COSMOLOGICAL MODELS

Cargèse Lectures 1998

George F R Ellis*

and

Henk van Elst†

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

March 15, 1999

Abstract

The aim of this set of lectures is a systematic presentation of a $1+3$ covariant approach to studying the geometry, dynamics, and observational properties of relativistic cosmological models. In giving (i) the basic $1+3$ covariant relations for a cosmological fluid, the present lectures cover some of the same ground as a previous set of Cargèse lectures [7], but they then go on to give (ii) the full set of corresponding tetrad equations, (iii) a classification of cosmological models with exact symmetries, (iv) a brief discussion of some of the most useful exact models and their observational properties, and (v) an introduction to the gauge-invariant and 1+3 covariant perturbation theory of almost-Friedmann–Lemaître–Robertson–Walker universes, with a fluid description for the matter and a kinetic theory description of the radiation.

LANL e-print archives: gr-qc/9812046

*Electronic address: ellis@maths.uct.ac.za
†Electronic address: henk@gmunu.mth.uct.ac.za

Contents

1 Basic relations

A cosmological model represents the universe at a particular scale. We will assume that on large scales, space-time geometry is described by General Relativity (see e.g. d'Inverno [1], Wald [2], Hawking and Ellis [3], or Stephani [4]). Then a cosmological model is defined by specifying [5]–[7]:

* the *space-time geometry* represented on some specific averaging scale and determined by the metric $g_{ij}(x^k)$, which — because of the requirement of compatibility with observations — must either have some expanding Robertson–Walker ('RW') geometries as a regular limit (see [8]), or else be demonstrated to have observational properties compatible with the major features of current astronomical observations of the universe;
* the *matter present*, represented on the same averaging scale, and its *physical behaviour* (the energy-momentum tensor of each matter component, the equations governing the behaviour of each such component, and the interaction terms between them), which must represent physically plausible matter (ranging from early enough times to the present day, this will include most of the interactions described by present-day physics); and
* the *interaction of the geometry and matter* — how matter determines the geometry, which in turn determines the motion of the matter (see e.g. [9]). We assume this is through the *Einstein gravitational field equations* ('EFE') given by[1]

$$G_{ab} \equiv R_{ab} - \tfrac{1}{2}\,R\,g_{ab} = T_{ab} - \Lambda\,g_{ab} \ , \tag{1}$$

which, because of the *twice-contracted Bianchi identities*, guarantee the conservation of total energy-momentum

$$\nabla_b G^{ab} = 0 \qquad \Rightarrow \qquad \nabla_b T^{ab} = 0 \ , \tag{2}$$

provided the *cosmological constant* Λ satisfies the relation $\nabla_a\Lambda = 0$, i.e., it is constant in time and space.

Together, these determine the combined dynamical evolution of the model and the matter in it. The description must be sufficiently complete to determine

* the *observational relations* predicted by the model for both discrete sources and background radiation, implying a well-developed theory of *structure growth* for very small and for very large physical scales (i.e. for light atomic nuclei and for galaxies and clusters of galaxies), and of *radiation absorbtion and emission*.

To be useful in an explanatory role, a cosmological model must be easy to describe — that means they have symmetries or special properties of some kind or other. The usual choices for the matter description will be some combination of

* a fluid with a physically well-motivated equation of state, for example a perfect fluid with specified equation of state (beware of imperfect fluids, unless they have well-defined and motivated physical properties);
* a mixture of fluids, usually with different 4-velocities;
* a set of particles represented by a kinetic theory description;
* a scalar field ϕ, with a given potential $V(\phi)$ (at early times);
* an electromagnetic field described by the Maxwell field equations.

As intimated above, the observational relations implied by cosmological models must be compared with astronomical observations. This determines those solutions that can usefully be considered as

[1]Throughout this review we employ geometrised units characterised by $c = 1 = 8\pi G/c^2$. Consequently, all geometrical variables occurring have physical dimensions that are integer powers of the dimension [length]. The index convention is such that space-time and spatial indices with respect to a general basis are denoted by $a, b, \ldots = 0, 1, 2, 3$ and $\alpha, \beta, \ldots = 1, 2, 3$, respectively, while space-time indices in a coordinate basis are $i, j, \ldots = 0, 1, 2, 3$.

viable cosmological models of the real universe. A major aim of the present lectures is to point out that this class is wider than just the standard Friedmann–Lemaître–Robertson–Walker ('FLRW') cosmologies; indeed those models cannot be realistic on all scales of description, but they may also be inaccurate on large scales, or at very early and very late times. To examine this, we need to consider the subspace of the space of all cosmological solutions that contains models with observational properties like those of the real universe at some stage of their histories. Thus we are interested in the *full* state space of solutions, allowing us to see how realistic models are related to each other and to higher symmetry models, including particularly the FLRW models.

These lectures develop general techniques for examining this, and describe some specific models of interest. The first part looks at exact general relations valid in all cosmological models, the second part at exact cosmological solutions of the EFE, and the third part at approximate equations and solutions: specifically, 'almost-FLRW' models, linearised about a FLRW geometry.

2 $1+3$ covariant description

Space-times can be described via

(a) the metric $g_{ij}(x^k)$ described in a particular set of coordinates, with its differential properties, as embodied by the connection, given through the Christoffel symbols;

(b) the metric described by means of particular tetrads, with its connection given through the Ricci rotation coefficients;

(c) $1+3$ covariantly defined variables. In anisotropic cases, tetrad vectors can be uniquely defined in a $1+3$ covariant way and this approach merges into (b).

Here we will concentrate on the $1+3$ covariant approach, based on [5, 10, 6, 7, 11], but dealing also with the tetrad approach which serves as a completion to the $1+3$ covariant approach. The basic point here is that because we have complete coordinate freedom in General Relativity, it is preferable where possible to describe physics and geometry by tensor relations and quantities; these then remain valid whatever coordinate system is chosen.

2.1 Variables

2.1.1 Average 4-velocity of matter

In a cosmological space-time $(\mathcal{M}, \mathbf{g})$, at late times there will be a family of preferred worldlines representing the average motion of matter at each point[2] (notionally, these represent the histories of clusters of galaxies, with associated 'fundamental observers'); at early times there will be uniquely defined notions of the average velocity of matter (at that time, interacting gas and radiation), and corresponding preferred worldlines. In each case their *4-velocity* is[3]

$$u^a = \frac{dx^a}{d\tau} \; , \qquad u_a u^a = -1 \; , \tag{3}$$

where τ is proper time measured along the fundamental worldlines. We assume this 4-velocity is unique: that is, there is a well-defined preferred motion of matter at each space-time event. At recent times this is taken to be the 4-velocity defined by the vanishing of the dipole of the cosmic microwave background radiation ('CBR'): for there is precisely one 4-velocity which will set this dipole to zero. It is usually assumed that this is the same as the average 4-velocity of matter in

[2] We are here assuming a fluid description can be used on a large enough scale [5, 6]. The alternative is that the matter distribution is hierarchically structured at all levels or fractal (see e.g. [12] and refrences there), so that a fluid description does not apply. The success of the FLRW models encourages us to use the approach taken here.

[3] Merging from the one concept to the other as structure formation takes place.

a suitably sized volume [6]; indeed this assumption is what underlies studies of large scale motions and the 'Great Attractor'.

Given u^a, there are defined unique *projection tensors*

$$U^a{}_b = -\,u^a\,u_b \qquad \Rightarrow \qquad U^a{}_c\,U^c{}_b = U^a{}_b \ , \ U^a{}_a = 1 \ , \ U_{ab}\,u^b = u_a \ , \tag{4}$$

$$h_{ab} = g_{ab} + u_a\,u_b \qquad \Rightarrow \qquad h^a{}_c\,h^c{}_b = h^a{}_b \ , \ h^a{}_a = 3 \ , \ h_{ab}\,u^b = 0 \ . \tag{5}$$

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

$$\eta_{abc} = u^d\,\eta_{dabc} \qquad \Rightarrow \qquad \eta_{abc} = \eta_{[abc]} \ , \ \eta_{abc}\,u^c = 0 \ , \tag{6}$$

where η_{abcd} is the 4-dimensional volume element ($\eta_{abcd} = \eta_{[abcd]}$, $\eta_{0123} = \sqrt{|\det\, g_{ab}\,|}$).

Moreover, *two derivatives* are defined: the covariant time derivative $\dot{}$ along the fundamental worldlines, where for any tensor $T^{ab}{}_{cd}$

$$\dot{T}^{ab}{}_{cd} = u^e\nabla_e T^{ab}{}_{cd} \ , \tag{7}$$

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

$$\tilde{\nabla}_e T^{ab}{}_{cd} = h^a{}_f\,h^b{}_g\,h^p{}_c\,h^q{}_d\,h^r{}_e\nabla_r\,T^{fg}{}_{pq} \ , \tag{8}$$

with total projection on all free indices. The tilde serves as a reminder that if u^a has *non-zero* vorticity, $\tilde{\nabla}$ is *not* a proper 3-dimensional covariant derivative (see Eq. (27) below). Finally, following [11] (and see also [13]), we use angle brackets to denote orthogonal projections of vectors and the orthogonally projected symmetric trace-free part of tensors:

$$v^{\langle a\rangle} = h^a{}_b\,v^b \ , \qquad T^{\langle ab\rangle} = [\ h^{(a}{}_c\,h^{b)}{}_d - \tfrac{1}{3}\,h^{ab}\,h_{cd}\ ]\,T^{cd} \ ; \tag{9}$$

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

$$\dot{v}^{\langle a\rangle} = h^a{}_b\,\dot{v}^b \ , \qquad \dot{T}^{\langle ab\rangle} = [\ h^{(a}{}_c\,h^{b)}{}_d - \tfrac{1}{3}\,h^{ab}\,h_{cd}\ ]\,\dot{T}^{cd} \ . \tag{10}$$

Exercise: Show that the projected time and space derivatives of U_{ab}, h_{ab} and η_{abc} all vanish.

2.1.2 Kinematical quantities

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

$$\nabla_a u_b = -\,u_a\,\dot{u}_b + \tilde{\nabla}_a u_b = -\,u_a\,\dot{u}_b + \tfrac{1}{3}\,\Theta\,h_{ab} + \sigma_{ab} + \omega_{ab} \ , \tag{11}$$

where the trace $\Theta = \tilde{\nabla}_a u^a$ is the *(volume) rate of expansion* of the fluid (with $H = \Theta/3$ the Hubble parameter); $\sigma_{ab} = \tilde{\nabla}_{\langle a}u_{b\rangle}$ is the trace-free symmetric *rate of shear* tensor ($\sigma_{ab} = \sigma_{(ab)}$, $\sigma_{ab}\,u^b = 0$, $\sigma^a{}_a = 0$), describing the rate of distortion of the matter flow; and $\omega_{ab} = \tilde{\nabla}_{[a}u_{b]}$ is the skew-symmetric *vorticity* tensor ($\omega_{ab} = \omega_{[ab]}$, $\omega_{ab}\,u^b = 0$),[4] describing the rotation of the matter relative to a non-rotating (Fermi-propagated) frame. The stated meaning for these quantities follows from the evolution equation for a relative position vector $\eta^a_\perp = h^a{}_b\eta^b$, where η^a is a deviation vector for the family of fundamental worldlines, i.e. $u^b\nabla_b\eta^a = \eta^b\nabla_b u^a$. Writing $\eta^a_\perp = \delta\ell\,e^a$, $e_a e^a = 1$, we find the relative distance $\delta\ell$ obeys the propagation equation

$$\frac{(\delta\ell)^{\cdot}}{\delta\ell} = \tfrac{1}{3}\,\Theta + (\sigma_{ab}e^a e^b) \ , \tag{12}$$

[4] The vorticity here is defined with respect to a right-handedly oriented spatial basis.

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

$$\dot{e}^{\langle a\rangle} = (\sigma^a{}_b - (\sigma_{cd}e^c e^d)\, h^a{}_b - \omega^a{}_b)\, e^b \ , \tag{13}$$

giving the observed rate of change of position in the sky of distant galaxies. Finally $\dot{u}^a = u^b \nabla_b u^a$ is the *relativistic acceleration* vector, representing the degree to which the matter moves under forces other than gravity plus inertia (which cannot be covariantly separated from each other in General Relativity: they are different aspects of the same effect). The acceleration vanishes for matter in free fall (i.e. moving under gravity plus inertia alone).

2.1.3 Matter tensor

The matter *energy-momentum tensor* T_{ab} can be decomposed relative to u^a in the form[5]

$$\begin{gathered} T_{ab} = \mu\, u_a\, u_b + q_a\, u_b + u_a\, q_b + p\, h_{ab} + \pi_{ab} \ , \\ q_a\, u^a = 0 \ , \ \pi^a{}_a = 0 \ , \ \pi_{ab} = \pi_{(ab)} \ , \ \pi_{ab}\, u^b = 0 \ , \end{gathered} \tag{14}$$

where $\mu = (T_{ab}u^a u^b)$ is the *relativistic energy density* relative to u^a, $q^a = -\,T_{bc}\, u^b\, h^{ca}$ is the *relativistic momentum density*, which is also the energy flux relative to u^a, $p = \frac{1}{3}\,(T_{ab}h^{ab})$ is the *isotropic pressure*, and $\pi_{ab} = T_{cd}\, h^c{}_{\langle a}\, h^d{}_{b\rangle}$ is the trace-free *anisotropic pressure* (stress).

The *physics* of the situation is in the *equations of state* relating these quantities; for example, the commonly imposed restrictions

$$q^a = \pi_{ab} = 0 \quad \Leftrightarrow \quad T_{ab} = \mu\, u_a\, u_b + p\, h_{ab} \tag{15}$$

characterise a 'perfect fluid' with, in general, equation of state $p = p(\mu, s)$. If in addition we assume that $p = 0$, we have the simplest case: pressure-free matter ('dust' or 'Cold Dark Matter'). Otherwise we must specify an equation of state determining p from μ and possibly other thermodynamical variables. Whatever these relations may be, we usually require that various *energy conditions* hold: one or all of

$$\mu > 0 \ , \quad (\mu + p) > 0 \ , \quad (\mu + 3p) > 0 \ , \tag{16}$$

(the latter, however, being violated by scalar fields in inflationary universe models), and additionally demand the *isentropic speed of sound* $c_s^2 = (\partial p/\partial\mu)_{s=\mathrm{const}}$ obeys

$$0 \le c_s^2 \le 1 \quad \Leftrightarrow \quad 0 \le \left(\frac{\partial p}{\partial \mu}\right)_{s=\mathrm{const}} \le 1 \ , \tag{17}$$

as required for local stability of matter (lower bound) and causality (upper bound), respectively.

2.1.4 Maxwell field strength tensor

The *Maxwell field strength tensor* F_{ab} of an electromagnetic field is split relative to u^a into *electric* and *magnetic field* parts by the relations (see [7])

$$E_a = F_{ab}\, u^b \quad \Rightarrow \quad E_a u^a = 0 \ , \tag{18}$$

$$H_a = \tfrac{1}{2}\, \eta_{abc}\, F^{bc} \quad \Rightarrow \quad H_a u^a = 0 \ . \tag{19}$$

[5] We should really write $\mu = \mu(u^a)$, etc; but usually assume this dependence is understood.

2.1.5 Weyl curvature tensor

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

$$E_{ab} = C_{acbd}\, u^c\, u^d \quad \Rightarrow \quad E^a{}_a = 0\ ,\ E_{ab} = E_{(ab)}\ ,\ E_{ab}\, u^b = 0\ , \tag{20}$$

$$H_{ab} = \tfrac{1}{2}\, \eta_{ade}\, C^{de}{}_{bc}\, u^c \quad \Rightarrow \quad H^a{}_a = 0\ ,\ H_{ab} = H_{(ab)}\ ,\ H_{ab}\, u^b = 0\ . \tag{21}$$

These represent the 'free gravitational field', enabling gravitational action at a distance (tidal forces, gravitational waves), and influence the motion of matter and radiation through the *geodesic deviation equation* for timelike and null vectors, respectively [14]–[18]. Together with the Ricci tensor R_{ab} (determined locally at each point by the matter tensor through the EFE (1)), these quantities completely represent the space-time *Riemann curvature tensor* R_{abcd}, which in fully 1+3-decomposed form becomes[6]

$$\begin{aligned}
R^{ab}{}_{cd} &= R_P^{ab}{}_{cd} + R_I^{ab}{}_{cd} + R_E^{ab}{}_{cd} + R_H^{ab}{}_{cd}\ ,\\
R_P^{ab}{}_{cd} &= \tfrac{2}{3}\,(\mu + 3p - 2\Lambda)\, u^{[a}\, u_{[c}\, h^{b]}{}_{d]} + \tfrac{2}{3}\,(\mu + \Lambda)\, h^a{}_{[c}\, h^b{}_{d]}\ ,\\
R_I^{ab}{}_{cd} &= -\,2\, u^{[a}\, h^{b]}{}_{[c}\, q_{d]} - 2\, u_{[c}\, h^{[a}{}_{d]}\, q^{b]} - 2\, u^{[a}\, u_{[c}\, \pi^{b]}{}_{d]} + 2\, h^{[a}{}_{[c}\, \pi^{b]}{}_{d]}\ ,\\
R_E^{ab}{}_{cd} &= 4\, u^{[a}\, u_{[c}\, E^{b]}{}_{d]} + 4\, h^{[a}{}_{[c}\, E^{b]}{}_{d]}\ ,\\
R_H^{ab}{}_{cd} &= 2\, \eta^{abe}\, u_{[c}\, H_{d]e} + 2\, \eta_{cde}\, u^{[a}\, H^{b]e}\ .
\end{aligned} \tag{22}$$

2.1.6 Auxiliary quantities

It is useful to define some associated kinematical quantities: the *vorticity vector*

$$\omega^a = \tfrac{1}{2}\, \eta^{abc}\, \omega_{bc} \qquad \Rightarrow \qquad \omega_a\, u^a = 0\ ,\ \omega_{ab}\, \omega^b = 0\ , \tag{23}$$

the magnitudes

$$\omega^2 = \tfrac{1}{2}\, (\omega_{ab}\omega^{ab}) \geq 0\ ,\quad \sigma^2 = \tfrac{1}{2}\, (\sigma_{ab}\sigma^{ab}) \geq 0\ , \tag{24}$$

and the *average length scale* S determined by

$$\frac{\dot{S}}{S} = \tfrac{1}{3}\, \Theta\ , \tag{25}$$

so the volume of a fluid element varies as S^3. Further it is helpful to define particular *spatial gradients* orthogonal to u^a, characterising the inhomogeneity of space-time:

$$X_a \equiv \tilde{\nabla}_a \mu\ , \qquad Z_a \equiv \tilde{\nabla}_a \Theta\ . \tag{26}$$

The energy density μ (and also Θ) satisfies the important *commutation relation* for the $\tilde{\nabla}$-derivative [19]

$$\tilde{\nabla}_{[a} \tilde{\nabla}_{b]} \mu = \eta_{abc}\, \omega^c\, \dot{\mu}\ . \tag{27}$$

This shows that if $\omega^a\, \dot{\mu} \neq 0$ in an open set then $X_a \neq 0$ there, so non-zero vorticity implies anisotropic number counts in an expanding universe [20] (this is because there are then no 3-surfaces orthogonal to the fluid flow; see [6]).

[6]Here P is the perfect fluid part, I the imperfect fluid part, E that due to the electric Weyl tensor, and H that due to the magnetic Weyl tensor. This obscures the similarities in these equations between E and π, and between H and q; however, this partial symmetry is broken by the field equations, so the splitting given here (due to M Shedden) is conceptually useful.

2.2 $1+3$ covariant propagation and constraint equations

There are three sets of equations to be considered, resulting from the EFE (1) and its associated integrability conditions.

2.2.1 Ricci identities

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

$$2\,\nabla_{[a}\nabla_{b]}u^c = R_{ab}{}^c{}_d\,u^d \ . \tag{28}$$

On substituting in from (11), using (1), and separating out the orthogonally projected part into trace, symmetric trace-free, and skew symmetric parts, and the parallel part similarly, we obtain three propagation equations and three constraint equations. The *propagation equations* are,

1. The *Raychaudhuri equation* [21]

$$\dot{\Theta} - \tilde{\nabla}_a\dot{u}^a = -\,\tfrac{1}{3}\,\Theta^2 + (\dot{u}_a\dot{u}^a) - 2\,\sigma^2 + 2\,\omega^2 - \tfrac{1}{2}\,(\mu+3p) + \Lambda \ , \tag{29}$$

which is the *basic equation of gravitational attraction* [5]–[7], showing the repulsive nature of a positive cosmological constant, leading to identification of $(\mu+3p)$ as the active gravitational mass density, and underlying the basic singularity theorem (see below).

2. The *vorticity propagation equation*

$$\dot{\omega}^{\langle a\rangle} - \tfrac{1}{2}\,\eta^{abc}\,\tilde{\nabla}_b\dot{u}_c = -\,\tfrac{2}{3}\,\Theta\,\omega^a + \sigma^a{}_b\,\omega^b \ ; \tag{30}$$

together with (38) below, showing how vorticity conservation follows if there is a perfect fluid with acceleration potential Φ [5, 7], since then, on using (27), $\eta^{abc}\,\tilde{\nabla}_b\dot{u}_c = \eta^{abc}\,\tilde{\nabla}_b\tilde{\nabla}_c\Phi = 2\,\omega^a\,\dot{\Phi}$,

3. The *shear propagation equation*

$$\dot{\sigma}^{\langle ab\rangle} - \tilde{\nabla}^{\langle a}\dot{u}^{b\rangle} = -\,\tfrac{2}{3}\,\Theta\,\sigma^{ab} + \dot{u}^{\langle a}\,\dot{u}^{b\rangle} - \sigma^{\langle a}{}_c\,\sigma^{b\rangle c} - \omega^{\langle a}\,\omega^{b\rangle} - (E^{ab} - \tfrac{1}{2}\,\pi^{ab}) \ , \tag{31}$$

the anisotropic pressure source term π_{ab} vanishing for a perfect fluid; this shows how the tidal gravitational field E_{ab} directly induces shear (which then feeds into the Raychaudhuri and vorticity propagation equations, thereby changing the nature of the fluid flow).

The *constraint equations* are,

1. The (0α)-*equation*

$$0 = (C_1)^a = \tilde{\nabla}_b\sigma^{ab} - \tfrac{2}{3}\,\tilde{\nabla}^a\Theta + \eta^{abc}\,[\ \tilde{\nabla}_b\omega_c + 2\,\dot{u}_b\,\omega_c\ ] + q^a \ , \tag{32}$$

showing how the momentum flux (zero for a perfect fluid) relates to the spatial inhomogeneity of the expansion;

2. The *vorticity divergence identity*

$$0 = (C_2) = \tilde{\nabla}_a\omega^a - (\dot{u}_a\omega^a) \ ; \tag{33}$$

3. The H_{ab}-*equation*

$$0 = (C_3)^{ab} = H^{ab} + 2\,\dot{u}^{\langle a}\,\omega^{b\rangle} + \tilde{\nabla}^{\langle a}\omega^{b\rangle} - (\mathrm{curl}\,\sigma)^{ab} \ , \tag{34}$$

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

2.2.2 Twice-contracted Bianchi identities

The second set of equations arise from the *twice-contracted Bianchi identities* which, by the EFE (1), imply the conservation equations (2). Projecting parallel and orthogonal to u^a, we obtain the propagation equations

$$\dot{\mu} + \tilde{\nabla}_a q^a = -\,\Theta\,(\mu + p) - 2\,(\dot{u}_a q^a) - (\sigma^a{}_b \pi^b{}_a) \tag{35}$$

and

$$\dot{q}^{\langle a\rangle} + \tilde{\nabla}^a p + \tilde{\nabla}_b \pi^{ab} = -\,\tfrac{4}{3}\,\Theta\,q^a - \sigma^a{}_b\,q^b - (\mu + p)\,\dot{u}^a - \dot{u}_b\,\pi^{ab} - \eta^{abc}\,\omega_b\,q_c\ , \tag{36}$$

respectively. For perfect fluids, characterised by Eq. (15), these reduce to

$$\dot{\mu} = -\,\Theta\,(\mu + p)\ , \tag{37}$$

the *energy conservation equation*, and one constraint equation

$$0 = \tilde{\nabla}_a p + (\mu + p)\,\dot{u}_a\ , \tag{38}$$

the *momentum conservation equation*. This shows that $(\mu + p)$ is the inertial mass density, and also governs the conservation of energy. It is clear that if this quantity is zero (an effective cosmological constant) or negative, the behaviour of matter will be anomalous.

Exercise: Examine what happens in the two cases (i) $(\mu + p) = 0$, (ii) $(\mu + p) < 0$.

2.2.3 Other Bianchi identities

The third set of equations arise from the *Bianchi identities*

$$\nabla_{[a} R_{bc]de} = 0\ . \tag{39}$$

Double contraction gives Eq. (2), already considered. On using the splitting of R_{abcd} into R_{ab} and C_{abcd}, the above $1 + 3$ splitting of those quantities, and the EFE, the once-contracted Bianchi identities give two further propagation equations and two further constraint equations, which are similar in form to the Maxwell field equations in an expanding universe (see [22, 7]).

