophys. J.* **453** 574
Zwicky F 1937a Nebulae as gravitational lenses *Phys. Rev.* **51** 290
——1937b On the probability of detecting nebulae which act as gravitational lenses *Phys. Rev.* **51** 679