The *propagation equations* are,

$$\begin{aligned}(\dot{E}^{\langle ab\rangle} + \tfrac{1}{2}\,\dot{\pi}^{\langle ab\rangle}) - (\mathrm{curl}\,H)^{ab} + \tfrac{1}{2}\,\tilde{\nabla}^{\langle a} q^{b\rangle} \;&=\; -\,\tfrac{1}{2}\,(\mu + p)\,\sigma^{ab} - \Theta\,(E^{ab} + \tfrac{1}{6}\,\pi^{ab}) \\ &\quad + 3\,\sigma^{\langle a}{}_c\,(E^{b\rangle c} - \tfrac{1}{6}\,\pi^{b\rangle c}) - \dot{u}^{\langle a}\,q^{b\rangle} \\ &\quad + \eta^{cd\langle a}\,[\,2\,\dot{u}_c\,H^{b\rangle}{}_d + \omega_c\,(E^{b\rangle}{}_d + \tfrac{1}{2}\,\pi^{b\rangle}{}_d)\,]\ ,\end{aligned} \tag{40}$$

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

$$\begin{aligned}\dot{H}^{\langle ab\rangle} + (\mathrm{curl}\,E)^{ab} - \tfrac{1}{2}(\mathrm{curl}\,\pi)^{ab} \;&=\; -\,\Theta\,H^{ab} + 3\,\sigma^{\langle a}{}_c\,H^{b\rangle c} + \tfrac{3}{2}\,\omega^{\langle a}\,q^{b\rangle} \\ &\quad - \eta^{cd\langle a}\,[\,2\,\dot{u}_c\,E^{b\rangle}{}_d - \tfrac{1}{2}\,\sigma^{b\rangle}{}_c\,q_d - \omega_c\,H^{b\rangle}{}_d\,]\ ,\end{aligned} \tag{41}$$

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

$$(\mathrm{curl}\,H)^{ab} = \eta^{cd\langle a}\,\tilde{\nabla}_c H^{b\rangle}{}_d\ , \tag{42}$$

$$(\mathrm{curl}\,E)^{ab} = \eta^{cd\langle a}\,\tilde{\nabla}_c E^{b\rangle}{}_d\ , \tag{43}$$

$$(\mathrm{curl}\,\pi)^{ab} = \eta^{cd\langle a}\,\tilde{\nabla}_c \pi^{b\rangle}{}_d\ . \tag{44}$$

These equations show how gravitational radiation arises: taking the time derivative of the $\dot{E}$-equation gives a term of the form $(\mathrm{curl}\,H)^{\dot{}}$; commuting the derivatives and substituting from the $\dot{H}$-equation eliminates H, and results in a term in $\ddot{E}$ and a term of the form $(\mathrm{curl}\,\mathrm{curl}\,E)$, which together give the wave operator acting on E [22, 23]; similarly the time derivative of the $\dot{H}$-equation gives a wave equation for H.

The *constraint equations* are

$$\begin{aligned} 0 \;=\; (C_4)^a \;=\; & \tilde{\nabla}_b(E^{ab} + \tfrac{1}{2}\,\pi^{ab}) - \tfrac{1}{3}\,\tilde{\nabla}^a\mu + \tfrac{1}{3}\,\Theta\,q^a - \tfrac{1}{2}\,\sigma^a{}_b\,q^b - 3\,\omega_b\,H^{ab} \\ & - \eta^{abc}\,[\,\sigma_{bd}\,H^d{}_c - \tfrac{3}{2}\,\omega_b\,q_c\,]\ , \end{aligned} \tag{45}$$

the $(\operatorname{div} E)$-*equation* with source the spatial gradient of the energy density, which can be regarded as a vector analogue of the Newtonian Poisson equation [24], enabling tidal action at a distance, and

$$\begin{aligned} 0 \;=\; (C_5)^a \;=\; & \tilde{\nabla}_b H^{ab} + (\mu + p)\,\omega^a + 3\,\omega_b\,(E^{ab} - \tfrac{1}{6}\,\pi^{ab}) \\ & + \eta^{abc}\,[\,\tfrac{1}{2}\,\tilde{\nabla}_b q_c + \sigma_{bd}\,(E^d{}_c + \tfrac{1}{2}\,\pi^d{}_c)\,]\ , \end{aligned} \tag{46}$$

the $(\operatorname{div} H)$-*equation*, with source the fluid vorticity. These equations show respectively that scalar modes will result in a non-zero divergence of E_{ab} (and hence a non-zero E-field), and vector modes in a non-zero divergence of H_{ab} (and hence a non-zero H-field).

2.2.4 Maxwell field equations

Finally, we turn for completeness to the $1+3$ decomposition of the *Maxwell field equations*

$$\nabla_b F^{ab} = j^a_{\rm e}\ , \qquad\qquad \nabla_{[a}F_{bc]} = 0\ . \tag{47}$$

As shown in [7], the *propagation equations* can be written as

$$\dot{E}^{\langle a\rangle} - \eta^{abc}\,\tilde{\nabla}_b H_c \;=\; -\,j^{\langle a\rangle}_{\rm e} - \tfrac{2}{3}\,\Theta\,E^a + \sigma^a{}_b\,E^b + \eta^{abc}\,[\,\dot{u}_b\,H_c + \omega_b\,E_c\,]\ , \tag{48}$$

$$\dot{H}^{\langle a\rangle} + \eta^{abc}\,\tilde{\nabla}_b E_c \;=\; -\,\tfrac{2}{3}\,\Theta\,H^a + \sigma^a{}_b\,H^b - \eta^{abc}\,[\,\dot{u}_b\,E_c - \omega_b\,H_c\,]\ , \tag{49}$$

while the *constraint equations* assume the form

$$0 \;=\; (C_E) \;=\; \tilde{\nabla}_a E^a - 2\,(\omega_a H^a) - \rho_{\rm e}\ , \tag{50}$$

$$0 \;=\; (C_H) \;=\; \tilde{\nabla}_a H^a + 2\,(\omega_a E^a)\ , \tag{51}$$

where $\rho_{\rm e} = (-j_{{\rm e}\,a}u^a)$.

2.3 Pressure-free matter ('dust')

A particularly useful dynamical restriction is

$$p = 0 = q^a = \pi_{ab} \quad \Rightarrow \quad \dot{u}_a = 0\,, \tag{52}$$

so the matter (often described as 'baryonic') is represented only by its 4-velocity u^a and its energy density $\mu > 0$. The implication follows from momentum conservation: (38) shows that the matter moves geodesically (as expected from the equivalence principle). This is the case of *pure gravitation*: it separates out the (non-linear) gravitational effects from all the fluid dynamical effects. The vanishing of the acceleration greatly simplifies the above set of equations.

2.4 Irrotational flow

If we have a barotropic perfect fluid:

$$q^a = \pi_{ab} = 0\ , \qquad p = p(\mu)\ , \qquad \Rightarrow \qquad \eta^{abc}\,\tilde{\nabla}_b\dot{u}_c = 0\ , \tag{53}$$

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

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

follows from the vorticity conservation equation (30) (and is true also in the special case $p = 0$). When the vorticity vanishes:

1. The fluid flow is hypersurface-orthogonal, and there exists a cosmic time function t such that $u_a = -\, g(x^b)\, \nabla_a t$; if additionally the acceleration vanishes, we can set $g = 1$;

2. The metric of the orthogonal 3-spaces is h_{ab},

3. From the Gauss equation and the Ricci identities for u^a, the Ricci tensor of these 3-spaces is given by [5, 6]

$$^3R_{ab} = -\,\dot{\sigma}_{\langle ab \rangle} - \Theta\,\sigma_{ab} + \tilde{\nabla}_{\langle a}\dot{u}_{b\rangle} + \dot{u}_{\langle a}\,\dot{u}_{b\rangle} + \pi_{ab} + \tfrac{1}{3}\,h_{ab}\,[\,2\,\mu - \tfrac{2}{3}\,\Theta^2 + 2\,\sigma^2 + 2\,\Lambda\,]\,, \quad (54)$$

which relates $^3R_{ab}$ to E_{ab} via (31), and their Ricci scalar is given by

$$^3R = 2\,\mu - \tfrac{2}{3}\,\Theta^2 + 2\,\sigma^2 + 2\,\Lambda\,, \quad (55)$$

which is a generalised Friedmann equation, showing how the matter tensor determines the 3-space average curvature. These equations fully determine the curvature tensor $^3R_{abcd}$ of the orthogonal 3-spaces [6].

2.5 Implications

Altogether, in general we have six propagation equations and six constraint equations; considered as a set of evolution equations for the $1+3$ covariant variables, they are a first-order system of equations. This set is determinate once the fluid equations of state are given; together they then form a complete set of equations (the system closes up, but is essentially infinite dimensional because of the spatial derivatives that occur). The total set is normal hyperbolic at least in the case of a perfect fluid, although this is not obvious from the above form; it is shown by completing the equations to tetrad form (see the next section) and then taking combinations of the equations to give a symmetric hyperbolic normal form (see [25, 26]). We can determine many of the properties of specific solutions directly from these equations, once the nature of these solutions has been prescribed in $1+3$ covariant form (see for example the FLRW and Bianchi Type I cases considered below).

The *key issue* that arises is consistency of the constraints with the evolution equations. It is believed that they are *generally consistent* for physically reasonable and well-defined equations of state, i.e. they are consistent if no restrictions are placed on their evolution other than implied by the constraint equations and the equations of state (this has not been proved in general, but is very plausible; however, it has been shown for irrotational dust [11, 27]). It is this that makes consistent the overall hyperbolic nature of the equations with the 'instantaneous' action at a distance implicit in the Gauss-like equations (specifically, the $(\mathrm{div}\,E)$-equation), the point being that the 'action at a distance' nature of the solutions to these equations is built into the initial data, which ensures that the constraints are satisfied initially, and are conserved thereafter because the time evolution preserves these constraints (cf. [28]). A particular aspect of this is that when $\omega^a = 0$, the generalised Friedmann equation (55) is an integral of the Raychaudhuri equation (29) and energy equation (37).

One must be very cautious with imposing simplifying assumptions (e.g. vanishing shear) in order to obtain solutions: this can lead to major restrictions on the possible flows, and one can be badly misled if their consistency is not investigated carefully [29, 24]. Cases of particular interest are shear-free fluid motion (see [30]–[32]) and various restrictions on the Weyl tensor, including the 'silent universes', characterised by $H_{ab} = 0$ (and $p = 0$) [33, 34], or models with $\tilde{\nabla}_b H^{ab} = 0$ [35].

2.5.1 Energy equation

It is worth commenting here that, because of the equivalence principle, there is no agreed energy conservation equation for the gravitational field itself, nor is there a definition of its entropy (indeed some people — Freeman Dyson, for example [36] — claim it has no entropy). Thus the above set of equations does not contain expressions for gravitational energy[7] or entropy, and the concept of energy conservation does not play the major role for gravitation that it does in the rest of physics, neither is there any agreed view on the growth of entropy of the gravitational field.[8] However, energy conservation of the matter content of space-time, expressed by the divergence equation $\nabla_b T^{ab} = 0$, is of course of major importance.

If we assume a perfect fluid with a (linear) γ-law equation of state, then (37) shows that

$$p = (\gamma - 1)\,\mu \ , \ \dot{\gamma} = 0 \quad \Rightarrow \quad \mu = M/S^{3\gamma} \ , \ \dot{M} = 0 \ . \tag{56}$$

One can approximate ordinary fluids in this way with $1 \leq \gamma \leq 2$ in order that the causality and energy conditions are valid, with 'dust' and Cold Dark Matter ('CDM') corresponding to $\gamma = 1 \Rightarrow \mu = M/S^3$, and radiation to $\gamma = \frac{4}{3} \Rightarrow \mu = M/S^4$.

Exercise: Show how to generalise this to more realistic equations of state, taking account of entropy and of matter pressure (see e.g. [5]–[7]).

In the case of a mixture of non-interacting matter, radiation and CDM having the *same* 4-velocity, represented as a single perfect fluid, the total energy density is simply the sum of these components: $\mu = \mu_{\rm dust} + \mu_{\rm CDM} + \mu_{\rm radn}$. (NB: This is only possible in universes with spatially homogeneous radiation energy density, because the matter will move on geodesics which by the momentum conservation equation implies $\tilde{\nabla}_a p_{\rm radn} = 0 \Leftrightarrow \tilde{\nabla}_a \mu_{\rm radn} = 0$. This will not be true for a general inhomogeneous or perturbed FLRW model, but will be true in exact FLRW and orthogonal Bianchi models.)

Exercise: The pressure can still be related to the energy density by a γ-law as in (56) in this case of non-interacting matter and radiation, but γ will no longer be constant. What is the equation giving the variation of (i) γ, (ii) the speed of sound, with respect to the scale factor in this case? (See [38].)

A scalar field has a perfect fluid energy-momentum tensor if the surfaces $\{\phi = \text{const}\}$ are spacelike and we choose u^a normal to these surfaces. Then it approximates the equation of state (56) in the 'slow-rolling' regime, with $\gamma \approx 0$, and in the velocity-dominated regime, with $\gamma \approx 2$. In the former case the energy conditions are no longer valid, so 'inflationary' behaviour is possible, which changes the nature of the attractors in the space of space-times in an important way.

Exercise: Derive expressions for μ, p, $(\mu + p)$, $(\mu + 3p)$ in this case. Under what conditions can a scalar field have (a) $(\mu + p) = 0$, (b) $(\mu + 3p) = 0$, (c) $(\mu + 3p) < 0$?

2.5.2 Basic singularity theorem

Using the definition (25) of S, the Raychaudhuri equation can be rewritten in the form (cf. [21])

$$3\,\frac{\ddot{S}}{S} = -\,2\,(\sigma^2 - \omega^2) + \tilde{\nabla}_a \dot{u}^a + (\dot{u}_a \dot{u}^a) - \tfrac{1}{2}\,(\mu + 3p) + \Lambda \ , \tag{57}$$

[7] There are some proposed 'super-energy' tensors, for example the Bel–Robinson tensor [37], but they do not play a significant role in the theory.

[8] Entropy is well understood in the case of black holes, but not for gravitational fields in the expanding universe.

showing how the curvature of the curve $S(\tau)$ along each worldline (in terms of proper time τ along that worldline) is determined by the kinematical quantities, the total energy density and pressure[9] in the combination $(\mu + 3p)$, and the cosmological constant Λ. This gives the basic

> **Singularity Theorem** [21, 5, 6, 7]: In a universe where $(\mu + 3p) > 0$, $\Lambda \leq 0$, and $\dot{u}^a = \omega^a = 0$ at all times, at any instant when $H_0 = \frac{1}{3}\Theta_0 > 0$, there must have been a time $t_0 < 1/H_0$ ago such that $S \to 0$ as $t \to t_0$; a space-time singularity occurs there, where $\mu \to \infty$ and $p \to \infty$ for ordinary matter (with $(\mu + p) > 0$).

The further singularity theorems of Hawking and Penrose (see [39, 3, 40]) utilize this result (and its null version) as an essential part of their proofs.

Closely related to this are two other results: the statements that (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). Proofs are left to the reader; they follow directly from (57).

2.5.3 Relations between important parameters

Given the definitions

$$H_0 = \frac{\dot{S}_0}{S_0}\,, \quad q_0 = -\,\frac{1}{H_0^2}\frac{\ddot{S}_0}{S_0}\,, \quad \Omega_0 = \frac{\mu_0}{3H_0^2}\,, \quad w_0 = \frac{p_0}{\mu_0}\,, \quad \Omega_\Lambda = \frac{\Lambda}{3H_0^2}\,, \tag{58}$$

for the present-day values of the Hubble parameter ('constant'), deceleration parameter, density parameter, pressure to density ratio, and cosmological constant parameter, respectively, then from (57) we obtain

$$q_0 = 2\,\frac{(\sigma^2 - \omega^2)_0}{H_0^2} - \frac{(\tilde{\nabla}_a \dot{u}^a)_0 + (\dot{u}_a \dot{u}^a)_0}{3H_0^2} + \tfrac{1}{2}\,\Omega_0\,(1 + 3w_0) - \Omega_\Lambda\,. \tag{59}$$

Now CBR anisotropies let us deduce that the first two terms on the right are very small today, certainly less than 10^{-3}, and we can reasonably estimate from the nature of the matter that $p_0 \ll \mu_0$ and the third term on the right is also very small, so we estimate that in realistic universe models, at the present time

$$q_0 \approx \tfrac{1}{2}\,\Omega_0 - \Omega_\Lambda\,. \tag{60}$$

(Note we can estimate the magnitudes of the terms which have been neglected in this approximation.) This shows that a cosmological constant can cause an acceleration (negative q_0); if it vanishes, as commonly assumed, the expression simplifies:

$$\Lambda = 0 \quad\Rightarrow\quad q_0 \approx \tfrac{1}{2}\,\Omega_0\,, \tag{61}$$

expressing how the matter density present causes a deceleration of the universe. If we assume no rotation ($\omega^a = 0$), then from (55) we can estimate

$${}^3\!R_0 \approx 6\,H_0^2\,(\,\Omega_0 - 1 + \Omega_\Lambda\,)\,, \tag{62}$$

where we have dropped a term $(\sigma_0/H_0)^2$. If $\Lambda = 0$, then ${}^3\!R_0 \approx 6\,H_0^2\,(\,\Omega_0 - 1\,)$, showing that $\Omega_0 = 1$ is the critical value separating irrotational universes with positive spatial curvature ($\Omega_0 > 1 \Rightarrow {}^3\!R_0 > 0$) from those with negative spatial curvature ($\Omega_0 < 1 \Rightarrow {}^3\!R_0 < 0$).

Present day values of these parameters are almost certainly in the ranges [41]: baryon density: $0.01 \leq \Omega_0^{\rm baryons} \leq 0.03$, total matter density: $0.1 \leq \Omega_0 \leq 0.3$ to 1 (implying that much matter may not be baryonic), Hubble constant: $45\,{\rm km/sec/Mpc} \leq H_0 \leq 80\,{\rm km/sec/Mpc}$, deceleration parameter: $-\,0.5 \leq q_0 \leq 0.5$, cosmological constant: $0 \leq \Omega_\Lambda \leq 1$.

[9] This form of the equation is valid for imperfect fluids also: the quantities q^a and π_{ab} do not directly enter this equation.

2.6 Newtonian case

Newtonian equations can be developed completely in parallel [42, 43, 6] and are very similar, but simpler; for example the Newtonian version of the Raychaudhuri equation is

$$\dot{\Theta} + \tfrac{1}{3}\,\Theta^2 + 2\,(\sigma^2 - \omega^2) - \mathrm{D}_\alpha a^\alpha + \tfrac{1}{2}\,\rho - \Lambda = 0\ , \tag{63}$$

where ρ is the matter density and $a_\alpha = \dot{v}_\alpha + \mathrm{D}_\alpha\Phi$ is the Newtonian analogue of the relativistic 'acceleration vector', with $\dot{}$ the convective derivative and Φ the Newtonian potential (with suitably generalised boundary conditions [44, 45]). The Newtonian analogue of E_{ab} is

$$E_{\alpha\beta} = \mathrm{D}_\alpha\mathrm{D}_\beta\Phi - \tfrac{1}{3}\,(\mathrm{D}_\gamma\mathrm{D}^\gamma\Phi)\,h_{\alpha\beta}\ , \tag{64}$$

where $h_{\alpha\beta}$ denotes the metric, and D_α the covariant derivative, of Euclidean space. For the latter $\mathrm{D}_\alpha h_{\beta\gamma} = 0$ and $[\,\mathrm{D}_\alpha, \mathrm{D}_\beta\,] = 0$. There is no analogue of H_{ab} in Newtonian theory [6], as shown by a strict limit process leading from relativistic to Newtonian solutions [46].

Exercise: Under what conditions will a relativistic cosmological solution allow a representation (64) for the electric part of the Weyl tensor? Will the potential Φ occurring here necessarily also relate to the acceleration of the reference timelike worldlines?

2.7 Solutions

Useful solutions are defined by considering appropriate restrictions on the kinematical quantities, Weyl tensor, or space-time geometry, for a specified plausible matter content. Given such restrictions,

(a) we need to understand the *dynamical evolution* that results, particularly fixed points, attractors, etc., in terms of suitable variables,

(b) we particularly seek to determine and characterise *involutive subsets* of the space of space-times: these are subspaces mapped into themselves by the dynamical evolution of the system, and so are left invariant by that evolution. The constraint and evolution equations must be consistent with each other on such subsets. A characterisation of these subspaces goes a long way to characterising the nature of self-consistent solutions of the full non-linear EFE.

As far as possible we aim to do this for the exact equations. We are also concerned with

(c) *linearisation* of the equations about known simple solutions, and determination of properties of the resulting linearised solutions, in particular considering whether they accurately represent the behaviour of the full non-linear theory in a neighbourhood of the background solution (the issue of *linearisation stability*),

(d) derivation of the *Newtonian limit* and its properties from the General Relativity equations, and understanding how accurately this represents the properties of the full relativistic equations (and of its linearised solutions); see [24] for a discussion.

3 Tetrad description

The 1+3 covariant equations are immediately transparent in terms of representing relations between $1 + 3$ covariantly defined quantities with clear geometrical and/or physical significance. However, they do *not* form a complete set of equations guaranteeing the existence of a corresponding metric and connection. For that we need to use a tetrad description. The equations determined will then form a complete set, which will contain as a subset all the $1 + 3$ covariant equations just derived (albeit presented in a slightly different form). For completeness we will give these equations for a general dissipative relativistic fluid (recent presentations, giving the following form of the equations, are [47, 26]). First we briefly summarize a generic tetrad formalism, and then its application to cosmological models (cf. [30, 48]).

3.1 General tetrad formalism

A tetrad is a set of four orthogonal unit basis vector fields $\{\, \mathbf{e}_a \,\}$, $a = 0, 1, 2, 3$, which can be written in terms of a local coordinate basis by means of the *tetrad components* ${e_a}^i(x^j)$:

$$\mathbf{e}_a = {e_a}^i(x^j)\,\frac{\partial}{\partial x^i} \quad \Leftrightarrow \quad \mathbf{e}_a(f) = {e_a}^i(x^j)\,\frac{\partial f}{\partial x^i}\,, \quad {e_a}^i = \mathbf{e}_a(x^i)\ , \tag{65}$$

(the latter stating that the i-th component of the a-th tetrad vector is just the directional derivative of the i-th coordinate in the direction $\mathbf{e}_a$). This can be thought of as just a general change of vector basis, leading to a change of tensor components of the standard tensorial form: ${T^{ab}}_{cd} = {e^a}_i\,{e^b}_j\,{e_c}^k\,{e_d}^l\,{T^{ij}}_{kl}$ with obvious inverse, where the inverse components ${e^a}_i(x^j)$ (note the placing of the indices!) are defined by

$${e_a}^i\,{e^a}_j = {\delta^i}_j \quad \Leftrightarrow \quad {e_a}^i\,{e^b}_i = {\delta^b}_a\ . \tag{66}$$

However, it is a change from an integrable basis to a non-integrable one, so non-tensorial relations (specifically: the form of the metric and connection components) are a bit different than when coordinate bases are used. A change of one tetrad basis to another will also lead to transformations of the standard tensor form for all tensorial quantities: if $\mathbf{e}_a = {\Lambda_a}^{a'}(x^i)\,\mathbf{e}_{a'}$ is a change of tetrad basis with inverse $\mathbf{e}_{a'} = {\Lambda_{a'}}^a(x^i)\,\mathbf{e}_a$ (each of these matrices representing a Lorentz transformation), then ${T^{ab}}_{cd} = {\Lambda_{a'}}^a\,{\Lambda_{b'}}^b\,{\Lambda_c}^{c'}\,{\Lambda_d}^{d'}\,{T^{a'b'}}_{c'd'}$. Again the inverse is obvious.[10]

The metric tensor components in the tetrad form are given by

$$g_{ab} = g_{ij}\,{e_a}^i\,{e_b}^j = \mathbf{e}_a\cdot\,\mathbf{e}_b = \eta_{ab}\ , \tag{67}$$

where $\eta_{ab} = \mathrm{diag}\,(\,-1,\,+1,\,+1,\,+1\,)$, showing that the basis vectors are unit vectors orthogonal to each other (because the components g_{ab} are just the scalar products of these vectors with each other). The inverse equation

$$g_{ij}(x^k) = \eta_{ab}\,{e^a}_i(x^k)\,{e^b}_j(x^k) \tag{68}$$

explicitly constructs the coordinate components of the metric from the (inverse) tetrad components ${e^a}_i(x^j)$. We can raise and lower tetrad indices by use of the metric $g_{ab} = \eta_{ab}$ and its inverse $g^{ab} = \eta^{ab}$.

The *commutation functions* related to the tetrad are the quantities ${\gamma^a}_{bc}(x^i)$ defined by the *commutators* of the basis vectors:[11]

$$[\,\mathbf{e}_a,\,\mathbf{e}_b\,] = {\gamma^c}_{ab}(x^i)\,\mathbf{e}_c \quad \Rightarrow \quad {\gamma^a}_{bc}(x^i) = -\,{\gamma^a}_{cb}(x^i)\ . \tag{69}$$

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

$${\gamma^a}_{bc}(x^i) = {e^a}_i\,(\,{e_b}^j\,\partial_j{e_c}^i - {e_c}^j\,\partial_j{e_b}^i\,) = -\,2\,{e_b}^i\,{e_c}^j\,\nabla_{[i}{e^a}_{j]}\ . \tag{70}$$

These quantities vanish iff the basis $\{\,\mathbf{e}_a\,\}$ is a coordinate basis: that is, there exist coordinates x^i such that $\mathbf{e}_a = {\delta_a}^i\,\partial/\partial x^i$, iff $[\,\mathbf{e}_a,\,\mathbf{e}_b\,] = 0 \Leftrightarrow {\gamma^a}_{bc} = 0$.

The *connection components* ${\Gamma^a}_{bc}$ for the tetrad ('Ricci rotation coefficients') are defined by the relations

$$\nabla_{\mathbf{e}_b}\mathbf{e}_a = {\Gamma^c}_{ab}\,\mathbf{e}_c \quad \Leftrightarrow \quad {\Gamma^c}_{ab} = {e^c}_i\,{e_b}^j\,\nabla_j{e_a}^i\ , \tag{71}$$

[10] The tetrad components of any quantity are invariant when the coordinate basis is changed (for a fixed tetrad), and coordinate components are invariant when a change of tetrad basis is made (for a fixed set of coordinates); however, either change will alter the tetrad components relative to the given coordinates.

[11] Remember that the commutator of any two vectors X, Y is $[X, Y] = XY - YX$.

i.e. it is the c-component of the covariant derivative in the b-direction of the a-vector. It follows that all covariant derivatives can be written out in tetrad components in a way completely analogous to the usual tensor form, for example $\nabla_a T_{bc} = \mathbf{e}_a(T_{bc}) - \Gamma^d{}_{ba}\, T_{dc} - \Gamma^d{}_{ca}\, T_{bd}$, where for any function f, $\mathbf{e}_a(f) = e_a{}^i\, \partial f/\partial x^i$ is the derivative of f in the direction $\mathbf{e}_a$. In particular, because $\mathbf{e}_a(g_{bc}) = 0$ for $g_{ab} = \eta_{ab}$, applying this to the metric gives

$$\nabla_a g_{bc} = 0 \quad \Leftrightarrow \quad -\Gamma^d{}_{ba}\, g_{dc} - \Gamma^d{}_{ca}\, g_{bd} = 0 \quad \Leftrightarrow \quad \Gamma_{(ab)c} = 0 \ , \tag{72}$$

— the rotation coefficients are skew in their first two indices, when we raise and lower the first indices only. We obtain from this and the assumption of vanishing torsion the tetrad relations that are the analogue of the usual Christoffel relations:

$$\gamma^a{}_{bc} = -\,(\Gamma^a{}_{bc} - \Gamma^a{}_{cb}) \ , \qquad \Gamma_{abc} = \tfrac{1}{2}\,(\,g_{ad}\,\gamma^d{}_{cb} - g_{bd}\,\gamma^d{}_{ca} + g_{cd}\,\gamma^d{}_{ab}\,) \ . \tag{73}$$

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

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

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

which follow from the definition of a commutator. Applying this to the basis vectors $\mathbf{e}_a$, $\mathbf{e}_b$ and $\mathbf{e}_c$ gives the identities

$$\mathbf{e}_{[a}(\gamma^d{}_{bc]}) + \gamma^e{}_{[ab}\,\gamma^d{}_{c]e} = 0 \ , \tag{74}$$

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

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

$$R^a{}_{bcd} = \mathbf{e}_c(\Gamma^a{}_{bd}) - \mathbf{e}_d(\Gamma^a{}_{bc}) + \Gamma^a{}_{ec}\,\Gamma^e{}_{bd} - \Gamma^a{}_{ed}\,\Gamma^e{}_{bc} - \Gamma^a{}_{be}\,\gamma^e{}_{cd} \ . \tag{75}$$

Contracting this on a and c, one obtains the EFE (for $\Lambda = 0$) in the form

$$R_{bd} = \mathbf{e}_a(\Gamma^a{}_{bd}) - \mathbf{e}_d(\Gamma^a{}_{ba}) + \Gamma^a{}_{ea}\,\Gamma^e{}_{bd} - \Gamma^a{}_{de}\,\Gamma^e{}_{ba} = T_{bd} - \tfrac{1}{2}\,T\,g_{bd} \ . \tag{76}$$

It is not immediately obvious that this is symmetric, but this follows because (74) implies $R_{a[bcd]} = 0 \Rightarrow R_{ab} = R_{(ab)}$.

3.2 Tetrad formalism in cosmology

For a cosmological model we choose $\mathbf{e}_0$ to be the unit tangent of the matter flow, u^a. This fixing implies that the initial six-parameter freedom of using Lorentz transformations has been reduced to a three-parameter freedom of rotations of the spatial frame $\{\,\mathbf{e}_\alpha\,\}$. The 24 algebraically independent frame components of the space-time connection $\Gamma^a{}_{bc}$ can then be split into the set (see [30, 49, 26])

$$\Gamma_{\alpha 00} = \dot{u}_\alpha \tag{77}$$
$$\Gamma_{\alpha 0\beta} = \tfrac{1}{3}\,\Theta\,\delta_{\alpha\beta} + \sigma_{\alpha\beta} - \epsilon_{\alpha\beta\gamma}\,\omega^\gamma \tag{78}$$
$$\Gamma_{\alpha\beta 0} = \epsilon_{\alpha\beta\gamma}\,\Omega^\gamma \tag{79}$$
$$\Gamma_{\alpha\beta\gamma} = 2\,a_{[\alpha}\,\delta_{\beta]\gamma} + \epsilon_{\gamma\delta[\alpha}\,n^\delta{}_{\beta]} + \tfrac{1}{2}\,\epsilon_{\alpha\beta\delta}\,n^\delta{}_\gamma \ . \tag{80}$$

The first two sets contain the kinematical variables. In the third is the rate of rotation Ω^α of the spatial frame $\{\,\mathbf{e}_\alpha\,\}$ with respect to a *Fermi-propagated* basis. Finally, the quantities a^α and

$n_{\alpha\beta} = n_{(\alpha\beta)}$ determine the 9 spatial rotation coefficients. The commutator equations (69) applied to any space-time scalar f take the form

$$[\,\mathbf{e}_0, \mathbf{e}_\alpha\,]\,(f) \;=\; \dot{u}_\alpha\;\mathbf{e}_0(f) - [\;\tfrac{1}{3}\,\Theta\,\delta_\alpha{}^\beta + \sigma_\alpha{}^\beta + \epsilon_\alpha{}^\beta{}_\gamma\,(\omega^\gamma + \Omega^\gamma)\;]\;\mathbf{e}_\beta(f) \tag{81}$$

$$[\,\mathbf{e}_\alpha, \mathbf{e}_\beta\,]\,(f) \;=\; 2\,\epsilon_{\alpha\beta\gamma}\,\omega^\gamma\;\mathbf{e}_0(f) + [\;2\,a_{[\alpha}\,\delta^\gamma{}_{\beta]} + \epsilon_{\alpha\beta\delta}\,n^{\delta\gamma}\;]\;\mathbf{e}_\gamma(f)\;; \tag{82}$$

The full set of equations for a gravitating fluid can be written as a set of constraints and a set of evolution equations, which include the tetrad form of the $1+3$ covariant equations given above, but complete them by giving all Ricci and Jacobi identities for the basis vectors. We now give these equations.

3.2.1 Constraints

The following set of relations does not contain any frame derivatives with respect to $\mathbf{e}_0$. Hence, we refer to these relations as 'constraints'. From the Ricci identities for u^a and the Jacobi identities we have the (0α)-equation $(C_1)^\alpha$, which, in Hamiltonian treatments of the EFE, is also referred to as the 'momentum constraint', the vorticity divergence identity (C_2) and the H_{ab}-equation $(C_3)^{\alpha\beta}$, respectively; the once-contracted Bianchi identities yield the $(\mathrm{div}\,E)$- and $(\mathrm{div}\,H)$-equations $(C_4)^\alpha$ and $(C_5)^\alpha$ [6, 47]; the constraint $(C_J)^\alpha$ again arises from the Jacobi identities while, finally, $(C_G)^{\alpha\beta}$ and (C_G) stem from the EFE. In detail,

$$\begin{aligned} 0 \;=\; (C_1)^\alpha \;=\; & (\mathbf{e}_\beta - 3\,a_\beta)\,(\sigma^{\alpha\beta}) - \tfrac{2}{3}\,\delta^{\alpha\beta}\,\mathbf{e}_\beta(\Theta) - n^\alpha{}_\beta\,\omega^\beta + q^\alpha \\ & + \epsilon^{\alpha\beta\gamma}\,[\;(\mathbf{e}_\beta + 2\,\dot{u}_\beta - a_\beta)\,(\omega_\gamma) - n_{\beta\delta}\,\sigma^\delta{}_\gamma\;] \end{aligned} \tag{83}$$

$$0 \;=\; (C_2) \;=\; (\mathbf{e}_\alpha - \dot{u}_\alpha - 2\,a_\alpha)\,(\omega^\alpha) \tag{84}$$

$$\begin{aligned} 0 \;=\; (C_3)^{\alpha\beta} \;=\; & H^{\alpha\beta} + (\delta^{\gamma\langle\alpha}\,\mathbf{e}_\gamma + 2\,\dot{u}^{\langle\alpha} + a^{\langle\alpha})\,(\omega^{\beta\rangle}) - \tfrac{1}{2}\,n^\gamma{}_\gamma\,\sigma^{\alpha\beta} + 3\,n^{\langle\alpha}{}_\gamma\,\sigma^{\beta\rangle\gamma} \\ & - \epsilon^{\gamma\delta\langle\alpha}\,[\;(\mathbf{e}_\gamma - a_\gamma)\,(\sigma^{\beta\rangle}{}_\delta) + n^{\beta\rangle}{}_\gamma\,\omega_\delta\;] \end{aligned} \tag{85}$$

$$\begin{aligned} 0 \;=\; (C_4)^\alpha \;=\; & (\mathbf{e}_\beta - 3\,a_\beta)\,(E^{\alpha\beta} + \tfrac{1}{2}\,\pi^{\alpha\beta}) - \tfrac{1}{3}\,\delta^{\alpha\beta}\,\mathbf{e}_\beta(\mu) + \tfrac{1}{3}\,\Theta\,q^\alpha - \tfrac{1}{2}\,\sigma^\alpha{}_\beta\,q^\beta - 3\,\omega_\beta\,H^{\alpha\beta} \\ & - \epsilon^{\alpha\beta\gamma}\,[\;\sigma_{\beta\delta}\,H^\delta{}_\gamma - \tfrac{3}{2}\,\omega_\beta\,q_\gamma + n_{\beta\delta}\,(E^\delta{}_\gamma + \tfrac{1}{2}\,\pi^\delta{}_\gamma)\;] \end{aligned} \tag{86}$$

$$\begin{aligned} 0 \;=\; (C_5)^\alpha \;=\; & (\mathbf{e}_\beta - 3\,a_\beta)\,(H^{\alpha\beta}) + (\mu + p)\,\omega^\alpha + 3\,\omega_\beta\,(E^{\alpha\beta} - \tfrac{1}{6}\,\pi^{\alpha\beta}) - \tfrac{1}{2}\,n^\alpha{}_\beta\,q^\beta \\ & + \epsilon^{\alpha\beta\gamma}\,[\;\tfrac{1}{2}\,(\mathbf{e}_\beta - a_\beta)\,(q_\gamma) + \sigma_{\beta\delta}\,(E^\delta{}_\gamma + \tfrac{1}{2}\,\pi^\delta{}_\gamma) - n_{\beta\delta}\,H^\delta{}_\gamma\;] \end{aligned} \tag{87}$$

$$0 \;=\; (C_J)^\alpha \;=\; (\mathbf{e}_\beta - 2\,a_\beta)\,(n^{\alpha\beta}) + \tfrac{2}{3}\,\Theta\,\omega^\alpha + 2\,\sigma^\alpha{}_\beta\,\omega^\beta + \epsilon^{\alpha\beta\gamma}\,[\;\mathbf{e}_\beta(a_\gamma) - 2\,\omega_\beta\,\Omega_\gamma\;] \tag{88}$$

$$0 \;=\; (C_G)^{\alpha\beta} \;=\; {}^*S^{\alpha\beta} + \tfrac{1}{3}\,\Theta\,\sigma^{\alpha\beta} - \sigma^{\langle\alpha}{}_\gamma\,\sigma^{\beta\rangle\gamma} - \omega^{\langle\alpha}\,\omega^{\beta\rangle} + 2\,\omega^{\langle\alpha}\,\Omega^{\beta\rangle} - (E^{\alpha\beta} + \tfrac{1}{2}\,\pi^{\alpha\beta}) \tag{89}$$

$$0 \;=\; (C_G) \;=\; {}^*R + \tfrac{2}{3}\,\Theta^2 - (\sigma^\alpha{}_\beta\sigma^\beta{}_\alpha) + 2\,(\omega_\alpha\omega^\alpha) - 4\,(\omega_\alpha\Omega^\alpha) - 2\,\mu - 2\,\Lambda\;, \tag{90}$$

where

$${}^*S_{\alpha\beta} \;=\; \mathbf{e}_{\langle\alpha}(a_{\beta\rangle}) + b_{\langle\alpha\beta\rangle} - \epsilon^{\gamma\delta}{}_{\langle\alpha}\,(\mathbf{e}_{|\gamma|} - 2\,a_{|\gamma|})\,(n_{\beta\rangle\delta}) \tag{91}$$

$${}^*R \;=\; 2\,(2\,\mathbf{e}_\alpha - 3\,a_\alpha)\,(a^\alpha) - \tfrac{1}{2}\,b^\alpha{}_\alpha \tag{92}$$

$$b_{\alpha\beta} \;=\; 2\,n_{\alpha\gamma}\,n^{\gamma}{}_{\beta} - n^{\gamma}{}_{\gamma}\,n_{\alpha\beta} \ . \qquad (93)$$

If $\omega^{\alpha} = 0$, so that u^{a} become the normals to a family of 3-spaces of constant time, the last two constraints in the set correspond to the symmetric trace-free and trace parts of the Gauss equation (54). In this case, one also speaks of (C_G) as the generalised Friedmann equation, alias the 'Hamiltonian constraint' or the 'energy constraint'.

3.2.2 Evolution of spatial commutation functions

The 9 spatial commutation functions a^{α} and $n_{\alpha\beta}$ are generally evolved by equations (40) and (41) given in [47]; these originate from the Jacobi identities. Employing each of the constraints $(C_1)^{\alpha}$ to $(C_3)^{\alpha\beta}$ listed in the previous paragraph, we can eliminate $\mathbf{e}_{\alpha}$ frame derivatives of the kinematical variables Θ, $\sigma_{\alpha\beta}$ and ω^{α} from their right-hand sides. Thus, we obtain the following equations for the evolution of the spatial connection components:

$$\begin{aligned} \mathbf{e}_0(a^{\alpha}) \;=\;& -\tfrac{1}{3}\,(\Theta\,\delta^{\alpha}{}_{\beta} - \tfrac{3}{2}\,\sigma^{\alpha}{}_{\beta})\,(\dot{u}^{\beta} + a^{\beta}) + \tfrac{1}{2}\,n^{\alpha}{}_{\beta}\,\omega^{\beta} - \tfrac{1}{2}\,q^{\alpha} \\ & - \tfrac{1}{2}\,\epsilon^{\alpha\beta\gamma}\,[\,(\dot{u}_{\beta} + a_{\beta})\,\omega_{\gamma} - n_{\beta\delta}\,\sigma^{\delta}{}_{\gamma} - (\mathbf{e}_{\beta} + \dot{u}_{\beta} - 2\,a_{\beta})\,(\Omega_{\gamma})\,] + \tfrac{1}{2}\,(C_1)^{\alpha} \end{aligned} \qquad (94)$$

$$\begin{aligned} \mathbf{e}_0(n^{\alpha\beta}) \;=\;& -\tfrac{1}{3}\,\Theta\,n^{\alpha\beta} - \sigma^{\langle\alpha}{}_{\gamma}\,n^{\beta\rangle\gamma} + \tfrac{1}{2}\,\sigma^{\alpha\beta}\,n^{\gamma}{}_{\gamma} - (\dot{u}^{\langle\alpha} + a^{\langle\alpha})\,\omega^{\beta\rangle} - H^{\alpha\beta} + (\delta^{\gamma\langle\alpha}\,\mathbf{e}_{\gamma} + \dot{u}^{\langle\alpha})\,(\Omega^{\beta\rangle}) \\ & - \tfrac{2}{3}\,\delta^{\alpha\beta}\,[\,2\,(\dot{u}_{\gamma} + a_{\gamma})\,\omega^{\gamma} - (\sigma^{\gamma}{}_{\delta}n^{\delta}{}_{\gamma}) + (\mathbf{e}_{\gamma} + \dot{u}_{\gamma})\,(\Omega^{\gamma})\,] \\ & - \epsilon^{\gamma\delta\langle\alpha}\,[\,(\dot{u}_{\gamma} + a_{\gamma})\,\sigma^{\beta\rangle}{}_{\delta} - (\omega_{\gamma} + 2\,\Omega_{\gamma})\,n^{\beta\rangle}{}_{\delta}\,] - \tfrac{2}{3}\,\delta^{\alpha\beta}\,(C_2) + (C_3)^{\alpha\beta} \ . \end{aligned} \qquad (95)$$

3.2.3 Evolution of kinematical variables

The evolution equations for the 9 kinematical variables Θ, ω^{α} and $\sigma_{\alpha\beta}$ are provided by the Ricci identities for u^{a}, i.e.

$$\begin{aligned} \mathbf{e}_0(\Theta) - \mathbf{e}_{\alpha}(\dot{u}^{\alpha}) \;=\;& -\tfrac{1}{3}\,\Theta^2 + (\dot{u}_{\alpha} - 2\,a_{\alpha})\,\dot{u}^{\alpha} - (\sigma^{\alpha}{}_{\beta}\sigma^{\beta}{}_{\alpha}) + 2\,(\omega_{\alpha}\omega^{\alpha}) \\ & - \tfrac{1}{2}\,(\mu + 3p) + \Lambda \end{aligned} \qquad (96)$$

$$\mathbf{e}_0(\omega^{\alpha}) - \tfrac{1}{2}\,\epsilon^{\alpha\beta\gamma}\,\mathbf{e}_{\beta}(\dot{u}_{\gamma}) \;=\; -\tfrac{2}{3}\,\Theta\,\omega^{\alpha} + \sigma^{\alpha}{}_{\beta}\,\omega^{\beta} - \tfrac{1}{2}\,n^{\alpha}{}_{\beta}\,\dot{u}^{\beta} - \tfrac{1}{2}\,\epsilon^{\alpha\beta\gamma}\,[\,a_{\beta}\,\dot{u}_{\gamma} - 2\,\Omega_{\beta}\,\omega_{\gamma}\,] \qquad (97)$$

$$\begin{aligned} \mathbf{e}_0(\sigma^{\alpha\beta}) - \delta^{\gamma\langle\alpha}\,\mathbf{e}_{\gamma}(\dot{u}^{\beta\rangle}) \;=\;& -\tfrac{2}{3}\,\Theta\,\sigma^{\alpha\beta} + (\dot{u}^{\langle\alpha} + a^{\langle\alpha})\,\dot{u}^{\beta\rangle} - \sigma^{\langle\alpha}{}_{\gamma}\,\sigma^{\beta\rangle\gamma} - \omega^{\langle\alpha}\,\omega^{\beta\rangle} \\ & - (E^{\alpha\beta} - \tfrac{1}{2}\,\pi^{\alpha\beta}) + \epsilon^{\gamma\delta\langle\alpha}\,[\,2\,\Omega_{\gamma}\,\sigma^{\beta\rangle}{}_{\delta} - n^{\beta\rangle}{}_{\gamma}\,\dot{u}_{\delta}\,] \ . \end{aligned} \qquad (98)$$

3.2.4 Evolution of matter and Weyl curvature variables

Finally, we have the equations for the 4 matter variables μ and q^{α} and the 10 Weyl curvature variables $E_{\alpha\beta}$ and $H_{\alpha\beta}$, which are obtained from the twice-contracted and once-contracted Bianchi identities, respectively:

$$\mathbf{e}_0(\mu) + \mathbf{e}_{\alpha}(q^{\alpha}) \;=\; -\,\Theta\,(\mu + p) - 2\,(\dot{u}_{\alpha} - a_{\alpha})\,q^{\alpha} - (\sigma^{\alpha}{}_{\beta}\pi^{\beta}{}_{\alpha}) \qquad (99)$$

$$\begin{aligned} \mathbf{e}_0(q^{\alpha}) + \delta^{\alpha\beta}\,\mathbf{e}_{\beta}(p) + \mathbf{e}_{\beta}(\pi^{\alpha\beta}) \;=\;& -\tfrac{4}{3}\,\Theta\,q^{\alpha} - \sigma^{\alpha}{}_{\beta}\,q^{\beta} - (\mu + p)\,\dot{u}^{\alpha} - (\dot{u}_{\beta} - 3\,a_{\beta})\,\pi^{\alpha\beta} \\ & - \epsilon^{\alpha\beta\gamma}\,[\,(\omega_{\beta} - \Omega_{\beta})\,q_{\gamma} - n_{\beta\delta}\,\pi^{\delta}{}_{\gamma}\,] \end{aligned} \qquad (100)$$

$$\mathbf{e}_0(E^{\alpha\beta} + \tfrac{1}{2}\,\pi^{\alpha\beta}) - \epsilon^{\gamma\delta\langle\alpha}\,\mathbf{e}_{\gamma}(H^{\beta\rangle}{}_{\delta}) \;+\; \tfrac{1}{2}\,\delta^{\gamma\langle\alpha}\,\mathbf{e}_{\gamma}(q^{\beta\rangle}) = -\,\tfrac{1}{2}\,(\mu + p)\,\sigma^{\alpha\beta}$$

$$
\begin{aligned}
& - \Theta \, (E^{\alpha\beta} + \tfrac{1}{6}\, \pi^{\alpha\beta}) + 3\, \sigma^{\langle\alpha}{}_{\gamma}\, (E^{\beta\rangle\gamma} - \tfrac{1}{6}\, \pi^{\beta\rangle\gamma}) \\
& + \tfrac{1}{2}\, n^{\gamma}{}_{\gamma}\, H^{\alpha\beta} - 3\, n^{\langle\alpha}{}_{\gamma}\, H^{\beta\rangle\gamma} - \tfrac{1}{2}\, (2\, \dot{u}^{\langle\alpha} + a^{\langle\alpha\rangle})\, q^{\beta\rangle} \\
& + \epsilon^{\gamma\delta\langle\alpha}\, [\, (2\, \dot{u}_{\gamma} - a_{\gamma})\, H^{\beta\rangle}{}_{\delta} \\
& \qquad + (\omega_{\gamma} + 2\, \Omega_{\gamma})\, (E^{\beta\rangle}{}_{\delta} + \tfrac{1}{2}\, \pi^{\beta\rangle}{}_{\delta}) + \tfrac{1}{2}\, n^{\beta\rangle}{}_{\gamma}\, q_{\delta}\,]
\end{aligned} \tag{101}
$$

$$
\begin{aligned}
\mathbf{e}_0(H^{\alpha\beta}) + \epsilon^{\gamma\delta\langle\alpha}\, \mathbf{e}_{\gamma}(E^{\beta\rangle}{}_{\delta} - \tfrac{1}{2}\, \pi^{\beta\rangle}{}_{\delta}) \;=\; & - \Theta\, H^{\alpha\beta} + 3\, \sigma^{\langle\alpha}{}_{\gamma}\, H^{\beta\rangle\gamma} + \tfrac{3}{2}\, \omega^{\langle\alpha}\, q^{\beta\rangle} \\
& - \tfrac{1}{2}\, n^{\gamma}{}_{\gamma}\, (E^{\alpha\beta} - \tfrac{1}{2}\, \pi^{\alpha\beta}) + 3\, n^{\langle\alpha}{}_{\gamma}\, (E^{\beta\rangle\gamma} - \tfrac{1}{2}\, \pi^{\beta\rangle\gamma}) \\
& + \epsilon^{\gamma\delta\langle\alpha}\, [\, a_{\gamma}\, (E^{\beta\rangle}{}_{\delta} - \tfrac{1}{2}\, \pi^{\beta\rangle}{}_{\delta}) - 2\, \dot{u}_{\gamma}\, E^{\beta\rangle}{}_{\delta} \\
& \qquad + \tfrac{1}{2}\, \sigma^{\beta\rangle}{}_{\gamma}\, q_{\delta} + (\omega_{\gamma} + 2\, \Omega_{\gamma})\, H^{\beta\rangle}{}_{\delta}\,]\ .
\end{aligned} \tag{102}
$$

Exercise: (a) Show how most of these equations are the tetrad version of corresponding $1+3$ covariant equations. For which of the tetrad equations is this not true? (b) Explain why there are no equations for $\mathbf{e}_0(\Omega^{\alpha})$ and $\mathbf{e}_0(\dot{u}^{\alpha})$. [Hint: What freedom is there in choosing the tetrad?]

3.3 Complete set

For a prescribed set of matter equations of state, this gives the complete set of tetrad relations, which can be used to characterise particular families of solutions in detail. It clearly contains all the $1+3$ covariant equations above, plus others required to form a complete set. It can be recast into a symmetric normal hyperbolic form [26], showing the hyperbolic nature of the equations and determining their characteristics. Detailed studies of exact solutions will need a coordinate system and vector basis, and usually it will be advantageous to use tetrads for this purpose, because the tetrad vectors can be chosen in physically preferred directions (see [30, 50] for the use of tetrads to study locally rotationally symmetric space-times, and [49, 51] for Bianchi universes; these cases are both discussed below).

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

4 FLRW universes and observational relations

A particularly important involutive subspace is that of the Friedmann–Lemaître (‘FL’) universes, based on the everywhere-isotropic Robertson–Walker (‘RW’) geometry. It is characterised by a perfect fluid matter tensor and the condition that *local isotropy* holds everywhere:

$$
0 = \dot{u}^a = \sigma_{ab} = \omega^a \quad \Leftrightarrow \quad 0 = E_{ab} = H_{ab} \quad \Rightarrow \quad 0 = X_a = Z_a = \tilde{\nabla}_a p\ , \tag{103}
$$

the first conditions stating the kinematical quantities are locally isotropic, the second that these universes are conformally flat, and the third that they are spatially homogeneous.

Exercise: Show that the implications in this relation follow from the $1+3$ covariant equations in the previous section when $p = p(\mu)$, thus showing that isotropy everywhere implies spatial homogeneity in this case.

4.1 Coordinates and metric

It follows then that (see [52]):

1. Comoving coordinates can be found[12] so that the metric takes the form

$$ds^2 = -\,dt^2 + S^2(t)\,(\,dr^2 + f^2(r)\,d\Omega^2\,)\ , \qquad u^a = \delta^a{}_0\ , \tag{104}$$

where $d\Omega^2 = d\theta^2 + \sin^2\theta\,d\phi^2$, $u_a = -\,\nabla_a t$, and $\dot{S}/S = \frac{1}{3}\,\Theta$, characterising $S(t)$ as the scale factor for distances between any pair of fundamental observers. The expansion of matter depends only on one scale length, so it is isotropic (there is no distortion or rotation).

2. The Ricci tensor ${}^3\!R_{ab}$ is isotropic, so the 3-spaces $\{t = \text{const}\}$ are 3-spaces of constant curvature k/S^2 where k can be normalised to $\pm\,1$, if it is non-zero. Using the geodesic deviation equation in these 3-spaces, one finds that (see [52, 53])

$$f(r) = \sin r\ ,\ r\ ,\ \sinh r \quad \text{if} \quad k = +\,1\ ,\ 0\ ,\ -\,1\ . \tag{105}$$

Thus when $k = +\,1$ the surface area $4\pi\,S^2(t)\,f^2(r)$ of a geodesic 2-sphere in these spaces, centred on the (arbitrary) point $r = 0$, increases to a maximum at $r = \pi/2$ and then decreases to zero again at the antipodal point $r = \pi$; hence the point at $r = 2\pi$ has to be the same point as $r = 0$, and these 3-spaces are necessarily closed, with finite total volume. In the other cases the 3-spaces are usually unbounded and the surface areas of these 2-spaces increase without limit; however, unusual topologies still allow the spatial sections to be closed [54].

Exercise: Find the obvious orthonormal tetrad associated with these coordinates, and determine their commutators and Ricci rotation coefficients.

4.2 Dynamical equations

The remaining non-trivial equations are the energy equation (37), the Raychaudhuri equation (29), which now takes the form

$$3\,\frac{\ddot{S}}{S} + \tfrac{1}{2}\,(\mu + 3p) = 0\ , \tag{106}$$

and the Friedmann equation that follows from (55):

$${}^3\!R = 2\,\mu - \tfrac{2}{3}\,\Theta^2 = \frac{6\,k}{S^2}\ , \tag{107}$$

where k is a constant. Any two of these equations imply the third if $\dot{S} \neq 0$ (the latter equation being a first integral of the other two). All one has to do then to determine the dynamics is to solve the Friedmann equation. The solution depends on what form is assumed for the matter: Usually it is taken to be a perfect fluid with equation of state $p = p(\mu)$, or as a sum of such fluids, or as a scalar field with given potential $V(\phi)$. For the γ-law discussed above, the energy equation integrates to give (56), which can then be used to represent μ in the Friedmann equation.

Exercise: Show that on using the tetrad found above, all the other $1 + 3$ covariant and tetrad equations are identically true when these equations are satisfied.

4.2.1 Basic parameters

As well as the parameters H_0, Ω_0, Ω_Λ and q_0, the FLRW models are characterised by the spatial curvature parameter $K_0 = k/S_0^2 = {}^3\!R_0/6$. These parameters are related by the equations (60) and (62), which are now exact rather than approximate relations.

[12] There are many other coordinate systems in use, for example with different definitions of the radial distance r.

4.2.2 Singularity and ages

The existence of the big bang, and age limits on the universe, follow directly from the Raychaudhuri equation, together with the energy assumption $(\mu + 3p) > 0$ (true at least when quantum fields do not dominate), because the universe is expanding today $(\Theta_0 > 0)$. That is, the singularity theorem above applies in particular to FLRW models. Furthermore, from the Raychaudhuri equation, in any FLRW model, the fundamental age relation holds (see e.g. [52]):

> **Age Theorem**: In an expanding FLRW universe with vanishing cosmological constant and satisfying the active gravitational mass density energy condition, ages are strictly constrained by the Hubble expansion rate: namely, at every instant, the age t_0 of the universe (the time since the big bang) is less than the inverse Hubble constant at that time:
> $$(\mu + 3p) > 0\ ,\ \ \Lambda = 0 \quad \Rightarrow \quad t_0 < 1/H_0\ . \tag{108}$$

More precise ages $t_0(H_0, \Omega_0)$ can be determined for any specific cosmological model from the Friedmann equation (107); in particular, in a matter-dominated early universe the same result will hold with a factor 2/3 on the right-hand side, while in a radiation dominated universe the factor will be 1/2. Note that this relation 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 [52].

The age limits are one of the central issues in modern cosmology [55, 41]. Hipparchos satellite measurements suggest a lowering of the age estimates of globular clusters to about 1.2×10^9 years, together with a decrease in the estimate of the Hubble constant to about $H_0 \approx 50\,\mathrm{km/sec/Mpc}$. This corresponds to a Hubble time $1/H_0$ of about 1.8×10^9 years, implying there is no problem, but red giant and Cepheid measurements suggest $H_0 \approx 72 - 77\,\mathrm{km/sec/Mpc}$ [56], implying the situation is very tight indeed. However, recent supernovae measurements [57] suggest a positive cosmological constant, allowing violation of the age constraint, and hence easing the situation. All these figures should still be treated with caution; the issue is fundamental to the viability of the FLRW models, and still needs resolution.

4.3 Exact and approximate solutions

If $\Lambda = 0$ and the energy conditions are satisfied, FLRW models expand forever from a big bang if $k = -1$ or $k = 0$, and recollapse in the future if $k = +1$. A positive value of Λ gives a much wider choice for behaviours [58, 59].

4.3.1 Simplest models

a) *Einstein static*: $S(t) = \text{const}$, $k = +1$, $\Lambda = \frac{1}{2}(\mu + 3p) > 0$, where everything is constant in space and time, and there is no redshift. This model is unstable (see above).

b) *de Sitter*: $S(t) = S_{\text{unit}} \exp(H\,t)$, $H = \text{const}$, $k = 0$, a steady state solution in a constant curvature space-time: it is empty, because $(\mu + p) = 0$, i.e. it does not contain ordinary matter, but rather a cosmological constant,[13] or a scalar field in the strict 'no-rolling' case. It has ambiguous redshift because the choice of families of worldlines and space sections is not unique in this case; see [60].

c) *Milne*: $S(t) = t$, $k = -1$. This is flat, empty space-time in expanding coordinates (again $(\mu + p) = 0$).

[13] A fluid with $(\mu + p) = 0$ is equivalent to a cosmological constant.

d) *Einstein–de Sitter*: the simplest non-empty expanding model, with

$$k = 0 = \Lambda: \quad S(t) = a\, t^{2/3}\ ,\ a = \text{const} \ \text{ if } \ p = 0\ .$$

$\Omega = 1$ is always identically true in this case (this is the critical density case that just manages to expand forever). The age of such a universe is $t_0 = 2/(3H_0)$; if the cosmological constant vanishes, higher density universes ($\Omega_0 > 1$) will have ages less than this, and lower density universes ($0 < \Omega_0 < 1$) ages between this value and (108). This is the present state of the universe if the standard inflationary universe theory is correct, the high value of Ω then implying that most of the matter in the universe is invisible (the 'dark matter' issue; see [41] for a summary of ways of estimating the matter content of the universe, leading to estimates that the detected matter in the universe in fact corresponds to $\Omega_0 \approx 0.2$ to 0.3). It is thus difficult to reconcile this model with observations (the universe could have flat space sections and a large cosmological constant; but then that is not the Einstein–de Sitter model).

4.3.2 Parametric solutions

Use conformal time $\tau = (1/S_0) \int dt/S(t)$ and rescale $S \to y = S(t)/S_0$. Then for a non-interacting mixture of pressure-free matter and radiation, we find in the three cases $k = +1,\ 0,\ -1$,

$$k = +1: \quad y = \quad \alpha\,(1 - \cos\tau) + \beta\,\sin\tau\ , \tag{109}$$

$$k = 0: \quad y = \quad \alpha\,\tau^2/2 + \beta\,\tau\ , \tag{110}$$

$$k = -1: \quad y = \quad \alpha\,(\cosh\tau - 1) + \beta\,\sinh\tau\ , \tag{111}$$

where $\alpha = S_0^2\, H_0^2\, \Omega_m/2$, $\beta = (S_0^2\, H_0^2\, \Omega_r)^{1/2}$, and, on setting $t = \tau = 0$ when $S = 0$,

$$k = +1: \quad t = \quad S_0\,[\ \alpha\,(\tau - \sin\tau) + \beta\,(1 - \cos\tau)\ ]\ , \tag{112}$$

$$k = 0: \quad t = \quad S_0\,[\ \alpha\,\tau^3/6 + \beta\,\tau^2/2\ ]\ , \tag{113}$$

$$k = -1: \quad t = \quad S_0\,[\ \alpha\,(\sinh\tau - \tau) + \beta\,(\cosh\tau - 1)\ ]\ . \tag{114}$$

It is interesting how in this parametrization the dust and radiation terms decouple; this solution includes as special cases the pure dust solutions, $\beta = 0$, and the pure radiation solution, $\alpha = 0$. The general case represents a smooth transition from a radiation dominated early era to a matter dominated later era, and (if $k \neq 0$) on to a curvature dominated era, recollapsing if $k = +1$.

4.3.3 Early-time solutions

At early times, when matter is relativistic or negligible compared with radiation, the equation of state is $p = \frac{1}{3}\,\mu$ and the curvature term can be ignored. The solution is

$$S(t) = c\, t^{1/2}\ ,\ c = \text{const}\ ,\ \mu = \tfrac{3}{4}\, t^{-2}\ ,\ T = \left(\frac{3}{4a}\right)^{1/4} \frac{1}{t^{1/2}}\ , \tag{115}$$

which determines the expansion time scale during nucleosynthesis and so the way the temperature T varies with time (and hence determines the element fractions produced), and has no adjustable parameters. Consequently the degree of agreement attained between nucleosynthesis theory based on this time scale and element abundance observations [61]–[63] may be taken as supporting both a FLRW geometry and the validity of the EFE at that epoch.

The standard thermal history of the hot early universe (e.g. [61]) follows; going back in time, the temperature rises indefinitely (at least until an inflationary or quantum-dominated epoch occurs), so that the very early universe is an opaque near-equilibrium mixture of elementary particles that combine to form nuclei, atoms, and then molecules after pair production ends and the mix cools down as the universe expands, while various forms of radiation (gravitational radiation, neutrinos, electromagnetic radiation) successively decouple and travel freely through the universe that has become transparent to them. This picture is very well supported by the detection of the extremely accurate black body spectrum of the CBR, together with the good agreement of nucleosynthesis observations with predictions based on the FLRW time scales (115) for the early universe.

Exercise: The early universe was radiation dominated but later became matter dominated (as at the present day). Determine at what values $S_{\rm equ}$ of the scale factor $S(t)$ matter–radiation equality occurs, as a function of Ω_0. For what values of Ω_0 does this occur before decoupling of matter and radiation? (Note that if the universe is dominated by Cold Dark Matter ('CDM') then equality of baryon and radiation density occurs after this time.) When does the universe become curvature dominated?

4.3.4 Scalar field

The inflationary universe models use many approximations to model a FLRW universe with a scalar field ϕ as the dominant contribution to the dynamics, so allowing accelerating models that expand quasi-exponentially through many efoldings at a very early time [64, 65], possibly leading to a very inhomogeneous structure on very large (super-particle-horizon) scales [66]. This then leads to important links between particle physics and cosmology, and there is a very large literature on this subject. If an inflationary period occurs in the very early universe, the matter and radiation densities drop very close to zero while the inflaton field dominates, but is restored during 'reheating' at the end of inflation when the scalar field energy converts to radiation.

This will not be pursued further here, except to make one point: because the potential $V(\phi)$ is unspecified (the nature of the inflaton is not known) and the initial value of the 'rolling rate' $\dot{\phi}$ can be chosen at will, it is possible to specify a precise procedure whereby any desired evolutionary history $S(t)$ is attained by appropriate choice of the potential $V(\phi)$ and the initial 'rolling rate' (see [67] for details). Thus, inflationary models may be adjusted to give essentially any desired results in terms of expansion history.

4.3.5 Kinetic theory

While a fluid description is used most often, it is also of interest to use a kinetic theory description of the matter in the universe [68]. The details of collisionless isotropic kinetic models in a FLRW geometry are given by Ehlers, Geren and Sachs [69]; this is extended to collisions in [70]. Curiously it is also possible to obtain exact anisotropic collisionless solutions in FLRW geometries; details are given in [71].

4.4 Phase planes

From these equations, as well as finding simple exact solutions, one can determine evolutionary phase planes for this family of models; see Refsdal and Stabell [59] for (Ω_m, q_0), Ehlers and Rindler [72] for $(\Omega_m, \Omega_r, q_0)$, Wainwright and Ellis [51] for (Ω_0, H_0), and Madsen and Ellis [38] for (Ω, S). The latter are based on the phase-plane equation

$$\frac{d\Omega}{dS} = -\left(3\gamma - 2\right)\frac{\Omega}{S}\left(1 - \Omega\right) . \tag{116}$$

This equation is valid for any γ, i.e. for arbitrary relations between μ and p, but gives a (Ω, S) phase plane flow if $\gamma = \gamma(\Omega, S)$, and in particular if $\gamma = \gamma(S)$ or $\gamma =$ const. Non-static solutions can be followed through turnaround points where $\dot{S} = 0$ (and so Ω is infinite). This enables one to attain complete (time-symmetric) phase planes for models with and without inflation; see [38] and [73] for details.

4.5 Observations

Astronomical observations are based on radiation travelling to us on the *geodesic null rays* that generate our *past light cone*. In the case of a FLRW universe, we may consider only radial null rays as these are generic (because of spatial homogeneity, we can choose the origin of coordinates on any light ray of interest; because of isotropy, light rays travelling in any direction are equivalent to those travelling in any other direction). Thus we may consider geodesic null rays travelling in the FLRW metric (104) such that $ds^2 = 0 = d\theta = d\phi$; then it follows that $0 = -\,dt^2 + S^2(t)\,dr^2$ on these geodesics. Hence, radiation emitted at E and received at O obeys the basic relations

$$r = \int_E^O dr = \int_{t_E}^{t_0} \frac{dt}{S(t)} = \int_{S_E}^{S_0} \frac{dS}{S(t)\,\dot{S}(t)}\ , \tag{117}$$

where the term $\dot{S}$ may be found from the Friedmann equation (107), once a suitable matter description has been chosen.

4.5.1 Redshift

The first fundamental quantity is *redshift*. Considering two successive pulses sent from E to O, each remaining at the same comoving coordinate position, it follows from (117) that the cosmological redshift in a FLRW model is given by

$$(1+z_c) = \frac{\lambda_0}{\lambda_E} = \frac{\Delta T_0}{\Delta T_E} = \frac{S(t_0)}{S(t_E)}\ , \tag{118}$$

and so directly measures the expansion of the universe between when light was emitted and when it is received. Two comments are in order. First, redshift is essentially a time-dilation effect, and will be apparent in all observations of a source, not just in its spectra; this characterisation has the important consequences that (i) redshift is achromatic — the fractional shift in wavelength is independent of wavelength, (ii) the width of any emitted frequency band $d\nu_E$ is altered proportional to the redshift when it reaches the observer, i.e. the observed width of the band is $d\nu_0 = (1+z)\,d\nu_E$, and (iii) the observed rate of emission of radiation and the rate of any time variation in its intensity will both also be proportional to $(1+z)$. Second, there can be local gravitational and Doppler contributions z_0 at the observer, and z_E at the emitter; observations of spectra tell us the overall redshift z, given by

$$(1+z) = (1+z_0)\,(1+z_c)\,(1+z_E)\ , \tag{119}$$

but cannot tell us what part is cosmological and what part is due to local effects at the source and the observer. The latter can be determined from the CBR anisotropy, but the former can only be estimated by identifying cluster members and subtracting off the mean cluster motion. The essential problem is in identifying which sources should be considered members of the same cluster. This is the source of the controversies between Arp et al. and the rest of the observational community (see e.g. Field et al [74]).

4.5.2 Areas

The second fundamental issue is apparent size. Considering light rays converging to the observer at time t_0 in a solid angle $d\Omega = \sin\theta\,d\theta\,d\phi$, from the metric form (104) the corresponding null rays[14] will be described by constant values of θ and ϕ and at the time t_E will encompass an area

[14] Bounded by geodesics located at (ϕ_0, θ_0), $(\phi_0 + d\phi, \theta_0)$, $(\phi_0, \theta_0 + d\theta)$, $(\phi_0 + d\phi, \theta_0 + d\theta)$.

$dA = S^2(t_E)f^2(r)d\Omega$ orthogonal to the light rays, where r is given by (117). Thus, on defining the *observer area distance* $r_0(z)$ by the standard area relation, we find

$$dA = r_0^2\, d\Omega \quad \Rightarrow \quad r_0^2 = S^2(t_E)\, f^2(r)\ . \tag{120}$$

Because these models are isotropic about each point, the *same* distance will relate the observed angle α corresponding to a linear length scale ℓ orthogonal to the light rays:

$$\ell = r_0\, \alpha\ . \tag{121}$$

One can now calculate r_0 from this formula together with (117) and the Friedmann equation, or from the geodesic deviation equation (see [53]), to obtain for a non-interacting mixture of matter and radiation [75],

$$r_0(z) = \frac{1}{H_0 q_0(q_0+\beta-1)}\,\frac{\left[\,(q_0-1)\left\{1+2q_0z+q_0z^2(1-\beta)\right\}^{1/2}-(q_0-q_0\beta z-1)\,\right]}{(1+z)^2}\ , \tag{122}$$

where β represents the matter to radiation ratio: $(1-\beta)\,\rho_{m0} = 2\,\beta\,\rho_{r0}$. The standard Mattig relation for pressure-free matter is obtained for $\beta = 1$ [76], and the corresponding radiation result for $\beta = 0$.

An important consequence of this relation is refocusing of the past light cone: the universe as a whole acts as a gravitational lens, so that there is a redshift z_* such that the area distance reaches a maximum there and then decreases for larger z; correspondingly, the apparent size of an object of fixed size would reach a minimum there and then increase as the object was moved further away [77]. As a specific example, in the simplest (Einstein–de Sitter) case with $p = \Lambda = k = 0$, we find

$$\beta = 1\ ,\ q_0 = \tfrac{1}{2} \quad \Rightarrow \quad r_0(z) = \frac{2}{H_0}\,\frac{1}{(1+z)^{3/2}}\,(\,\sqrt{1+z}-1\,)\ , \tag{123}$$

which refocuses at $z_* = 5/4$ [78]; objects further away will look the same size as much closer objects. For example, an object at a redshift $z_1 = 1023$ (i.e. at about last scattering) will appear the same angular size as an object of identical size at redshift $z_2 = 0.0019$ (which is very close — it corresponds to a speed of recession of about $570\,\mathrm{km/sec}$). In a low density universe, refocusing takes place further out, at redshifts up to $z \approx 4$, depending on the density, and with apparent sizes depending on possible source size evolution [79].

The predicted (*angular size, distance*)–relations are difficult to test observationally because objects of more or less fixed size (such as spherical galaxies) do not have sharp edges that can be used for measuring angular size and so one has rather to measure isophotal diameters (see e.g. [80]), while objects with well-defined linear dimensions, such as double radio sources, are usually rapidly evolving and so one does not know their intrinsic size. Thus, these tests, while in principle clean, are in fact difficult to use in practice.

4.5.3 Luminosity and reciprocity theorem

There is a remarkable relation between upgoing and downgoing bundles of null geodesics connecting the source at t_E and the observer at t_0. Define *galaxy area distance* r_G as above for observer area distance but for the upgoing rather than downgoing bundle of null geodesics. The expression for this distance will be exactly the same as (120) except that the times t_E and t_0 will be interchanged. Consequently, on using the redshift relation (118),

Reciprocity Theorem: The observer area distance and galaxy area distance are identical up to redshift factors:

$$\frac{r_0^2}{r_G^2} = \frac{1}{(1+z)^2} \; . \tag{124}$$

This is true in any space-time as a consequence of the standard first integral of the geodesic deviation equation [81, 6].

Now from photon conservation, the *flux of light* received from a source of *luminosity* L at time t_E will be measured to be

$$F = \frac{L(t_E)}{4\pi}\,\frac{1}{(1+z)^2}\,\frac{1}{r_G^2} \; ,$$

with r given by (117), and the two factors $(1+z)$ coming from photon redshift and time dilation of the emission rate, respectively. On using the reciprocity result this becomes

$$F = \frac{L(t_E)}{4\pi}\,\frac{1}{(1+z)^4}\,\frac{1}{r_0^2} \; , \tag{125}$$

where $r_0(q_0, z)$ is given by (122). On taking logarithms, this gives the standard (*luminosity, redshift*)–relation of observational cosmology [77]. Observations of this Hubble relation basically agree with these predictions, but are not accurate enough to distinguish between the various FLRW models. The hopes that this relation would determine q_0 from galaxy observations have faded away because of the major problem of *source evolution*: we do not know what the source luminosity would have been at the time of evolution. We lack standard candles of known luminosity (or equivalently, rigid objects of known linear size, from which apparent size measurements would give the answer). Various other distance estimators such as the Tully–Fisher relation have helped considerably, but not enough to give a definitive answer. Happily it now seems that Type Ia supernovae may provide the answer in the next decade, because their luminosity can be determined from their light curves, which should depend only on local physics rather than their evolutionary history. This is an extremely promising development at the present time (see e.g. [57]).

4.5.4 Specific intensity

In practice, we measure (a) in a limited waveband rather than over all wavelengths as the 'bolometric' calculation above suggests; and (b) real detectors measure specific intensity (radiation received per unit solid angle) at each point of an image, rather than total source luminosity. Putting these together, we see that if the *source spectrum* is $\mathcal{I}(\nu_E)$, i.e. a fraction $\mathcal{I}(\nu_E)\,d\nu_E$ of the source radiation is emitted in the frequency range $d\nu_E$, then the observed *specific intensity* at each image point is given by[15]

$$I_\nu \, d_\nu = \frac{B_E}{(1+z)^3}\,\mathcal{I}(\nu(1+z))\,d\nu \; , \tag{126}$$

where B_E is the *surface brightness* of the emitting object, and the area distance $r_0(z)$ has canceled out (because of the reciprocity theorem). This tells us the apparent intensity of radiation detected in each direction — which is independent of (area) distance, and dependent only on the source redshift, spectrum, and surface brightness. Together with the angular diameter relation (121), this determines what is actually measured by a detector [80].

An immediate application is black body radiation: if any radiation is emitted as black body radiation at temperature T_E, it follows (*Exercise!*) from the black body expression $\mathcal{I}_\nu = \nu^3\, b(\nu/T_E)$

[15] Absorption effects will modify this if there is sufficient absorbing matter present; see [6] for relevant formulae.

that the received radiation will also be black body (i.e. have this same black body form) but with a measured temperature of

$$T_0 = \frac{T_E}{(1+z)} \ . \tag{127}$$

Note this is true in all cosmologies: the result does not depend on the FLRW symmetries. The importance of this, of course, is that it applies to the observed CBR.

4.5.5 Number counts

If we observe sources in a given solid angle $d\Omega$ in the distance range $(r, r+dr)$, the corresponding volume is $dV = S^3(t_E)\, r_0^2\, dr\, d\Omega$, so if the source density is $n(t_E)$ and the probability of detection is p, the number of sources observed will be

$$dN = p\, n(t_E)\, dV = p \left[\frac{n(t_E)}{(1+z)^3} \right] S^3(t_0)\, f^2(r)\, dr\, d\Omega \ , \tag{128}$$

with r given by (117). This is the *basic number count relation*, where dr can be expressed in terms of observable quantities such as dz; the quantity in brackets is constant if source numbers are conserved in a FLRW model, that is

$$n(t_E) = n(t_0)(1+z)^3 \, . \tag{129}$$

The FLRW predictions agree with observations only if we allow for source number and/or luminosity evolution (cf. the discussion of spherically symmetric models in the next section); but we have no good theory for source evolution.

The additional problem is that there are many undetectable objects in the sky, including entire galaxies, because they lie below the detection threshold; thus we face the problem of *dark matter*, which is very difficult to detect by cosmological observations except by its lensing effects (if it is clustered) and its effects on the age of the universe (if it is smoothly distributed). The current view is that there is indeed such dark matter, detected particularly through its dynamical effects in galaxies and clusters of galaxies (see [41] for a summary), with the present day total matter density most probably in the range $0.1 \leq \Omega_0 \leq 0.3$, while the baryon density is of the order of $0.01 \leq \Omega_0^{\mathrm{baryons}} \leq 0.03$ (from nucleosynthesis arguments). Thus, most of the dark matter is probably non-baryonic.

To properly deal with source statistics in general, and number counts in particular, one should have a reasonably good model of detection limits. It is highly misleading to represent such limits as depending on source apparent magnitude alone (see Disney [82]); this does not take into account the possible occurrence of low-surface brightness galaxies. A useful model based on both source apparent size and magnitude is presented in Ellis, Perry and Sievers [83], summarized in [52]. One should note from this particularly that if there is an evolution in source size, this has a more important effect on source detectability than an evolution in surface brightness.

Exercise: Explain why an observer in a FLRW model may at late times of its evolution see a situation that looks like an island universe (see [84]).

4.6 Observational limits

The first basic observational limit is that we cannot observe anything outside our past light cone, given by (117). Combined with the finite age of the universe, this leads to a maximum comoving coordinate distance from the origin for matter with which we can have had any causal connection: namely

$$r_{ph}(t_0) = \int_0^{t_0} \frac{dt}{S(t)} \ , \tag{130}$$

which converges for any ordinary matter. Matter outside is not visible to us, indeed we cannot have had any causal contact with it. Consequently (see Rindler [85]), the particles at this comoving coordinate value define the *particle horizon*: they separate that matter which can have had any causal contact with us since the origin of the universe from that which cannot. This is most clearly seen by using Penrose's conformal diagrams, obtained on using as coordinates the comoving radius and conformal time; see Penrose [86] and Tipler, Clarke and Ellis [40]. The present day distance to the horizon is

$$D_{ph}(t_0) = S(t_0)\, r_{ph} = S(t_0) \int_0^{t_0} \frac{dt}{S(t)} \ . \tag{131}$$

From (117), this is a sphere corresponding to infinite measured redshift (because $S(t) \to 0$ as $t \to 0$).

Exercise: Show that once comoving matter has entered the particle horizon, it cannot leave it (i.e. once causal contact has been established in a FLRW universe, it cannot cease).

Actually we cannot even see as far as the particle horizon: on our past light cone information rapidly fades with redshift (because of (126)); and because the early universe is opaque, we can only see (by means of any kind of electromagnetic radiation) to the *visual horizon* (Ellis and Stoeger [87]), which is the sphere at comoving coordinate distance

$$r_{vh}(t_0) = \int_{t_d}^{t_0} \frac{dt}{S(t)} \ , \tag{132}$$

where t_d is the time of *decoupling of matter and radiation*, when the universe became transparent (at about a redshift of $z = 1100$). The matter we see at that time is the matter which emitted the CBR we measure today with a present temperature of $2.73\,^{\circ}$K; its present distance from us is

$$D_{vh}(t_0) = S(t_0)\, r_{vh} \ . \tag{133}$$

If we evaluate these quantities in an Einstein–de Sitter universe, we find an interesting paradox: (re-establishing the fundamental constant c,) the present day distance to the particle horizon is $D_{ph}(t_0) = 3\,ct_0$. The question is how can this be bigger than ct_0 (the distance corresponding to the age of the universe). This suggests that the matter comprising the particle horizon has been moving away from us at *faster* than the speed of light in order to reach that distance. How can this be? To investigate this (Ellis and Rothman [78]), note that the *proper distance* from the origin to a galaxy at comoving coordinate r at time t is $D(t,r) = S(t)\,r$. Its velocity away from us is thus given by the *Hubble law*

$$v = \dot{D} = \dot{S}\, r = \frac{\dot{S}}{S}\, D = H\, D \ . \tag{134}$$

Thus, at any time t, $v = c$ when $D_c(t) = c/H(t)$; this is the *speed of light sphere*, where galaxies are (at the present time) receding away from us at the speed of light; those galaxies at a larger distance will be (instantaneously) moving away at a speed greater than c. In the case of an Einstein–de Sitter universe, it occurs when $D_c = 3\,ct_0/2$. This is precisely half the present distance to the horizon, which is thus *not* the distance where points are moving away from us at the speed of light (however, it *is* the surface of infinite redshift).

To see that this is compatible with local causality, change from Lagrangian coordinates (t, r) to Eulerian coordinates (t, D), where D is the instantaneous proper distance, as above. Then we find the (non-comoving) metric form

$$ds^2 = -\,[\,1 - \frac{(\dot{S}/c)^2}{S^2}\,D^2\,]\,(dct)^2 - 2\,\frac{\dot{S}/c}{S}\,D\,dct\,dD + dD^2 + S^2(t)\,f^2(r(t,D))\,d\Omega^2 \ . \qquad (135)$$

It follows that the local light cones are given by

$$\frac{dD_\pm}{dt} = \frac{\dot{S}}{S}\,D \pm c \ . \qquad (136)$$

It is easily seen then that there is no violation of local causality. We also find from this that the past light cone of $t = t_0$ intersects the family of speed of light spheres at its *maximum distance* from the origin (the place where the past light cone starts refocusing), i.e. at

$$t_* = \frac{8}{27}\,t_0 \ , \qquad D_* = \frac{4}{9}\,ct_0 = \frac{8}{27}\,\frac{c}{H_0} \ , \qquad S(t_*) = \frac{4}{9}\,S(t_0) \ , \qquad z_* = 1.25 \ . \qquad (137)$$

At that intersection, $dD_-/dt = 0$ (maximum distance!), $dD_+/dt = 2c$, so there is no causality violation by the matter moving at speed c relative to the central worldline. That matter is presently at a distance ct_0 from us. By contrast, the matter comprising the visual horizon was moving away from us at a speed $v = 61c$ when it emitted the CBR, and was at a distance of about 10^7 light years from our past worldline at that time. Hence, it is the *fastest moving matter* we shall ever see, but was *not* at the greatest proper distance to which we can see (which is D_*, see (137)). For a full investigation of these matters see [78].

Finally it should be noted that an early inflationary era will move the particle horizon out to very large distances, thus [64, 65] solving the causal problem presented by the isotropy of CBR arriving here from causally disconnected regions (see [87, 78] for the relevant causal diagrams), but it will have no effect on the visual horizon. Thus, it changes the causal limitations, but does not affect the visual limits on the part of the universe we can see.

Exercise: Determine the angular size seen today for the horizon distance $D_{ph}(t_d)$ at the time of decoupling. What is the physical significance of this distance? How might this relate to CBR anisotropies?

4.6.1 Small universes

The existence of visual horizons represent absolute limits on what we can ever know; because of them, we can only hope to investigate a small fraction of all the matter in the universe. Furthermore, they imply we do not in fact have the data needed to predict to the future, for at any time gravitational radiation from as yet unseen objects (e.g. domain walls in a chaotic inflationary universe) may cross the visual horizon and undermine any predictions we may have made [88]. However, there is one exceptional situation: it is possible we live in a small universe, with a spatially closed topology on such a length scale (say, 300 to 800 Mpc) that we have already seen around the universe many times, thus already having seen all the matter there is in the universe. The effect is like being in a room with mirrors on the floor, ceiling, and all walls; images from a finite number of objects seem to stretch to infinity. There are many possible topologies, whatever the sign of k [54]; the observational result — best modelled by considering many identical copies of a basic cell attached to each other in an infinitely repeating pattern[16] — can be very like the real universe (Ellis and Schreiber [89]). In this case we would be able to see our own galaxy many times over, thus being able to observationally examine its historical evolution once we had identified which images of distant galaxies were in fact

[16] In mathematical terms, the universal covering space.

repeated images of our own galaxy.

It is possible the real universe is like this. Observational tests can be carried out by trying to identify the same cluster of galaxies, QSO's [90], or X-ray sources in different directions in the sky [91]; or by detecting circles of identical temperature variation in the CBR sky (Cornish et al [92]). If no such circles are detected, this will be a reasonably convincing proof that we do not live in such a small universe — which has various philosophical advantages over the more conventional models with infinite spatial sections [88]. Inter alia they give some degree of mixing of CBR modes so giving a potentially powerful explanation of the low degree of CBR anisotropy (but this effect is not as strong as some have claimed; see [93]).

4.7 FLRW universes as cosmological models

These models are very successful in explaining the major features of the observed universe — its expansion from a hot big bang leading to the observed galactic redshifts and remnant black body radiation, tied in well with element abundance predictions and observations (Peebles et al [94]). However, these models do not describe the real universe well in an essential way, in that the highly idealized degree of symmetry does not correspond to the lumpy real universe. Thus, they can serve as basic models giving the largest-scale smoothed out features of the observable physical universe, but one needs to perturb them to get realistic ('almost-FLRW') universe models that can be used to examine the inhomogeneities and anisotropies arising during structure formation, and that can be compared in detail with observations. This is the topic of the last sections.

However, there is a major underlying issue: because of their high symmetry, these models are infinitely improbable in the space of all possible universes. This high symmetry represents a very high degree of fine tuning of initial conditions, which is extraordinarily improbable, unless we can show physical reasons why it should develop from much more general conditions. In order to examine that question, one needs to look at much more general models and see if they do indeed evolve towards the FLRW models because of physical processes (this is the chaotic cosmology programme, initiated by Misner [95, 96], and taken up much later by the inflationary universe proposal of Guth [64]). Additionally, while the FLRW models seem good models for the observed universe at the present time, one can ask (a) are they the only possible models that will fit the observations? (b) does the universe necessarily have the same symmetries on very large scales (outside the particle horizon), or at very early and/or very late times?

To study these issues, we need to look at more general models, developing some understanding of their geometry and dynamics. This is the topic of the next section. We will find there is a range of models in addition to the FLRW models that can fulfill all present day observational requirements. Nevertheless, it is important to state that the family of perturbed FLRW models can meet all present observational requirements, *provided* we allow suitable evolution of source properties back in the past. They also provide a powerful theoretical framework for considering the nature of and effects of cosmic evolution. Hence, they are justifiably the standard models of cosmology. No evidence stands solidly against them.

Exercise: Apart from the detection of major anisotropy, there are a series of other observations which could, if they were ever observed, decisively disprove this family of standard models. What are these observations? (See [97] for some.)

4.8 General observational relations

Before moving to that section, we briefly consider general observational relations. The present section has focussed on the observational relations holding in FLRW universe models. However, corresponding generic relations can be found determining observations in arbitrary cosmologies (see Kristian and Sachs [98] and Ellis [6]). The essential points are as follows.

In a general model, observations take place on our *past light cone*, which will develop many cusps and caustics at early times because of gravitational lensing, but is still locally generated by geodesic null rays. The information we receive comes to us along these null rays, with tangent vector

$$k^a = \frac{dx^a}{dv} \ , \quad k_a k^a = 0 \ , \quad k_a = \nabla_a \phi \quad \Rightarrow \quad k^b \nabla_b k^a = 0 \ , \quad k^a \nabla_a \phi = 0 \ . \tag{138}$$

The phase factor ϕ determines the local light cone $\{\phi = \text{const}\}$. Relative to an observer with 4-velocity u^a the null vector k^a determines a redshift factor $(-k_a u^a)$ and a direction e_a:

$$k^a = (-k_b u^b)\,(u^a + e^a) \ , \quad e_a u^a = 0 \ , \quad e_a e^a = 1 \ . \tag{139}$$

Considering the observed variation $\dot{\phi} = u^a \nabla_a \phi$ of the phase ϕ, we see that the observed cosmological redshift z for comoving matter[17] is given by

$$(1+z) = \frac{\lambda_O}{\lambda_E} = \frac{(k_a u^a)_E}{(k_b u^b)_O} \ . \tag{140}$$

Taking the derivative of this equation along k^a, we get the fundamental equation [5, 6]

$$\frac{d\lambda}{\lambda} = -\frac{d(k_a u^a)}{(k_b u^b)} = \left[\ \tfrac{1}{3}\,\Theta + (\dot{u}_a e^a) + (\sigma_{ab} e^a e^b)\ \right] dl \ , \tag{141}$$

where $dl = (-k_a u^a)\,dv$. This shows directly the isotropic and anisotropic contributions to redshift from the expansion and shear, respectively, and the gravitational redshift contribution from the acceleration.[18] In a FLRW universe, the last two contributions will vanish.

Area distances are defined as before, and because of the geodesic deviation equation, the reciprocity theorem holds unchanged [6].[19] Consequently, the same surface brightness results as discussed above hold generically; specifically, Eq. (126) holds in any anisotropic or inhomogeneous cosmology. The major difference from the isotropic case is that due to the effect of the electric and magnetic Weyl tensors in the geodesic deviation equation, distortions occur in bundles of null geodesics which then cause focusing, resulting both in strong lensing (multiple images, Einstein rings, and arcs related to cusps and caustics in the past light cone) and weak lensing (systematic distortion of images in an observed area); see the lectures by Y Mellier and F Bernardeau, the book by Schneider, Ehlers and Falco [99], and the work by Holz and Wald [100].

Power series equations showing how the kinematical quantities and the electric and magnetic Weyl tensors affect cosmological observations have been given in a beautiful paper by Kristian and Sachs [98]. The generalisation of those relations to generic cosmologies has been investigated by Ellis et al [101], showing how in principle cosmological observations can directly determine the space-time structure on the past null cone, and thence off it. Needless to say, major practical observational difficulties make this a formidable task, but some progress in this direction is possible (see e.g. [88],

[17] Cf. the comment on cosmological and local sources of redshift above.

[18] In a static gravitational field, this will be given by an acceleration potential: $\dot{u}_a = \tilde{\nabla}_a \Phi$; see [5].

[19] Because of the first integrals of the geodesic deviation equation; this result can also be shown from use of Liouville's theorem in kinetic theory.

[102] and [103]).

Exercise: Apart from area distances, distortions, matter densities, and redshifts, a crucial data set needed to completely determine the space-time geometry from the EFE is the transverse velocities of the matter we observe on the past light cone [101]. Consider how one might try to measure these velocity components, and what are the best limits one might place on them by practical measurement techniques. [Hint: One possible route is by solar system interferometry. Another is by the Sunyaev–Zel'dovich effect [104].]

5 Solutions with symmetries

5.1 Symmetries of cosmologies

Symmetries of a space or a space-time (generically, 'space') are transformations of the space into itself that leave the metric tensor and all physical and geometrical properties invariant. We deal here only with continuous symmetries, characterised by a continuous group of transformations and associated vector fields [105].

5.1.1 Killing vectors

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

$$(L_\xi g)_{ij} = 0 \quad \Leftrightarrow \quad \nabla_{(i}\xi_{j)} = 0 \quad \Leftrightarrow \quad \nabla_i \xi_j = -\nabla_j \xi_i \ , \tag{142}$$

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

$$\nabla_i \nabla_j \xi_k = R^m{}_{ijk}\, \xi_m \ , \tag{143}$$

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

$$\nabla_l \nabla_i \nabla_j \xi_k = (\nabla_l R^m{}_{ijk})\, \xi_m + R^m{}_{ijk}\, \nabla_l \xi_m \ . \tag{144}$$

The set of all KV's forms a Lie algebra with a basis $\{\, \xi_a \,\}$, $a = 1, 2, \ldots, r$, of dimension $r \leq \frac{1}{2}\, n\, (n-1)$. ξ^i_a denote the components with respect to a local coordinate basis, a, b, c label the KV basis, and i, j, k the coordinate components. Any KV can be written in terms of this basis, with *constant coefficients*. Hence: if we take the commutator $[\, \xi_a, \, \xi_b \,]$ of two of the basis KV's, this is also a KV, and so can be written in terms of its components relative to the KV basis, which will be constants. We can write the constants as $C^c{}_{ab}$, obtaining[20]

$$[\, \xi_a, \, \xi_b \,] = C^c{}_{ab}\, \xi_c \ , \qquad C^a{}_{bc} = C^a{}_{[bc]} \ . \tag{145}$$

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

$$C^a{}_{e[b} C^e{}_{cd]} = 0 \ , \tag{146}$$

(which is just equation (74) specialized to the case of a set of vectors with constant commutation functions). These are the integrability conditions that must be satisfied in order that the Lie algebra exist in a consistent way. The transformations generated by the Lie algebra form a Lie group of the same dimension (see Eisenhart [105] or Cohn [106]).

[20]Cf. equation (69).

Arbitrariness of the basis: We can change the basis of KV's in the usual way;

$$\xi_{a'} = \Lambda_{a'}{}^{a}\,\xi_a \qquad \Leftrightarrow \qquad \xi^i_{a'} = \Lambda_{a'}{}^{a}\,\xi^i_a \ , \tag{147}$$

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

$$C^{c'}{}_{a'b'} = \Lambda^{c'}{}_{c}\,\Lambda_{a'}{}^{a}\,\Lambda_{b'}{}^{b}\,C^{c}{}_{ab} \ . \tag{148}$$

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

5.1.2 Groups of isometries

The isometries of a space of dimension n must be a group, as the identity is an isometry, the inverse of an isometry is an isometry, and the composition of two isometries is an isometry. Continuous isometries are generated by the Lie algebra of KV's. The group structure is determined locally by the Lie algebra, in turn characterised by the structure constants [106]. The action of the group is characterised by the nature of its orbits in space; this is only partially determined by the group structure (indeed the same group can act as a space-time symmetry group in quite different ways).

5.1.3 Dimensionality of groups and orbits

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

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

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

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

The *dimension r of the group of symmetries* of a space of dimension n is $r = s + q$ (translations plus rotations). From the above limits , $0 \leq r \leq n + \frac{1}{2}\,n\,(n-1) = \frac{1}{2}\,n\,(n+1)$ (the maximal number of translations and of rotations). This shows the Lie algebra of KV's is finite dimensional.

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

$$R_{ijkl} = K\,(\,g_{ik}\,g_{jl} - g_{il}\,g_{jk}\,)\ , \tag{149}$$

with K a constant; and K necessarily *is* a constant if this equation is true and $n \geq 3$. One cannot get $q = \frac{1}{2}\,n\,(n-1) - 1$ so $r \neq \frac{1}{2}\,n\,(n+1) - 1$.

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

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

5.2 Classification of cosmological symmetries

We consider non-empty perfect fluid models, i.e. (15) holds with $(\mu + p) > 0$.

For a cosmological model, because space-time is 4-dimensional, the possibilities for the dimension of the surface of transitivity are $s = 0, 1, 2, 3, 4$. As to isotropy, we assume $(\mu + p) \neq 0$; then $q = 3$, 1, or 0 because u^a is invariant and so the isotropy group at each point has to be a sub-group of the rotations acting orthogonally to u^a (and there is no 2-dimensional subgroup of $O(3)$.) The dimension q of the isotropy group can vary over the space (but not over an orbit): it can be greater at special points (e.g. an axis centre of symmetry) where the dimension s of the orbit is less, but r (the dimension of the total symmetry group) must stay the same everywhere. Thus the possibilities for isotropy at a general point are:

a) Isotropic: $q = 3$, the Weyl tensor vanishes, kinematical quantities vanish except Θ. All observations (at every point) are isotropic. This is the FLRW family of geometries;

b) Local Rotational Symmetry ('LRS'): $q = 1$, the Weyl tensor is of algebraic Petrov type D, kinematical quantities are rotationally symmetric about a preferred spatial direction. All observations at every general point are rotationally symmetric about this direction. All metrics are known in the case of dust [30] and a perfect fluid (see [50] and also [107]).

c) Anisotropic: $q = 0$; there are no rotational symmetries. Observations in each direction are different from observations in each other direction.

Putting this together with the possibilities for the dimensions of the surfaces of transitivity, we have the following possibilities (see Figure 1):

5.2.1 Space-time homogeneous models

These models with $s = 4$ are unchanging in space and time, hence μ is a constant, so by the energy conservation equation (37) they cannot expand: $\Theta = 0$. Thus by (140) they cannot produce an almost isotropic redshift, and are not useful as models of the real universe. Nevertheless they are of some interest.

Dimension Isotropy Group	Dim invariant variety: s = 2 inhomogeneous	s = 3 spatially homogeneous	s = 4 space-time homogeneous
q = 0 aniso- tropic	generic metric form known. Spatially self-similar, Abelian G_2 on 2-d spacelike surfaces, non-Abelian G_2	Bianchi: orthogonal, tilted	Osvath/Kerr
q = 1 LRS	Lemaitre-Tolman- Bondi family	Kantowski-Sachs, LRS Bianchi	G"odel
q = 3 isotropic	none (cannot happen)	Friedmann	Einstein static
	two non-ignorable coordinates	one non-ignorable coordinate	algebraic EFE (no redshift)

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

Figure 1: Classification of cosmological models (with $(\mu + p) > 0$) by isotropy and homogeneity.

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

Exercise: What other features make this space-time problematic as a cosmological model?

The *LRS case* $q = 1$ ($\Rightarrow r = 5$) is the Gödel stationary rotating universe [108], also with no redshifts. This model was important because of the new understanding it brought as to the nature of time in General Relativity (see [3, 40, 109]). Inter alia, it is a model in which causality is violated (there exist closed timelike lines through each space-time point) and there exists no cosmic time function whatsoever.

The anisotropic models $q = 0$ ($\Rightarrow r = 4$) are all known, [110], but are interesting only for the light they shed on Mach's principle; see [111].

5.2.2 Spatially homogeneous universes

These models with $s = 3$ are the major models of theoretical cosmology, because they express mathematically the idea of the 'cosmological principle': all points of space at the same time are equivalent to each other [112].

The *isotropic case* $q = 3$ ($\Rightarrow r = 6$) is the family of FLRW models, the standard models of cosmology discussed above that have the comoving metric form (104).

The *LRS case* $q = 1$ ($\Rightarrow r = 4$) is the family of Kantowski–Sachs universes [113]–[115] plus the LRS orthogonal [49] and tilted [116] Bianchi models. The simplest are the Kantowski–Sachs family, with comoving metric form

$$ds^2 = -\, dt^2 + A^2(t)\, dr^2 + B^2(t)\, (\, d\theta^2 + f^2(\theta)\, d\phi^2\,) \ , \tag{150}$$

where $f(\theta)$ is given by (105).

The *anisotropic case* $q = 0$ ($\Rightarrow r = 3$) is the family of Bianchi universes with a group of isometries G_3 acting simply transitively on spacelike surfaces. They can be orthogonal or tilted; the simplest class is the Bianchi Type I family, discussed later in this section. The family as a whole has quite complex properties; these models are discussed in the following section.

5.2.3 Spatially inhomogeneous universes

These models have $s \leq 2$.

The *LRS cases* ($q = 1 \Rightarrow s = 2, r = 3$) are the spherically symmetric family with comoving metric form

$$ds^2 = -\, C^2(t,r)\, dt^2 + A^2(t,r)\, dr^2 + B^2(t,r)\, (\, d\theta^2 + f^2(\theta)\, d\phi^2\,) \ , \tag{151}$$

where $f(\theta)$ is given by (105). In the dust case, we can set $C(t,r) = 1$ and can integrate the EFE analytically; for $k = +1$, these are the Lemaître–Tolman–Bondi ('LTB') spherically symmetric models [117]–[119], discussed later in this section. They may have a centre of symmetry (a timelike worldline), and can even allow two such centres, but they cannot be isotropic about a general point (because isotropy everywhere implies spatial homogeneity; see the discussion of FLRW models).

The *anisotropic cases* ($q = 0 \Rightarrow s \leq 2, r \leq 2$) include solutions admitting an Abelian or non-Abelian group of isometries G_2, and spatially self-similar models (see e.g. [51]).

Solutions with no symmetries at all have $r = 0 \Rightarrow s = 0, q = 0$. The real universe, of course, belongs to this class; all the others are intended as approximations to this unique universe. Remarkably, we know some exact solutions without symmetries, specifically (a) the Szekeres quasi-spherical models [120, 121], that are in a sense non-linear FLRW perturbations [122], with comoving metric form

$$ds^2 = -\,dt^2 + e^{2A}\,dx^2 + e^{2B}(dy^2 + dz^2)\ , \qquad A = A(t,x,y,z)\ , \qquad B = B(t,x,y,z)\ , \tag{152}$$

(b) Stephani's conformally flat models [123, 124], and (c) Oleson's type N solutions (for a discussion of these and all the other inhomogeneous models, see Krasiński [8] and Kramer et al [125]). One further interesting family without global symmetries are the Swiss-Cheese models made by cutting and pasting segments of spherically symmetric models. These are discussed below.

We now discuss the simplest useful anisotropic and inhomogeneous models, before turning to the Bianchi models in the next section.

5.3 Bianchi Type I universes ($s = 3$)

These are the simplest anisotropically expanding universe models. The metric can be given in comoving coordinates in the form [126]

$$ds^2 = -\,dt^2 + X^2(t)\,dx^2 + Y^2(t)\,dy^2 + Z^2(t)\,dz^2\ , \qquad u^a = \delta^a{}_0\ . \tag{153}$$

This is the simplest generalisation of the flat FLRW models to allow for different expansion factors in three orthogonal directions; the corresponding *average expansion scale factor* is $S(t) = (XYZ)^{1/3}$. They are spatially homogeneous, being invariant under an Abelian group of isometries G_3 simply transitive on spacelike surfaces $\{t = \text{const}\}$, so $s = 3$; in general $q = 0 \Rightarrow r = 3$, but there are LRS and isotropic subcases (the latter being the Einstein–de Sitter universe). The space sections $\{t = \text{const}\}$ are flat (when $t = t_0$, all the metric coefficients are constant), and all invariants depend only on the time coordinate t. The fluid flow (orthogonal to these homogeneous surfaces) is necessarily geodesic and irrotational. Thus these models obey the restrictions

$$0 = \dot{u}^a = \omega^a\ , \qquad 0 = X_a = Z_a = \tilde{\nabla}_a p\ , \qquad 0 = {}^3R_{ab}\ . \tag{154}$$

The $1+3$ covariant equations obeyed by these models follow from the $1+3$ covariant equations in subsection 2.2 on making these restrictions. We can find a tetrad in the obvious way from the above coordinates ($e_1{}^i = X(t)^{-1}\,\delta_1{}^i$, etc.); then the tetrad equations of the subsection 3.2 hold with

$$0 = \dot{u}^\alpha = \omega^\alpha = \Omega^\alpha\ , \qquad 0 = a^\alpha = n_{\alpha\beta}\ , \qquad 0 = \mathbf{e}_\alpha(\Theta) = \mathbf{e}_\alpha(\sigma_{\beta\gamma})\ , \qquad 0 = \mathbf{e}_\alpha(\mu) = \mathbf{e}_\alpha(p)\ . \tag{155}$$

It follows that the (0α)-equation (32), which is $(C_1)^\alpha$ in the tetrad form, is identically satisfied, and also that $H_{ab} = 0$ and $\tilde{\nabla}_b E^{ab} = 0$. From the Gauss equation (54), the shear obeys

$$(S^3\sigma_{ab})^{\cdot} = 0 \qquad \Rightarrow \qquad \sigma_{ab} = \frac{\Sigma_{ab}}{S^3}\ , \qquad (\Sigma_{ab})^{\cdot} = 0\ , \tag{156}$$

which implies

$$\sigma^2 = \frac{\Sigma^2}{S^6}\ , \qquad \Sigma^2 = \tfrac{1}{2}\,\Sigma_{ab}\Sigma^{ab}\ , \qquad (\Sigma^2)^{\cdot} = 0\ . \tag{157}$$

All the EFE will then be satisfied if the conservation equation (37), the Raychaudhuri equation (29), and the Friedmann-like equation (55) are satisfied. As in the FLRW case, the latter is the first integral of the other two. Assuming a γ-law equation of state, (56) will be satisfied and, using (157), equation (55) becomes the generalised Friedmann equation,

$$3\,\frac{\dot{S}^2}{S^2} = \frac{\Sigma^2}{S^6} + \frac{M}{S^{3\gamma}}\ . \tag{158}$$

This shows that no matter how small the shear today, it will (for ordinary matter) dominate the very early evolution of the universe, which will then approximate the Kasner vacuum solution [125].

On writing out the tetrad components of the shear equation (156), using the commutator relations (81) to determine the shear components, one finds that the individual length scales are given by,

$$X(t) = S(t)\,\exp(\Sigma_1 W(t))\ ,\quad Y(t) = S(t)\,\exp(\Sigma_2 W(t))\ ,\quad Z(t) = S(t)\,\exp(\Sigma_3 W(t))\ ,$$

where

$$W(t) = \int \frac{dt}{S^3(t)}\ , \tag{159}$$

and the constants Σ_α satisfy

$$\Sigma_1 + \Sigma_2 + \Sigma_3 = 0\ ,\quad \Sigma_1^2 + \Sigma_2^2 + \Sigma_3^2 = 2\Sigma^2\ .$$

These relations can be satisfied by setting

$$\Sigma_\alpha = (2\Sigma/3)\sin\alpha_\alpha\ ,\quad \alpha_1 = \alpha\ ,\quad \alpha_2 = \alpha + \frac{2\pi}{3}\ ,\quad \alpha_3 = \alpha + \frac{4\pi}{3}\ , \tag{160}$$

and α is a constant. Thus the solution is given by choosing a value for γ, and then integrating successively (158) and (159).

Exercise: Show that, on using the obvious tetrad associated with the coordinates above, all the tetrad (and $1+3$ covariant) equations are then satisfied.

For example, in the case of dust ($\gamma = 1$):

$$S(t) = (\frac{9}{2}M\,t^2 + \sqrt{3}\Sigma\,t\,)^{1/3}\ ,\quad W(t) = \frac{1}{\sqrt{3}\Sigma}\ln\left(\frac{t}{\frac{3}{4}Mt + \sqrt{3}\Sigma}\right)\ ,$$

so

$$X(t) = S(t)\left(\frac{t^2}{S(t)^3}\right)^{\frac{2}{3}\sin\alpha_1}\ ,\quad Y(t) = S(t)\left(\frac{t^2}{S(t)^3}\right)^{\frac{2}{3}\sin\alpha_2}\ ,\quad Z(t) = S(t)\left(\frac{t^2}{S(t)^3}\right)^{\frac{2}{3}\sin\alpha_3}\ .$$

The generic case is anisotropic; LRS cases occur when $\alpha = \pi/6$ and $\alpha = \pi/2$ in (160), and isotropy when $\Sigma = 0$.

At late times this isotropizes to give the Einstein–de Sitter model, and, hence, as mentioned above, can be a good model of the real universe if Σ is chosen appropriately. However, at early times, the situation is quite different. As $t \to 0$, provided $\Sigma \neq 0$, then $S(t) \to (\sqrt{3}\Sigma)^{1/3}\,t^{1/3}$ and

$$X(t) \to X_0\,t^{\frac{1}{3}(1+2\sin\alpha_1)}\ ,\quad Y(t) \to Y_0\,t^{\frac{1}{3}(1+2\sin\alpha_2)}\ ,\quad Z(t) \to Z_0\,t^{\frac{1}{3}(1+2\sin\alpha_3)}\ .$$

Plotting the function $f(\alpha) = \frac{2}{3}\,(\frac{1}{2} + \sin\alpha)$, we see that the generic behaviour occurs for $\alpha \neq \pi/2$; in this case two of the powers are positive but one is negative, so going backwards in time, the collapse along the preferred axis reverses and changes to a (divergent) expansion, while collapse continues (divergently) along the two orthogonal direction; the singularity is a *cigar singularity.* Going forward in time, a collapse along the preferred axis stops and reverses to become an expansion. However when $\alpha = \pi/2$, one exponent is positive but the other two are zero. Hence, going back in time, collapse continues divergently along the preferred direction in these LRS solutions back to the singularity, but in the orthogonal directions it slows down and halts; this is a *pancake singularity.* An important consequence in this special case is that horizons are broken in the preferred direction —

communication is possible to arbitrary distance in a cylinder around this axis [3].

One can work out detailed observational relations in these models. Because of the high symmetry, the null geodesics can be found explicitly; those along the three preferred axes are particularly simple. Redshift along each of these axes simply scales with the expansion ratio in that direction. Area distances can be found explicitly [127, 128]. An interesting feature is that all observations will show an eight-fold discrete isotropy symmetry about the preferred axes. One can also work out helium production and CBR anisotropy in these models, following the pioneering paper by Thorne [129]. Because the shear can dominate the dynamics at nucleosynthesis or baryosynthesis time, causing a speeding up of the expansion, one can get quite different results than in the FLRW models. Consequently, one can use the nucleosynthesis observations to limit the shear constant Σ, but still allowing extra freedom at the time of baryosynthesis. The CBR quadrupole anisotropy will directly measure the difference in expansion along the three principal axes since last scattering, and, hence, may also be used to limit the anisotropy parameter Σ. Nucleosynthesis gives stronger limits, because it probes to earlier times. These models have also been investigated in the case of viscous fluids and kinetic theory solutions (Misner [130]), with electromagnetic fields, and also the effects of 'reheating' on the CBR anisotropy and spectrum have been examined; see Rees [131].

Thus these models can have arbitrarily small shear at the present day, thus can be arbitrarily close to an Einstein–de Sitter universe since decoupling, but can be quite different early on.

Exercise: Show how the solutions will be altered by (i) a fluid with simple viscosity: $\pi_{ab} = -\,\eta\,\sigma_{ab}$ with constant viscocity coefficient η, (ii) freely propagating neutrinos [130].

5.4 Lemaître–Tolman–Bondi family ($s = 2$)

The simplest inhomogeneous models are those that are spherically symmetric. In general they are time-dependent, with 2-dimensional spherical surfaces of symmetry: $s = 2$, $q = 1 \Rightarrow r = 3$. The geometry of this family (including the closely related models with flat and negatively curved 2-surfaces of symmetry), is examined in a $1+3$ covariant way by van Elst and Ellis [107], and a tetrad analysis is given by Ellis [30] (the pressure-free case) and Stewart and Ellis [50] (for perfect fluids). Here we only consider the dust case, because then a simple analytic solution is possible; the perfect fluid case includes spherical stellar models and collapse solutions (see e.g. Misner, Thorne and Wheeler [132]).

The general spherically symmetric metric for an irrotational dust matter source in synchronous comoving coordinates is the Lemaître–Tolman–Bondi ('LTB') metric [117]–[119]

$$ds^2 = -\,dt^2 + X^2(t,r)\,dr^2 + Y^2(t,r)\,d\Omega^2 \;, \qquad u^a = \delta^a{}_0 \;. \tag{161}$$

The function $Y = Y(t,r)$ is the areal radius, since the proper area of a sphere of coordinate radius r on a time slice of constant t is $4\pi\,Y^2$ (upon re-establishing factors of π). Solving the EFE [119] shows

$$ds^2 = -\,dt^2 + \frac{[\,Y'(t,r)\,]^2}{1+2E(r)}\,dr^2 + Y^2(t,r)\,d\Omega^2 \;, \tag{162}$$

where $Y'(t,r) = \partial Y(t,r)/\partial r$, and $d\Omega^2 = d\theta^2 + \sin^2\theta\,d\phi^2$, with $Y(t,r)$ obeying a generalised Friedmann equation,

$$\dot{Y}(t,r) = \pm\sqrt{\frac{2\,M(r)}{Y(t,r)} + 2\,E(r)} \;, \tag{163}$$

and the density given by

$$4\pi\,\mu(t,r) = \frac{M'(r)}{Y^2(t,r)\,Y'(t,r)} \;. \tag{164}$$

Equation (163) can be solved in terms of a parameter $\eta = \eta(t, r)$:

$$Y(t,r) = \frac{M(r)}{\mathcal{E}(r)}\,\phi_0(t,r)\ , \qquad \xi(t,r) = \frac{[\,\mathcal{E}(r)\,]^{3/2}\,(t - t_B(r))}{M(r)}\ , \tag{165}$$

where[21]

$$\mathcal{E}(r) = \begin{cases} 2E(r), \\ 1, \\ -2E(r), \end{cases} \qquad \phi_0 = \begin{cases} \cosh\eta - 1, \\ (1/2)\eta^2, \\ 1 - \cos\eta, \end{cases} \qquad \xi = \begin{cases} \sinh\eta - \eta, \\ (1/6)\eta^3, \\ \eta - \sin\eta, \end{cases} \qquad \text{when} \begin{cases} E > 0 \\ E = 0 \\ E < 0 \end{cases}\ , \tag{166}$$

for hyperbolic, parabolic and elliptic solutions, respectively.

The LTB model is characterised by three arbitrary functions of comoving coordinate radius r. $E = E(r) \geq -1$ has a geometrical role, determining the local 'embedding angle' of spatial slices, and also a dynamical role, determining the local energy per unit mass of the dust particles, and, hence, the type of evolution of Y. $M = M(r)$ is the effective gravitational mass with coordinate radius r. $t_B = t_B(r)$ is the local time at which $Y = 0$, i.e., the local time of the big bang — we have a non-simultaneous bang surface. Specification of these three arbitrary functions — $M(r)$, $E(r)$ and $t_B(r)$ — fully determines the model, and whilst all have some type of physical or geometrical interpretation, they admit a freedom to choose the radial coordinate, leaving two physically meaningful choices, e.g. $r = r(M)$, $E = E(M)$, $t_B = t_B(M)$. For particular choices of this initial data, one obtains FLRW models, which, of course, are special cases of these spherical models with very specific initial data. In fact, the FLRW models are obtained if one sets

$$2\,E(r) = -\,k\,r^2\ , \qquad Y(t,r) = S(t)\,r\ , \qquad M(r) = \frac{4\pi}{3}\,\mu(t)\,Y^3\ . \tag{167}$$

The LTB models have been used in a number of interesting ways in cosmology:

* to give simple models of structure formation [133, 134], for example by looking at evolution of a locally open region in a closed universe [135] and evolution of density contrast [136],
* to give universe models that are inhomogeneous on a cosmological scale [137, 138],
* to examine inhomogeneous big bang structures [139],
* to examine CBR anisotropies [140]–[142],
* to investigate observational conditions for spatial homogeneity [143]–[147],
* to trace the effect of averaging on spatial inhomogeneities [148], and
* to look at the relationship between cosmic evolution and closure of the universe [149].

These aspects are discussed in Krasiński's book [8], Part III. Here I will only summarize one interesting result: namely, regarding observational tests of whether the real universe is more like a LTB inhomogeneous model, or a FLRW model. In Mustapha, Hellaby and Ellis [144], the following result is shown:

> **Isotropic Observations Theorem (1)**: Any given isotropic set of source observations $n(z)$ and $m(z)$, together with any given source luminosity and number evolution functions $L(z)$ and $N(z)$, can be fitted by a spherically symmetric dust cosmology — a LTB model — in which observations are spherically symmetric about us because we are located near the central worldline.

[21] Strictly speaking, the hyperbolic, parabolic and elliptic solutions obtain when $YE/M > 0$, $= 0$ and < 0, respectively, since $E = 0$ at a spherical origin in both hyperbolic and elliptic models.

This shows that any spherically symmetric observations we may eventually make can be accommodated by appropriate inhomogeneities in a LTB model — irrespective of what source evolution may occur. In particular, one can find such a model that will fit the observations if there is no source evolution. The following result also holds:

> **Isotropic Observations Theorem (2)**: Given any spherically symmetric geometry and any spherically symmetric set of observations, we can find evolution functions that will make the model compatible with the observations. This applies in particular if we want to fit observations to a FLRW model.

The point of the first result is that it shows that these models — spherical inhomogeneous generalisations of the FLRW models — are viable models of the real universe since decoupling, because they cannot be observationally disproved (at least not in any simple way). The usual response to this is: but the FLRW models are confirmed by observations. The second result clarifies this: yes they are, provided you allow an evolution function to be chosen specifically so that the initially discrepant number counts fit the FLRW model predictions. That is, we *assume* the FLRW geometry, and then determine what source evolution makes this assumption compatible with observations [150]. Without this freedom, the FLRW models are contradicted by radio source observations, for example, which without evolution are better fitted by a flat (Euclidean) model. Thus the FLRW model fit is obtained only because of this freedom, allowed because we do not understand source luminosity and number evolution. The inhomogeneous LTB models provide an alternative understanding of the data; the observations do not contradict them.

Exercise: Suppose (a) observations are isotropic, and (b) we knew the source evolution function and selection function, and (c) we were to observationally show that after taking them into account, the area distance relation has precisely the FLRW form (122) for $\beta = 1$ and the number count relation implies the FLRW form (129). Assuming the space-time matter content is dust, (i) prove from this that the space-time is a FLRW space-time. Now (ii) explain the observational difficulties that prevent us using this exact result to prove spatial homogeneity in practice.

An alternative approach to proving homogeneity is via the Postulate of Uniform Thermal Histories ('PUTH') — that is, the assumption that because we see similar kinds of objects at great distances and nearby, they must have had similar thermal histories. One might then hope that from this one could deduce spatial homogeneity of the space-time geometry (for otherwise the thermal histories would have been different). Unfortunately the argument here is not watertight, as can be shown by a counterexample based on the LTB models (Bonnor and Ellis [143]). Proving — rather than assuming — spatial homogeneity remains elusive. We cannot observationally disprove spatial inhomogeneity. However, we can give a solid argument for it via the EGS theorem discussed below.

5.5 Swiss-Cheese models

Finally, an interesting family of inhomogeneous models is the Swiss-Cheese family of models, obtained by repeatedly cutting out a spherical region from a FLRW model and filling it in with another spherical model: Schwarzschild or LTB, for example. This requires:

(i) locating the 3-dimensional timelike *junction surfaces* $\Sigma_\pm$ in each of the two models;

(ii) defining a proposed *identification* Φ between Σ_+ and Σ_-;

(iii) determining the *junction conditions* that (a) the 3-dimensional metrics of Σ_+ and Σ_- (the first fundamental forms of these surfaces) be isometric under this identification, so that there be no discontinuity when we glue them together — we arrive at the same metric from both sides — and (b) the second fundamental forms of these surfaces (i.e. the covariant first derivatives of the 3-dimensional metrics along the spacelike normal directions) must also be isometric when we make this identification, so that they too are continuous in the resultant space-time — equivalently, there is no discontinuity in the direction of the spacelike unit normal vector as we cross the junction surface Σ (this is the condition that there be no surface layer on Σ once we make the join; see Israel [151]).

Satisfying these junction conditions involves deciding how the 3-dimensional junction surfaces $\Sigma_\pm$ should be placed in the respective background space-times. It follows from them that 4 of the 10 components of T_{ab} must be continuous: if n^a denotes the spacelike (or, in some other matching problems, timelike) unit normal to Σ and $p^a{}_b$ the tensor projecting orthogonal to n^a, then $(T_{ab}n^a n^b)$ and $T_{bc}\, n^b\, p^c{}_a$ must be continuous, but the other 6 components $T_{cd}\, p^c{}_a\, p^d{}_b$ can be discontinuous (thus at the surface of a star, in which case n^a will be spacelike, the pressure is continuous but the energy density can be discontinuous). Conservation of the energy-momentum tensor across the junction surface Σ will then be satisfied by the constraints $0 = (G_{ab} - T_{ab})\, n^a n^b$ and $0 = (G_{bc} - T_{bc})\, n^b\, p^c{}_a$.

(iv) Having determined that these junction conditions can be satisfied for some particular identification of points, one can then proceed to identify these corresponding points in the two surfaces Σ_+ and Σ_-, thus gluing an interior Schwarzschild part to an exterior FLRW part, for example. Because of the reciprocal nature of the junction conditions, it is then clear we could have joined them the other way also, obtaining a well-matched FLRW interior and Schwarzschild exterior.

(v) One can continue in this way, obtaining a family of holes of different sizes in a FLRW model with different interior fillings, with further FLRW model segments fitted into the interiors of some of these regions, obtaining a Swiss-Cheese model. One can even obtain a hierarchically structured family of spherically symmetric vacuum and non-vacuum regions in this way.

It is important to note that one *cannot* match arbitrary masses. It follows from the junction conditions that the Schwarzschild mass in the interior of a combined FLRW–Schwarzschild solution must be the same as the mass that has been removed: $M_{\rm Schw} = (4\pi/3)\, (\mu\, S^3\, r^3)_{\rm FLRW}$. If the masses were wrongly matched, there would be an excess gravitational field from the mass in the interior that would not fit the exterior gravitational field, and the result would be to distort the FLRW geometry in the exterior region — which then would no longer be a FLRW model. Alternatively viewed, the reason this matching of masses is needed is that otherwise we will have fitted the wrong background geometry to the inhomogeneous Swiss-Cheese model — averaging the masses in that model will not give the correct background average [152], and they could not have arisen from rearranging uniformly distributed masses in an inhomogeneous way (this is the content of the Traschen integral constraints [153]). Consequently, there can be *no* long-range effects of such matching: the Schwarzschild mass cannot cause large-scale motions of matter in the FLRW region.

These models were originally developed by Einstein and Strauss [154] (see also Schücking [155]) to examine the effect of the expansion of the universe on the solar system (can we measure the expansion of the universe by laser ranging within the solar system?). Their matching of a Schwarzschild interior to a FLRW exterior showed that this expansion has no effect on the motion of planets in the Schwarzschild region. It does not, however, answer the question as to *where* the boundary between the regions should be placed — which determines which regions are affected by the universal expansion. Subsequent uses of these models have included:

* examining Oppenheimer–Snyder collapse in an expanding universe [156]–[158],
* examining gravitational lensing effects on area distances [159],
* investigating CBR anisotropies [160]–[162],

* modelling voids in large-scale structure [163, 164], perhaps using surface-layers [165],
* modelling the universe as a patchwork of domains of different curvature $k = 0, \pm 1$ [166].

Exercise: Show how appropriate choice of initial data in a LTB model can give an effective Swiss-Cheese model with one centre surrounded by a series of successive FLRW and non-FLRW spherical regions. Can you include (i) flat, (ii) vacuum (Schwarzschild) regions in this construction?

One of the most intriguing questions is what non-spherically symmetric models can be joined regularly onto a FLRW model. Bonnor has shown that some Szekeres anisotropic and inhomogeneous models can be matched to a dust FLRW model across a comoving spherical junction surface [167]. Dyer et al [168] have shown that one can match FLRW and LRS Kasner (anisotropic vacuum Bianchi Type I) models across a flat 3-dimensional timelike junction surface. Optical properties of these Cheese-slice models have been investigated in depth [169].

6 Bianchi universes ($s = 3$)

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

Orthogonal models, with the fluid flow lines orthogonal to the surfaces of homogeneity (Ellis and MacCallum [49], see also [51]);
Tilted models, with the fluid flow lines not orthogonal to the surfaces of homogeneity; the components of the fluid peculiar velocity enter as further variables (King and Ellis [116], see also [170]).

Rotating models *must* be tilted (cf. Eq. (27)), and are much more complex than non-rotating models.

6.1 Constructing Bianchi universes

There are essentially three direct ways of constructing the orthogonal models, all based on properties of a tetrad of vectors $\{ \mathbf{e}_a \}$ that commute with the basis of KV's $\{ \xi_\alpha \}$, and usually with the timelike basis vector chosen parallel to the unit normal $n_a = -\nabla_a t$ to the surfaces of homogeneity, i.e. $\mathbf{e}_0 = \mathbf{n}$.

The *first approach* (Taub [171], Heckmann and Schücking [126]) puts all the time variation in the metric components:

$$ds^2 = -\,dt^2 + \gamma_{\alpha\beta}(t)\,(e^\alpha{}_i(x^\gamma)\,dx^i)\,(e^\beta{}_j(x^\delta)\,dx^j)\ , \tag{168}$$

where $e^\alpha{}_i(x^\gamma)$ are 1-forms inverse to the spatial triad vectors $e_\alpha{}^i(x^\gamma)$, which have the same commutators $C^\alpha{}_{\beta\gamma}$, $\alpha, \beta, \gamma, \ldots = 1, 2, 3$, as the structure constants of the group of isometries and commute with the unit normal vector $\mathbf{e}_0$ to the surfaces of homogeneity; i.e. $\mathbf{e}_\alpha = e_\alpha{}^i\,\partial/\partial x^i$, $\mathbf{e}_0 = \partial/\partial t$ obey the commutator relations

$$[\,\mathbf{e}_\alpha, \mathbf{e}_\beta\,] = C^\gamma{}_{\alpha\beta}\,\mathbf{e}_\gamma\ , \quad [\,\mathbf{e}_0, \mathbf{e}_\alpha\,] = 0\ , \tag{169}$$

where the $C^\gamma{}_{\alpha\beta}$ are the Lie algebra structure constants satisfying the Jacobi identities (146). The EFE (1) become ordinary differential equations for the metric functions $\gamma_{\alpha\beta}(t)$.

The *second approach* is based on use of the automorphism group of the symmetry group with time-dependent parameters. We will not consider it further here (see Collins and Hawking [172], Jantzen [173, 174] and Wainwright and Ellis [51] for a discussion).

The *third approach* (Ellis and MacCallum [49]), which is in our view the preferable one, uses an orthonormal tetrad based on the normals to the surfaces of homogeneity (i.e. $\mathbf{e}_0 = \mathbf{n}$, the unit normal vector to these surfaces). The tetrad is chosen to be invariant under the group of isometries, i.e. the tetrad vectors commute with the KV's, and the metric components in the tetrad are space-time constants, $g_{ab} = \eta_{ab}$; now the dynamical time variation is in the commutation functions for the basis vectors, which then determine the time-(and space-)dependence in the basis vectors themselves. Thus we have an orthonormal basis $\{\, \mathbf{e}_a \,\}$, $a = 0, 1, 2, 3$, such that

$$[\, \mathbf{e}_a, \, \mathbf{e}_b \,] = \gamma^c{}_{ab}(t) \, \mathbf{e}_c \; . \tag{170}$$

The commutation functions $\gamma^a{}_{bc}(t)$, together with the matter variables, are then treated as the dynamical variables. The EFE (1) are first-order equations for these quantities, supplemented by the Jacobi identities for the $\gamma^a{}_{bc}(t)$, which are also first-order equations. It is sometimes useful to introduce also the Weyl tensor components as auxiliary variables, but this is not necessary in order to obtain solutions. Thus the equations needed are just the tetrad equations given in section 3, specialised to the case

$$\dot{u}^\alpha = \omega^\alpha = 0 = \mathbf{e}_\alpha(\gamma^a{}_{bc}) \; . \tag{171}$$

The spatial commutation functions $\gamma^\alpha{}_{\beta\gamma}(t)$ can be decomposed into a time-dependent matrix $n_{\alpha\beta}(t)$ and vector $a^\alpha(t)$ (see (80)), and are equivalent to the structure constants $C^\alpha{}_{\beta\gamma}$ of the symmetry group at each point.[22] In view of (171), the Jacobi identities (88) now take the simple form

$$n^{\alpha\beta} \, a_\beta = 0 \; . \tag{172}$$

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

Class A: $a = 0$,
Class B: $a \neq 0$.

Following Schücking, the adaptation of the Bianchi classification of G_3 group types used is as in Figure 2. Given a specific group type at one instant, this type will be preserved by the evolution equations for the quantities $n_\alpha(t)$ and $a(t)$. This is a consequence of a generic property of the EFE: they will always preserve symmetries in initial data (within the Cauchy development of that data); see Hawking and Ellis [3].

In some cases, the Bianchi groups allow higher symmetry subcases: isotropic (FLRW) or LRS models. Figure 3 gives the Bianchi symmetry groups admitted by FLRW and LRS solutions [49], i.e. these are the simply transitive 3-dimensional subgroups allowed by the full G_6 of isometries (in the FLRW case) and the G_4 of isometries (in the LRS case). The only LRS models not allowing a simply transitive subgroup G_3 are the Kantowski–Sachs models for $k = 1$.

Tilted models can be constructed similarly, as discussed below, or by using non-orthogonal bases in various ways [116]; those possibilities will not be pursued further here.

[22] That is, they can be brought to the canonical forms of the $C^\alpha{}_{\beta\gamma}$ by a suitable change of group-invariant basis (the final normalisation to ± 1 may require changing from normalised basis vectors); the transformation to do so is different at each point and at each time.

Class	Type	n_1	n_2	n_3	a	
A	I	0	0	0	0	Abelian
	II	+ve	0	0	0	
	VI_0	0	+ve	-ve	0	
	VII_0	0	+ve	+ve	0	
	VIII	-ve	+ve	+ve	0	
	IX	+ve	+ve	+ve	0	
B	V	0	0	0	+ve	
	IV	0	0	+ve	+ve	
	VI_h	0	+ve	-ve	+ve	h < 0
	III	0	+ve	-ve	n2n3	same as VI_1
	VII_h	0	+ve	+ve	+ve	h > 0

Figure 2: Canonical structure constants for different Bianchi types. The Class B parameter h is defined as $h = a^2/n_2 n_3$ (see e.g. [51]).

6.2 Dynamics of Bianchi universes

The set of tetrad equations (section 3) with these restrictions will determine the evolution of all the commutation functions and matter variables, and, hence, determine the metric and also the evolution of the Weyl tensor (these are regarded as auxiliary variables). In the case of *orthogonal models* — the fluid 4-velocity u^a is *parallel* to the normal vectors n^a — the matter variables will be just the fluid density and pressure [49]; in the case of *tilted models* — the fluid 4-velocity is *not parallel* to the normals — we also need the peculiar velocity of the fluid relative to the normal vectors [116], determining the fluid energy-momentum tensor decomposition relative to the normal vectors (a perfect fluid will appear as an imperfect fluid in that frame). Various papers relate these equations to variational principles and Hamiltonian formalisms, thus expressing them in terms of a potential formalism that gives an intuitive feel for what the evolution will be like [48, 175]. There have also been many numerical investigations of these dynamical equations and the resulting solutions. We will briefly consider three specific aspects here, then the relation to observations, and finally the related dynamical systems approach.

6.2.1 Chaos in these universes?

An ongoing issue since Misner's discovery of the 'Mixmaster' behaviour of the Type IX universes has been whether or not these solutions show chaotic behaviour as they approach the initial singularity (see [176] and Hobill in [51]). The potential approach represents these solutions as bouncing in an expanding approximately triangular shaped potential well, with three deep troughs attached to the corners. The return map approximation (a series of Kasner-like epochs, separated by collisions with the potential walls and consequent change of the Kasner parameters) suggests the motion is chaotic, but the question is whether this map represents the solutions of the differential equations

```
============================================================
                       Isotropic Bianchi models

 FLRW k = +1: Bianchi IX [two commuting groups]
 FLRW k =  0: Bianchi  I, Bianchi VII_0
 FLRW k = -1: Bianchi  V, Bianchi VII_h

============================================================
                          LRS Bianchi models

Orthogonal             c = 0                c \neq 0

   Taub-NUT I        [KS +1: no subgroup]   Bianchi IX
   Taub-NUT 3        Bianchi I, VII_0       Bianchi II
   Taub-NUT 2        Bianchi III [KS -1]    Bianchi VII_h,
                                            III
Tilted

                     Bianchi V, VII_h
                     Farnsworth,
                     Collins-Ellis

============================================================
```

Figure 3: The Bianchi models permitting higher symmetry subcases. The parameter c is zero iff the preferred spatial vector is hypersurface-orthogonal.

well enough to reach this conclusion (for example, the potential walls are represented as flat in this approximation; and there are times when the solution moves up one of the troughs and then reflects back, but the return map does not represent this part of the motion). Part of the problem is that the usual definitions of chaos in terms of a Lyapunov parameter depend on the definition of time variable used, and there is a good case for changing to conformal Misner time in these investigations.

The issue may have been solved now by an analysis of the motion in terms of the attractors in phase space given by Cornish and Levin [177], suggesting that the motion is indeed chaotic, independent of the definition of time used. There may also be chaos in Type XIII solutions. Moreover, chaotic behaviour near the initial singularity was observed in solutions when a source-free magnetic Maxwell field is coupled to fluid space-times of Type I [178] and Type VI_0 [179].

6.2.2 Horizons and whimper singularities

In tilted Class B models, it is possible for there to be a dramatic change in the nature of the solution. This occurs where the surfaces of homogeneity change from being spacelike (at late times) to being timelike (at early times), these regions being separated by a null surface $\mathcal{H}$, the horizon associated with this change of symmetry. At earlier times the solution is no longer spatially homogeneous — it is inhomogeneous and stationary.[23] Associated with the horizon is a singularity where all scalar quantities are finite but components of the matter energy-momentum tensor diverge when measured in a parallelly propagated frame as one approaches the boundary of space-time (this happens because the parallelly propagated frame gets infinitely rescaled in a finite proper time relative to a family of KV's, which in the limit have this singularity as a fixed point). The matter itself originates at an anisotropic big bang singularity at the origin of the universe in the stationary inhomogeneous region.

Details of how this happens are given in Ellis and King [180], and phase plane diagrams for the simplest models in which this occurs — tilted LRS Type V models — in Collins and Ellis [170]. These models isotropise at late times, and can be arbitrarily similar to a low density FLRW model at the present day.

6.2.3 Isotropisation properties

An issue of importance is whether these models tend to isotropy at early or late times. An important paper by Collins and Hawking [172] shows that for ordinary matter many Bianchi models become anisotropic at very late times, even if they are very nearly isotropic at present. Thus isotropy is unstable in this case. On the other hand, a paper by Wald [181] showed that Bianchi models will tend to isotropise at late times if there is a positive cosmological constant present, implying that an inflationary era can cause anisotropies to die away. The latter work, however, while applicable to models with non-zero tilt angle, did not show this angle dies away, and indeed it does not do so in general (Goliath and Ellis [182]). Inflation also only occurs in Bianchi models if there is not too much anisotropy to begin with (Rothman and Ellis [183]), and it is not clear that shear and spatial curvature are in fact removed in all cases [184]. Hence, some Bianchi models isotropise due to inflation, but not all.

To study this kinds of question properly needs the use of phase planes. These will be discussed after briefly considering observations.

[23] This kind of change happens also in the maximally extended Schwarzschild solution at the event horizon.

6.3 Observational relations

Observational relations in these universes have been examined in detail.

(a) Redshift, area distance, and galaxy observations ((M, z) and (N, z) relations) are considered in MacCallum and Ellis [128]. Anisotropies can occur in all these relations, but many of the models will display discrete isotropies in the sky.

(b) The effect of tilt is to make the universe look inhomogeneous, even though it is spatially homogeneous (King and Ellis [116]). This will be reflected in particular in a dipole anisotropy in number counts, which will thus occur in rotating universes [20].[24]

(c) Element formation will be altered primarily through possible changes in the expansion time scale at the time of nucleosynthesis ([129, 186, 187]). This enables us to put limits on anisotropy from measured element abundances in particular Bianchi types. This effect could in principle go either way, so a useful conjecture [188] is that in fact the effect of anisotropy will always — despite the possible presence of rotation — be to speed up the expansion time scale in Bianchi models.

(d) CBR anisotropies will result in anisotropic universe models, for example many Class B Bianchi models will show a hot-spot and associated spiral pattern in the CBR sky [189]–[191].[25] This enables us to put limits on anisotropy from observed CBR anisotropy limits (Collins and Hawking [189], Bunn et al [192]). If 'reheating' takes place in an anisotropic universe, this will mix anisotropic temperatures from different directions, and hence distort the CBR spectrum [131].

Limits on present-day anisotropy from the CBR and element abundance measurements are very stringent: $|\sigma_0|/\Theta_0 \leq 10^{-6}$ to 10^{-12}, depending on the model. However, because of the anisotropies that can build up in both directions in time, this does not imply that the very early universe (before nucleosynthesis) or late universe will also be isotropic. The conclusion applies back to last scattering (CBR measurements) and to nucleosynthesis (element abundances). In both cases the conclusion is quite model dependent. Although very strong limits apply to some Bianchi models, they are much weaker for other types. Hence, one should be a bit cautious in what one claims in this regard.

6.4 Dynamical systems approach

The most illuminating description of the evolution of families of Bianchi models is a dynamical systems approach based on the use of orthonormal tetrads, presented in detail in Wainwright and Ellis [51]. The main variables used are essentially the commutation functions mentioned above, but rescaled by a common time dependent factor.

6.4.1 Reduced differential equations

The basic idea (Collins [193], Wainwright [194]) is to write the EFE in a way that enables one to study the evolution of the various physical and geometrical quantities *relative to the overall rate of expansion of the universe*, as described by the rate of expansion scalar Θ, or equivalently *the Hubble parameter* $H = \frac{1}{3}\Theta$. The remaining freedom in the choice of orthonormal tetrad needs to be eliminated by specifying the variables Ω^α implicitly or explicitly (for example by specifying them as functions of the $\sigma_{\alpha\beta}$). This also simplifies the other quantities (for example choice of a shear eigenframe will result in the tensor $\sigma_{\alpha\beta}$ being represented by two diagonal terms). One so obtains a

[24] They will also occur in FLRW models seen from a reference frame that is not comoving; hence, they should occur in the real universe if the standard interpretation of the CBR anisotropy as due to our motion relative to a FLRW universe is correct; see Ellis and Baldwin [185].

[25] This result is derived in a gauge-dependent way; it would be useful to have a gauge-invariant version.

reduced set of variables, consisting of H and the remaining commutation functions, which we denote symbolically by

$$\mathbf{x} = (\gamma^a{}_{bc}|_{\text{reduced}}) \ . \tag{173}$$

The physical state of the model is thus described by the vector $(H, \mathbf{x})$. The details of this reduction differ for the Class A and B models, and in the latter case there is an algebraic constraint of the form

$$g(\mathbf{x}) = 0 \ , \tag{174}$$

where g is a homogeneous polynomial.

The idea is now to normalise $\mathbf{x}$ with the Hubble parameter H. We denote the resulting variables by a vector $\mathbf{y} \in R^n$, and write:

$$\mathbf{y} = \frac{\mathbf{x}}{H} \ . \tag{175}$$

These new variables are *dimensionless*, and will be referred to as *expansion-normalised variables*. It is clear that each dimensionless state $\mathbf{y}$ determines a 1-parameter family of physical states $(\mathbf{x}, H)$. The evolution equations for the $\gamma^a{}_{bc}$ lead to evolution equations for H and $\mathbf{x}$ and hence for $\mathbf{y}$. In deriving the evolution equations for $\mathbf{y}$ from those for $\mathbf{x}$, the *deceleration parameter* q plays an important role. The Hubble parameter H can be used to define a scale factor S according to (25)

$$H = \frac{\dot{S}}{S} \ , \tag{176}$$

where $\cdot$ denotes differentiation with respect to t. The deceleration parameter, defined by $q = -\ddot{S}\,S/\dot{S}^2$ (see (58)), is related to $\dot{H}$ according to

$$\dot{H} = -\,(1+q)\,H^2 \ . \tag{177}$$

In order that the evolution equations define a flow, it is necessary, in conjunction with the rescaling (175), to introduce a *dimensionless time variable* τ according to

$$S = S_0\, e^{\tau} \ , \tag{178}$$

where S_0 is the value of the scale factor at some arbitrary reference time. Since S assumes values $0 < S < +\infty$ in an ever-expanding model, τ assumes all real values, with $\tau \to -\infty$ at the initial singularity and $\tau \to +\infty$ at late times. It follows from equations (176) and (178) that

$$\frac{dt}{d\tau} = \frac{1}{H} \ , \tag{179}$$

and the evolution equation (177) for H can be written

$$\frac{dH}{d\tau} = -\,(1+q)\,H \ . \tag{180}$$

Since the right-hand sides of the evolution equations for the $\gamma^a{}_{bc}$ are homogeneous of degree 2 in the $\gamma^a{}_{bc}$, the change (179) of the time variable results in H canceling out of the evolution equation for $\mathbf{y}$, yielding an autonomous differential equation ('DE'):

$$\frac{d\mathbf{y}}{d\tau} = \mathbf{f}(\mathbf{y}) \ , \quad \mathbf{y} \in R^n \ . \tag{181}$$

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

$$g(\mathbf{y}) = 0 \ , \tag{182}$$

which is preserved by the DE. The functions $\mathbf{f} : R^n \to R^n$ and $g : R^n \to R$ are polynomial functions in $\mathbf{y}$. An essential feature of this process is that the evolution equation for H, namely (180), decouples from the remaining equations (181) and (182). In other words, the DE (181) describes the evolution of the non-tilted Bianchi cosmologies, the transformation (175) essentially scaling away the effects of the overall expansion. An important consequence is that the new variables are bounded near the initial singularity.

6.4.2 Equations and orbits

The first step in the analysis is to formulate the EFE, using expansion-normalised variables, as a DE (181) in R^n, possibly subject to a constraint (182). Thus one uses the tetrad equations presented above, now adapted to apply to the variables rescaled in this way. Since τ assumes all real values (for models which expand indefinitely), the solutions of (181) are defined for all τ and hence define a *flow* $\{\phi_\tau\}$ on R^n. The evolution of the cosmological models can thus be analyzed by studying the orbits of this flow in the physical region of state space, which is a subset of R^n defined by the requirement that the matter energy density μ be non-negative, i.e.

$$\Omega(\mathbf{y}) = \frac{\mu}{3H^2} \geq 0 \ , \tag{183}$$

where the density parameter Ω (see (58)) is a dimensionless measure of μ.

The *vacuum boundary*, defined by $\Omega(\mathbf{y}) = 0$, describes the evolution of vacuum Bianchi models, and is an invariant set which plays an important role in the qualitative analysis because vacuum models can be asymptotic states for perfect fluid models near the big-bang or at late times. There are other invariant sets which are also specified by simple restrictions on $\mathbf{y}$ which play a special role: the subsets representing each Bianchi type (Figure 2), and the subsets representing higher symmetry models, specifically the FLRW models and the LRS Bianchi models (according to Figure 3).

It is desirable that the dimensionless state space D in R^n is a *compact set*. In this case each orbit will have non-empty future and past limit sets, and hence there will exist a past attractor and a future attractor in state space. When using expansion-normalised variables, compactness of the state space has a direct physical meaning for ever-expanding models: if the state space is compact, then at the big-bang no physical or geometrical quantity diverges more rapidly than the appropriate power of H, and at late times no such quantity tends to zero less rapidly than the appropriate power of H. This will happen for many models; however, the state space for Bianchi Type VII$_0$ and Type VIII models is non-compact. This lack of compactness manifests itself in the behaviour of the Weyl tensor at late times.

6.4.3 Equilibrium points and self-similar cosmologies

Each ordinary orbit in the dimensionless state space corresponds to a one-parameter family of physical universes, which are conformally related by a constant rescaling of the metric. On the other hand, for an equilibrium point $\mathbf{y}^*$ of the DE (181) (which satisfies $\mathbf{f}(\mathbf{y}^*) = \mathbf{0}$), the deceleration parameter q is a constant, i.e. $q(\mathbf{y}^*) = q^*$, and we find

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

In this case, however, the parameter H_0 is no longer essential, since it can be set to unity by a translation of τ, $\tau \to \tau + \text{const}$; then (179) implies that

$$H\, t = \frac{1}{1+q^*} \ , \tag{184}$$

so that by (173) and (175) the commutation functions are of the form (const) $\times\, t^{-1}$. It follows that the resulting cosmological model is *self-similar*. It then turns out that *to each equilibrium point of the DE (181) there corresponds a unique self-similar cosmological model.* In such a model the physical states at different times differ only by an overall change in the length scale. Such models are expanding, but in such a way that their dimensionless state does not change. They include the flat FLRW model ($\Omega = 1$) and the Milne model ($\Omega = 0$). All vacuum and non-tilted perfect fluid self-similar Bianchi solutions have been given by Hsu and Wainwright [195]. The equilibrium points determine the asymptotic behaviour of other more general models.

6.4.4 Phase planes

Many phase planes can be constructed explicitly. The reader is referred to Wainright and Ellis [51] for a comprehensive presentation and survey of results attained so far. Several interesting points emerge:

* *Relation to lower dimensional spaces*: it seems that the lower dimensional spaces, delineating higher symmetry models, can be skeletons guiding the development of the higher dimensional spaces (the more generic models). This is one reason why study of the exact higher symmetry models is of significance. The way this occurs is the subject of ongoing investigation (the key issue being how the finite dimensional dynamical systems corresponding to models with symmetry are imbedded in and relate to the infinite dimensional dynamical system describing the evolution of models without symmetry).

* *Identification of models in state space*: the analysis of the phase planes for Bianchi models shows that the procedure sometimes adopted of identifying all points in state space corresponding to the same model is not a good idea. For example the Kasner ring that serves as a framework for evolution of many other Bianchi models contains multiple realizations of the same Kasner model. To identify them as the same point in state space would make the evolution patterns very difficult to follow. It is better to keep them separate, but to learn to identify where multiple realizations of the same model occur (which is just the *equivalence problem* for cosmological models).

* *Isotropisation* is a particular issue that can be studied by use of these planes [196, 182]. It turns out that even in the classes of non-inflationary Bianchi models that contain FLRW models as special cases, not all models isotropise at some period of their evolution; and of those that do so, most become anisotropic again at late times. Only an inflationary equation of state will lead to such isotropisation for a fairly general class of models (but in the tilted case it is not clear that the tilt angle will die away [182]); once it has turned off, anisotropic modes will again occur.

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

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

(This follows because the saddle point is a fixed point of the phase flow; consequently the phase flow vector becomes arbitrarily close to zero at all points in a small enough open region around the

FLRW point in state space.)

Consequently, although these models are quite unlike FLRW models at very early and very late times, there is an open set of them that are observationally indistinguishable from a FLRW model (choose L long enough to encompass from today to last coupling or nucleosynthesis, and ϵ to correspond to current observational bounds). Thus there exist many such models that are viable as models of the real universe in terms of compatibility with astronomical observations.

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

(Proof is by linearising the equations above (see the following section) to obtain the Bianchi equations linearised about the FLRW models that occur at the saddle point leading to the intermediate isotropisation. These modes will be the solutions in a small neighbourhood about the saddle point permitted by the linearised equations (given existence of solutions to the non-linear equations, linearisation will not prevent corresponding linearised solutions existing).)

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

Exercise: Most studies of CBR anisotropies and nucleosynthesis are carried out for the Bianchi types that allow FLRW models as special cases (see Figure 3). Show that Bianchi models can approximate FLRW models for extended periods even if they do not belong to those types. What kinds of CBR anisotropies can occur in these models? (See e.g. [51].)

7 Almost-FLRW universes

The real universe is *not* FLRW because of all the structure it contains, and (because of the non-linearity of the EFE) the other exact solutions we can attain have higher symmetry than the real universe. Thus in order to obtain realistic models we can compare with detailed observations we need to approximate, aiming to obtain 'almost-FLRW' models representing a universe that is FLRW-like on a large scale but allowing for generic inhomogeneities on a small scale.

7.1 Gauge problem

The major problem in studying perturbed models is the *gauge problem*, due to the fact that there is *no* identifiable fixed background model in General Relativity. One can start with a unique FLRW universe model with metric $\overline{g}_{ab}$ in some coordinate system, and perturb it to obtain a more realistic model: $\overline{g}_{ab} \to g_{ab} = \overline{g}_{ab} + \delta g_{ab}$, but then the process has no unique inverse: the background model $\overline{g}_{ab}$ is not uniquely determined by the lumpy universe model g_{ab} (no unique tensorial averaging process has been defined that will recover $\overline{g}_{ab}$ from g_{ab}). Many choices can be made. However, the usual variables describing perturbations depend on the way the (fictitious) background model $\overline{g}_{ab}$

is fitted to the real universe g_{ab}; these variables can be given any values one wants by changing this correspondence.[26] For example, the *dimensionless density contrast* δ representing a density perturbation is usually defined by

$$\delta(x) = \frac{\mu(x) - \overline{\mu}(x)}{\mu(x)} \,, \tag{185}$$

where $\mu(x)$ is the actual value of the density at the point x, while $\overline{\mu}(x)$ is the (fictitious) background value there, determined by the chosen mapping of the background model into the realistic lumpy model (Ellis and Bruni [197]). This quantity can be given any value we desire by altering that map; we can for example set it to zero by choosing the real surfaces of constant density to be the background surfaces of constant time and, hence, of constant density. Consequently, perturbation equations written in terms of this variable have as solution both physical modes and gauge modes, the latter corresponding to variation of gauge choice rather than to physical variation.

One way to solve this is by very carefully keeping track of the gauge used and the resulting gauge freedom; see Ma and Bertschinger [198] and Prof. Bertschinger's lectures here. The alternative is to use gauge-invariant variables. A widely used and fundamentally important set of such variables are those introduced by Bardeen [199], and used for example by Bardeen, Steinhardt and Turner [200]. Another possibility is use of *gauge-invariant and* $1+3$ *covariant* ('GIC') *variables*, i.e. variables that are gauge-invariant and also $1+3$ covariantly defined so that they have a clear geometrical meaning, and can be examined in any desired (spatial) coordinate system. That is what will be pursued here.

In more detail: our aim is to examine perturbed models by using $1+3$ covariant variables defined in the real space-time (not the background), deriving exact equations for these variables in that space-time, and then approximating by linearising about a RW geometry to get the linearised equations describing the evolution of density inhomogeneities in almost-FLRW universes. How do we handle gauge invariance in this approach? We rely on the

> **Gauge Invariance Lemma** (Stewart and Walker [201]): If a quantity $T^{\cdots}{}_{\dots}$ vanishes in the background space-time, then it is gauge-invariant (to <u>all</u> orders).

[The proof is straightforward : If $\overline{T}^{\cdots}{}_{\dots} = 0$, then $\delta T^{\cdots}{}_{\dots} = T^{\cdots}{}_{\dots} - \overline{T}^{\cdots}{}_{\dots} = T^{\cdots}{}_{\dots}$, which is manifestly independent of the mapping Φ from $\overline{S}$ to S (it does not matter how we map $\overline{T}^{\cdots}{}_{\dots}$ from $\overline{S}$ to S when $\overline{T}^{\cdots}{}_{\dots}$ vanishes).] The application to almost-FLRW models follows (see Ellis and Bruni [197]), where we use an order-of-magnitude notation as follows: Given a smallness parameter ϵ, $\mathcal{O}[n]$ denotes $\mathcal{O}(\epsilon^n)$, and $A \approx B$ means $A - B = \mathcal{O}[2]$ (i.e. these variables are equivalent to $\mathcal{O}[1]$). When $A \approx 0$ we shall regard A as vanishing when we linearise (for it is zero to the accuracy of relevant first-order calculations). Then,

- Zero-order variables are μ, p, Θ, and their time derivatives, $\dot{\mu}$, $\dot{p}$, $\dot{\Theta}$,
- First-order variables are $\dot{u}^a$, σ_{ab}, ω^a, q^a, π_{ab}, E_{ab}, H_{ab}, X_a, Z_a, and their time and space derivatives.

As these first-order variables all vanish in exact FLRW universes, provided u^a is uniquely defined in the realistic (lumpy) almost-FLRW universe model, they are all uniquely defined GIC variables. Thus this set of variables provides what we wanted: 1+3 covariant variables characterising departures from a FLRW geometry (and, in particular, the spatial inhomogeneity of a universe) that are gauge-invariant when the universe is almost-FLRW. Because they are tensors defined in the real space-time, we can evaluate them in any coordinate system we like in that space-time.

[26] This is often represented implicitly rather than explicitly, by assuming that points with the same coordinate values in the background space and more realistic model map to each other; then the gauge freedom is contained in the coordinate freedom available in the realistic universe model.

7.1.1 Key variables

Two simple gauge-invariant quantities give us the information we need to discuss the time evolution of density fluctuations. The basic quantities we start with are the orthogonal projections of the energy density gradient, i.e. the vector $X_a \equiv \tilde{\nabla}_a \mu$, and of the expansion gradient, i.e. the vector $Z_a \equiv \tilde{\nabla}_a \Theta$. The first can be determined (a) from virial theorem estimates and large-scale structure observations (as, e.g., in the POTENT programme), (b) by observing gradients in the numbers of observed sources and estimating the mass-to-light ratio (Kristian and Sachs [98], Eq. (39)), and (c) by gravitational lensing observations. However, these do not directly correspond to the quantities usually calculated; but two closely related quantities do. The first is the *comoving fractional density gradient*:

$$\mathcal{D}_a \equiv S\,\frac{X_a}{\mu}\ , \tag{186}$$

which is gauge-invariant and dimensionless, and represents the spatial density variation over a fixed comoving scale. Note that S, and so $\mathcal{D}_a$, is defined only up to a constant by equation (25); this allows it to represent the density variation between any neighbouring worldlines. The vector $\mathcal{D}_a$ can be separated into a magnitude $\mathcal{D}$ and direction e_a

$$\mathcal{D}_a = \mathcal{D}\,e_a\ ,\ e_a e^a = 1\ ,\ e_a u^a = 0 \qquad \Rightarrow \qquad \mathcal{D} = (\mathcal{D}_a \mathcal{D}^a)^{1/2}\ . \tag{187}$$

The magnitude $\mathcal{D}$ is the gauge-invariant variable[27] that most closely corresponds to the intention of the usual $\delta = \delta\mu/\mu$ given in (185). The crucial difference from the usual definition is that $\mathcal{D}$ represents a (real) spatial fluctuation, rather than a (fictitious) time fluctuation, and does so in a GIC manner. An important auxiliary variable in what follows is the *comoving spatial expansion gradient*:

$$\mathcal{Z}_a \equiv S\,Z_a\ . \tag{188}$$

The issue now is, can we find a set of equations determining how these variables evolve? Yes we can; they follow from the exact $1+3$ covariant equations of subsection 2.2.

7.2 Dynamical equations

We can determine exact propagation equations along the fluid flow lines for the quantities defined in the previous section, and then linearise these to the almost-FLRW case. The basic linearised equations are given by Hawking [22] (see his equations (13) to (19)); we add to them the linearised propagation equations for the gauge-invariant spatial gradients defined above [197].

7.2.1 Growth of inhomogeneity

Taking the spatial gradient of the equation of energy conservation (37) (for the case of a perfect fluid), we find [197]

$$\tilde{\nabla}_a(\dot{\mu}) + \Theta\,\tilde{\nabla}_a(\mu+p) + (\mu+p)\,\tilde{\nabla}_a\Theta = 0\ ,$$

i.e.

$$h^b{}_a \nabla_b(u^c\nabla_c\mu) + \Theta\,(X_a + \tilde{\nabla}_a p) + (\mu+p)\,Z_a = 0\ .$$

Using Leibniz' Rule and changing the order of integration in the second-derivative term (and noting that the pressure-gradient term cancels on using the momentum conservation equation (38)), we obtain the *fundamental equation for the growth of inhomogeneity*:

$$\dot{X}^{\langle a\rangle} + \tfrac{4}{3}\,\Theta\,X^a = -\,(\mu+p)\,Z^a - \sigma^a{}_b\,X^b + \eta^{abc}\,\omega_b\,X_c\ , \tag{189}$$

[27] Or, equivalently in the linear case, the spatial divergence of $\mathcal{D}_a$.

with source term Z^a. On taking the gradient of the Raychaudhuri equation (29) we find the companion equation for that source term:

$$\dot{Z}^{\langle a\rangle} + \Theta\, Z^a = -\tfrac{1}{2}\, X^a - \sigma^a{}_b\, Z^b + \eta^{abc}\,\omega_b\, Z_c + \dot{u}^a\,\mathcal{R} - 2\,\tilde{\nabla}^a(\sigma^2 - \omega^2) + \tilde{\nabla}^a(\tilde{\nabla}_b \dot{u}^b + \dot{u}_b \dot{u}^b)\;, \quad (190)$$

where

$$\mathcal{R} = -\tfrac{1}{3}\,\Theta^2 + \tilde{\nabla}_a \dot{u}^a + (\dot{u}_a \dot{u}^a) - 2\,\sigma^2 + 2\,\omega^2 + \mu + \Lambda\;. \quad (191)$$

These equations contain no information not implied by the others already given; nevertheless, they are useful in that they are exact equations directly giving the rate of growth of inhomogeneity in the generic (perfect fluid) case, the second, together with the evolution equations above, giving the rate of change of all the source terms in the first.

The procedure now is to systematically approximate all the dynamical equations, and, in particular, the structure growth equations given above, by dropping all terms of second order or higher in the implicit[28] expansion variable ϵ. Thus we obtain the linearised equations as approximations to the exact equations above by noting that in the almost-FLRW case,

$$\dot{X}^{\langle a\rangle} + \tfrac{4}{3}\,\Theta\, X^a = -\,(\mu + p)\, Z^a + \mathcal{O}[2] \quad (192)$$

$$\dot{Z}^{\langle a\rangle} + \Theta\, Z^a = -\,\tfrac{1}{2}\, X^a + \dot{u}^a\,\mathcal{R} + \tilde{\nabla}^a(\tilde{\nabla}_b \dot{u}^b) + \mathcal{O}[2]\;, \quad (193)$$

where

$$\mathcal{R} = -\tfrac{1}{3}\,\Theta^2 + \tilde{\nabla}_a \dot{u}^a + \mu + \Lambda + \mathcal{O}[2]\;. \quad (194)$$

Then we linearise the equations by dropping the terms $\mathcal{O}[2]$, *so from now on in this section '=' means equal up to terms of order* ϵ^2.

7.3 Dust

In the case of dust, $p = 0 \Rightarrow \dot{u}^a = 0$, and the equations (192) and (193) for growth of inhomogeneity become

$$S^{-4}\, h^a{}_b (S^4 X^b)^{\cdot} = -\,\mu\, Z^a \quad (195)$$

$$S^{-3}\, h^a{}_b (S^3 Z^b)^{\cdot} = -\,\tfrac{1}{2}\, X^a\;, \quad (196)$$

This closes up to give a second-order equation (take the time derivative of the first and substitute from the second and the energy conservation equation (37)). To compare with the usual equations, change to the variables $\mathcal{D}_a$ and $\mathcal{Z}_a$ (see (186) and (188)). Then the equations become

$$\dot{\mathcal{D}}^{\langle a\rangle} = -\,\mathcal{Z}^a \quad (197)$$

$$\dot{\mathcal{Z}}^{\langle a\rangle} = -\,\tfrac{2}{3}\,\Theta\,\mathcal{Z}^a - \tfrac{1}{2}\,\mu\,\mathcal{D}^a\;. \quad (198)$$

These directly imply the second-order equation (take the time derivative of the first equation!)

$$0 = \ddot{\mathcal{D}}_{\langle a\rangle} + \tfrac{2}{3}\,\Theta\,\dot{\mathcal{D}}_{\langle a\rangle} - \tfrac{1}{2}\,\mu\,\mathcal{D}_a\;, \quad (199)$$

which is the usual equation for growth of density inhomogeneity in dust universes, and has the usual solutions: when $k = 0$, then[29] $S(t) \propto t^{2/3}$, and we obtain

$$\mathcal{D}_a(t, x^\alpha) = d_{+a}(x^\alpha)\; t^{2/3} + d_{-a}(x^\alpha)\; t^{-1}\;, \qquad \dot{d}_{ia} = 0\;, \quad (200)$$

[28] One could make this variable explicit, but there does not seem to be much gain in doing so.

[29] This is the background rather than the real value of this quantity, which is what should really be used here; it is determined in the real space-time by $\dot{S}/S = \frac{1}{3}\,\Theta$. However, as we are linearising, the difference makes only a second-order change in the coefficients in the equations, which we can neglect.

(where t is proper time along the flow lines).[30] This shows the growing mode that leads to structure formation and the decaying mode that dissipates previously existing inhomogeneities. It has been obtained in a GIC way: all the first-order variables, including in particular those in this equation, are gauge-invariant, and there are no gauge modes. Furthermore, we have available the fully non-linear equations, Eqs. (189) and (190), and so can estimate the errors in the neglected terms, and set up a systematic higher-order approximation scheme for solutions of these equations. Solutions for other background models (with $k = \pm 1$ or $\Lambda \neq 0$) can be obtained by substituting the appropriate values in (199) for the background variables Θ and μ, possibly changing to conformal time to simplify the calculation.

Exercise: What is the growth rate at late times in a low-density universe, when the expansion is curvature dominated and so is linear: $S(t) = c_0\, t$? What if there is a cosmological constant, so the late time expansion is exponential: $S(t) = c_0 \exp(H_0\, t)$, where c_0 and H_0 are constants?

7.3.1 Other quantities

We have concentrated here on the growth of inhomogeneities. However, all the $1+3$ covariant equations of the previous sections apply and can be linearised in a straightforward way to the almost-FLRW case (and one can find suitable coordinates and tetrads in order to employ the complete set of tetrad equations in this context, too). Doing so, one can in particular study vorticity perturbations and gravitational wave perturbations; see the pioneering paper by Hawking [22]. We will not consider these further here; however, a series of interesting issues arise. The GIC approach to gravitational waves is examined in Hogan and Ellis [202] and Dunsby, Bassett and Ellis [23], and the possibility of longitudinal waves besides the familiar transverse ones in van Elst and Ellis [26].

A further important issue is the effect of perturbations on observations (apart from the CBR anisotropies, discussed below). It has been known for a long time that anisotropies can affect area distances as well as redshifts (see Bertotti [203], Kantowski [204]); the Dyer–Roeder formula ([205], see also [99]) can be used at any redshift for those many rays that propagate in the lower-density regions between inhomogeneities; however, this formula is not accurate for those ray bundles that pass very close to matter, where shearing becomes important. This is closely related to the *averaging problem* (see e.g. Ellis [88] or Boersma [206]): how can dynamics and observations of a universe which is basically empty almost everywhere average out correctly to give the same dynamics and observations as a universe which is exactly spatially homogeneous? What differences will there be from the FLRW case? We will not pursue this further here except to state that it is believed this does in fact work out OK: in the fully inhomogeneous case it is the Weyl tensor that causes distortions in the empty spaces between astrophysical objects (as in gravitational lensing), and hence causes convergence of both timelike and null geodesics; these, however, average out to give zero average distortion and the same convergence effect as a FLRW space-time with zero Weyl tensor, but with the Ricci tensor causing focusing of these curves (see [207] for a discussion of the null case; however, some subtleties arise here in terms of the way areas are defined when strong lensing takes place [208, 145, 209]).

7.4 Perfect fluids

A similar GIC analysis to the dust case has been given by Ellis, Hwang and Bruni determining FLRW perturbations for the perfect fluid case [210, 19]. This gives the single-fluid equation for growth of structure in the universe (again, derived in a GIC manner), and includes as special cases a fully $1+3$ covariant derivation of the Jeans length and of the speed of sound for barotropic perfect

[30] We can also see that if $\Theta = 0$, there will be an exponential rather than power-law growth.

fluids. To evaluate the last two terms in (193) when $p \neq 0$, using (38) we see that, to first order,

$$\tilde{\nabla}_a(\tilde{\nabla}_b \dot{u}^b) = -\frac{\tilde{\nabla}_a(\tilde{\nabla}_b \tilde{\nabla}^b p)}{(\mu + p)} \ . \tag{201}$$

But, for simplicity considering only the case of vanishing vorticity,[31] we have

$$\tilde{\nabla}_b \tilde{\nabla}^b \tilde{\nabla}_a p = \tilde{\nabla}_b \tilde{\nabla}_a \tilde{\nabla}^b p \ , \tag{202}$$

and, on using the Ricci identities for the $\tilde{\nabla}$-derivatives and the zero-order relation ${}^3R_{ab} = \frac{1}{3}\,{}^3R\,h_{ab}$ for the 3-dimensional Ricci tensor, we obtain

$$\tilde{\nabla}_a \tilde{\nabla}_b \tilde{\nabla}^b p = \tilde{\nabla}_b \tilde{\nabla}_a \tilde{\nabla}^b p - \tfrac{1}{3}\,{}^3R\ \tilde{\nabla}_a p \ . \tag{203}$$

Thus, on using ${}^3R = \mathcal{K} = 6\,k/S^2$, we find

$$\tfrac{1}{2}\,\mathcal{K}\,\dot{u}_a = -\frac{1}{(\mu+p)}\,\frac{3\,k}{S^2}\,\tilde{\nabla}_a p \ , \qquad \tilde{\nabla}_a(\tilde{\nabla}_b \dot{u}^b) = \frac{1}{(\mu+p)}\left(\frac{2k}{S^2}\,\tilde{\nabla}_a p - \tilde{\nabla}^2 \tilde{\nabla}_a p\right) \ , \tag{204}$$

introducing the notation $\tilde{\nabla}^2 \tilde{\nabla}_a p \equiv \tilde{\nabla}_b \tilde{\nabla}^b \tilde{\nabla}_a p$. In performing this calculation, note that there will *not* be 3-spaces orthogonal to the fluid flow if $\omega^a \neq 0$, but still we can calculate the 3-dimensional orthogonal derivatives as usual (by using the projection tensor $h^a{}_b$); the difference from when $\omega^a = 0$ will be that the quantity we calculate as a curvature tensor, using the usual definition from commutation of second derivatives, will not have all the usual curvature tensor symmetries. Nevertheless, the zero-order equations, representing the curvature of the 3-spaces orthogonal to the fluid flow in the background model, will agree with the linearised equations up to the required accuracy.

Now if $p = p(\mu, s)$, where s is the entropy per particle, we find

$$\tilde{\nabla}_a p = \left(\frac{\partial p}{\partial \mu}\right)_{s=\text{const}} \tilde{\nabla}_a \mu + \left(\frac{\partial p}{\partial s}\right)_{\mu=\text{const}} \tilde{\nabla}_a s \ . \tag{205}$$

We assume we can ignore the second term (pressure variations caused by spatial entropy variations) relative to the first (pressure variations caused by energy density variations) and spatial variations in the scale function S (which would at most cause second-order variations in the propagation equations). Then (ignoring terms due to the spatial variation of $\partial p/\partial \mu$, which will again cause second-order variations) we find in the zero-vorticity case,

$$S\,[\ \tfrac{1}{2}\,\mathcal{K}\,\dot{u}_a + \tilde{\nabla}_a(\tilde{\nabla}_b \dot{u}^b)\ ] = -\frac{1}{(1+p/\mu)}\left(\frac{\partial p}{\partial \mu}\right)\left(\frac{k}{S^2}\,\mathcal{D}_a + \tilde{\nabla}^2 \mathcal{D}_a\right) \ . \tag{206}$$

This is the result that we need in proceeding with (193).

7.4.1 Second-order equations

The equations for propagation can now be used to obtain second-order equations for $\mathcal{D}_a$.[32] For easy comparison, we follow Bardeen [199] by defining

$$w = \frac{p}{\mu} \ , \qquad c_s^2 = \frac{\partial p}{\partial \mu} \qquad \Rightarrow \qquad \left(\frac{p}{\mu}\right)^{\cdot} \equiv \ \dot{w} = -\,\Theta\,(1+w)\,(c_s^2 - w) \ . \tag{207}$$

[31] When it is non-zero, (27) must be taken into account when commuting derivatives; see [19].
[32] And the variable $\Phi_a = \mu\,S^2\,\mathcal{D}_a$ that corresponds more closely to Bardeen's variable; see [19].

Now differentiation of (192), projection orthogonal to u^a, and linearisation gives a second-order equation for $\mathcal{D}_a$ (we use Eqs. (29), (193), (207) and (206) in the process). We find

$$\begin{aligned} 0 &= \ddot{\mathcal{D}}_{\langle a\rangle} + (\tfrac{2}{3} - 2w + c_s^2)\,\Theta\,\dot{\mathcal{D}}_{\langle a\rangle} - [\,(\tfrac{1}{2} + 4w - \tfrac{3}{2}w^2 - 3c_s^2)\,\mu + (c_s^2 - w)\,\frac{12\,k}{S^2}\,]\,\mathcal{D}_a \\ &\quad + c_s^2\,[\,\frac{2\,k}{S^2}\,\mathcal{D}_a - \tilde{\nabla}^2\mathcal{D}_a\,]\,. \end{aligned} \tag{208}$$

This equation is the *basic result of this subsection*; the rest of the discussion examines its properties and special cases. It is a second-order equation determining the evolution of the GIC density variation variable $\mathcal{D}_a$ along the fluid flow lines. It has the form of a *wave equation* with extra terms due to the expansion of the universe, gravitation and the spatial curvature.[33] We bracket the last two terms together, because when we make a harmonic decomposition these terms together give the harmonic eigenvalues n^2.

This form of the equations allows for a variation of $w = p/\mu$ with time. However, if $w = \text{const}$, then from (207) $c_s^2 = w$, and the equation simplifies to

$$0 = \ddot{\mathcal{D}}_{\langle a\rangle} + (\tfrac{2}{3} - w)\,\Theta\,\dot{\mathcal{D}}_{\langle a\rangle} - \tfrac{1}{2}\,(1 - w)\,(1 + 3w)\,\mu\,\mathcal{D}_a + w\,[\,\frac{2\,k}{S^2}\,\mathcal{D}_a - \tilde{\nabla}^2\mathcal{D}_a\,]\,. \tag{209}$$

The matter source term vanishes if $w = 1$ (the case of 'stiff matter' $\Leftrightarrow p = \mu$) or $w = -\,\frac{1}{3}$ (the case $p = -\,\frac{1}{3}\,\mu$, corresponding to matter with no active gravitational mass). Between these two limits ('ordinary matter'), the matter term is positive and tends to cause the density gradient to increase ('gravitational aggregation'); outside these limits, the term is negative and tends to cause the density gradient to decrease ('gravitational smoothing'). Also the sign of the damping term (giving the adiabatic decay of inhomogeneities) is positive if $\frac{2}{3} > w$ (that is, $2\mu > 3p$) but negative otherwise (they adiabatically grow rather than decay in this case). The equation reduces correctly to the corresponding dust equation in the case $w = 0$. Two other cases of importance are:

* *Speed of sound*: when Θ, μ, and k/S^2 can be neglected, we see directly from (208) that c_s introduced above *is* the speed of sound (and that imaginary values of c_s, that is, negative values of $\partial p/\partial\mu$, then lead to exponential growth or decay rather than oscillations).

* *Radiation*: In the case of pure radiation, $\gamma = \frac{4}{3}$ and $w = \frac{1}{3} = c_s^2$. Then we find from (209)

$$0 = \ddot{\mathcal{D}}_{\langle a\rangle} + \tfrac{1}{3}\,\Theta\,\dot{\mathcal{D}}_{\langle a\rangle} - \tfrac{2}{3}\,(\tfrac{1}{3}\,\Theta^2 + \frac{3\,k}{S^2})\,\mathcal{D}_a + \tfrac{1}{3}\,[\,\frac{2\,k}{S^2}\,\mathcal{D}_a - \tilde{\nabla}^2\mathcal{D}_a\,]\,. \tag{210}$$

7.4.2 Harmonic decomposition

It is standard (see, e.g., [22] and [199]) to decompose the variables harmonically, thus effectively separating out the time and space variations by turning the differential equations for time variation of the perturbations as a whole into separate time variation equations for each component of spatial variation, characterised by comoving wavenumber. This conveniently represents the idea of a comoving wavelength for the matter inhomogeneities. In our case we do so by writing $\mathcal{D}_a$ in terms of harmonic vectors $Q_a^{(n)}$, from which the background expansion has been factored out.

We start with the defining equations of the $1+3$ covariant *scalar harmonics* $Q^{(n)}$,

$$\dot{Q}^{(n)} = 0\;, \qquad \tilde{\nabla}^2 Q^{(n)} = -\,\frac{n^2}{S^2}\,Q^{(n)}\;, \tag{211}$$

[33] We have dropped Λ in these equations; it can be represented by setting $w = -\,1$.

corresponding to Bardeen's scalar Helmholtz equation (2.7) [199], but expressed $1+3$ covariantly following Hawking [22]. From these quantities we define the $1+3$ covariant *vector harmonics* (cf. [199], equations (2.8) and (2.10); we do not divide by the wavenumber, however, so our equations are valid even if $n=0$)

$$Q_a^{(n)} \equiv S\,\tilde{\nabla}_a Q^{(n)} \quad \Rightarrow \quad Q_a^{(n)}\,u^a = 0\;, \quad \dot{Q}_{\langle a\rangle}^{(n)} \approx 0\;, \quad \tilde{\nabla}^2 Q_a^{(n)} = -\,\frac{(n^2-2k)}{S^2}\,Q_a^{(n)}\;, \tag{212}$$

(the factor S ensuring these vector harmonics are approximately covariantly constant along the fluid flow lines in the almost-FLRW case). We write $\mathcal{D}_a$ in terms of these harmonics:

$$\mathcal{D}_a = \sum_n \mathcal{D}^{(n)}\,Q_a^{(n)}\;, \qquad \tilde{\nabla}_a\mathcal{D}^{(n)} \approx 0\;, \tag{213}$$

where $\mathcal{D}^{(n)}$ is the harmonic component of $\mathcal{D}_a$ corresponding to the *comoving wavenumber* n, containing the time variation of that component; to first order, $\mathcal{D}^{(n)} \equiv \mu^{(n)}/\mu$. Putting this decomposition in the linearised equations (208) and (209), the harmonics decouple. Thus, for example, we obtain from (209) the n-th harmonic equation

$$0 = \ddot{\mathcal{D}}^{(n)} + (\tfrac{2}{3}-w)\,\Theta\,\dot{\mathcal{D}}^{(n)} - [\;\tfrac{1}{2}\,(1-w)\,(1+3w)\,\mu - w\,\frac{n^2}{S^2}\;]\;\mathcal{D}^{(n)}\;, \tag{214}$$

(valid for each $n \geq 0$), where one can, if one wishes, substitute for μ from the Friedmann equation in terms of Θ and k/S^2. This equation shows how the growth of the inhomogeneity depends on the comoving wavelength. For the case of radiation, $w = \frac{1}{3}$, this is

$$0 = \ddot{\mathcal{D}}^{(n)} + \tfrac{1}{3}\,\Theta\,\dot{\mathcal{D}}^{(n)} - [\;\tfrac{2}{3}\,\mu - \tfrac{1}{3}\,\frac{n^2}{S^2}\;]\;\mathcal{D}^{(n)}\;. \tag{215}$$

7.5 Implications

To determine the solutions explicitly, we have to substitute for μ, Θ and S from the zero-order equations. The most important issue is which terms dominate.

7.5.1 Jeans instability

Jeans' criterion is that *gravitational collapse* will tend to occur if the combination of the matter term and the term containing the Laplace operator in (208) or (209) is *positive* [211]; that is, if

$$\tfrac{1}{2}\,(1-w)\,(1+3w)\,\mu\,\mathcal{D}_a > w\,[\;\frac{2\,k}{S^2}\,\mathcal{D}_a - \tilde{\nabla}^2\mathcal{D}_a\;]\;, \tag{216}$$

when $c_s^2 = w$. Using the harmonic decomposition, this can be expressed in terms of an equivalent scale: from (214) gravitational collapse tends to occur for a mode $\mathcal{D}^{(n)}$ if

$$\tfrac{1}{2}\,(1-w)\,(1+3w)\,\mu > w\,\frac{n^2}{S^2}\;, \tag{217}$$

that is, if

$$n_J \equiv \left[\;(1-w)\left(\frac{1}{w}+3\right)\frac{\mu(t)}{2}\;\right]^{1/2} S(t) > n\;. \tag{218}$$

In terms of wavelengths the *Jeans length* is defined by

$$\lambda_J \equiv \frac{2\pi\,S(t)}{n_J} = c_s\,c\,\sqrt{\frac{\pi}{G\mu(t)}\,\frac{1}{(1-w)(1+3w)}}\;, \tag{219}$$

where we have re-established the fundamental constants c and G ($w = (c_s/c)^2$). Thus, *gravitational collapse* will occur for small n (wavelengths longer than λ_J), but not for large n (wavelengths less than λ_J), for the pressure gradients are then large enough to resist the collapse and lead to *oscillations* instead.

For *non-relativistic matter*, $|w| \ll 1$ and $\mu = \rho_m c^2$, where ρ_m is the mass density of the matter, so

$$\lambda_J = c_s \sqrt{\frac{\pi}{G\rho_m(t)}} \ . \tag{220}$$

Then the *Jeans mass* will be

$$M_J = \frac{4\pi}{3}\,\rho_m\,\lambda_J^3 = \frac{4\pi}{3}\,c_s^3\left(\frac{\pi}{G}\right)^{3/2}\rho_m^{-1/2} \ . \tag{221}$$

For *radiation*, where $w = \frac{1}{3}$ and $\mu = \rho_r c^2$, collapse will occur if

$$(2\mu)^{1/2} < \frac{n_J}{S} \qquad \Leftrightarrow \qquad \lambda > \lambda_J = c_s \sqrt{\frac{3\pi}{4G\rho_r(t)}} \ . \tag{222}$$

The corresponding *Jeans mass* for matter coupled to the radiation will be

$$M_J = \frac{4\pi}{3}\,\rho_m\,\lambda_J^3 = \frac{4\pi}{3}\,c_s^3\left(\frac{3\pi}{4G\rho_r}\right)^{3/2}\rho_m \ . \tag{223}$$

7.5.2 Short-wavelength solutions

For wavelengths much shorter than the Jeans length, equation (214) becomes the damped harmonic equation

$$0 = \ddot{\mathcal{D}}^{(n)} + (\tfrac{2}{3} - w)\,\Theta\,\dot{\mathcal{D}}^{(n)} + w\,\frac{n^2}{S^2}\,\mathcal{D}^{(n)} \ , \tag{224}$$

giving oscillations. In the early universe, during radiation dominated expansion before decoupling, the tight coupling of the dominant radiation and matter leads to a fluid with $w = \frac{1}{3}$; then the short-wavelength equation becomes

$$0 = \ddot{\mathcal{D}}^{(n)} + \tfrac{1}{3}\,\Theta\,\dot{\mathcal{D}}^{(n)} + \tfrac{1}{3}\,\frac{n^2}{S^2}\,\mathcal{D}^{(n)} \ , \tag{225}$$

giving the *acoustic oscillations* during that era for modes such that $\lambda < \lambda_J = c_s c\,(3\pi/4G\mu)^{1/2}$.

7.5.3 Long-wavelength solutions

For wavelengths much longer than the Jeans length, we can drop the Laplace operator terms in (214) to obtain

$$0 = \ddot{\mathcal{D}}^{(n)} + (\tfrac{2}{3} - w)\,\Theta\,\dot{\mathcal{D}}^{(n)} - \tfrac{1}{2}\,(1-w)\,(1+3w)\,\mu\,\mathcal{D}^{(n)} \ . \tag{226}$$

Thus, the second-order propagation equations become ordinary differential equations along the fluid flow lines, easily solved for particular equations of state. In the case of radiation ($w = \frac{1}{3}$) we find

$$0 = \ddot{\mathcal{D}}^{(n)} + \tfrac{1}{3}\,\Theta\,\dot{\mathcal{D}}^{(n)} - \tfrac{2}{3}\left(\tfrac{1}{3}\,\Theta^2 + \frac{3\,k}{S^2}\right)\mathcal{D}^{(n)} \ , \tag{227}$$

when $\lambda > \lambda_J = c_s c\,(3\pi/4G\mu)^{1/2}$. When $k = 0$, then $S(t) \propto t^{1/2}$, and we obtain in this long-wavelength limit

$$\mathcal{D}_a = d_{+a}(x^\alpha)\;t + d_{-a}(x^\alpha)\;t^{-1/2} \ , \qquad \dot{d}_{ia} = 0 \ , \tag{228}$$

(where t is proper time along the flow lines). The corresponding standard results in the synchronous and comoving proper time gauges differ, being modes proportional to t and to $t^{1/2}$; we obtain the same growth law as derived in the comoving time orthogonal gauge and equivalent gauges. As our variables are GIC, we believe they show the latter gauges represent the physics more accurately than any other. Note that, moreover, we obtain no fictitious modes (proportional to t^{-1}) because we are using GIC variables.

7.5.4 Change of behaviour with time

Any particular inhomogeneity will have a constant comoving size and, hence, constant comoving wavelength λ and constant comoving wavenumber n as defined above. However, the Jeans length will *vary with time*.

During the radiation era,[34] $S \propto t^{1/2}$ and $\mu = \frac{3}{4}\, t^{-2}$ (see (115)), so (dropping the dimensional constants) $\rho_m \propto t^{-3/2}$, and the comoving Jeans length

$$\lambda_J = \sqrt{\tfrac{1}{3}\, \pi\, t^2} \propto t \tag{229}$$

will steadily grow to a value $\lambda_J^{\max}$ at matter–radiation equality, while the Jeans mass of coupled matter will grow as

$$M_J = \frac{4\pi}{3}\, \rho_m\, \lambda_J^3 \propto t^{-3/2}\, t^3 = t^{3/2}\ . \tag{230}$$

Thereafter, until recombination, the Jeans mass stays constant: the matter and radiation are still tightly coupled but now the universe is matter dominated and the speed of sound of the coupled fluid depends on the matter density: $c_s = (c/\sqrt{3})\, (\, 1 + (3\rho_m/4\rho_r)\,)$ (see Rees [212]). After recombination the Jeans length and mass will rapidly die away because as the matter and radiation decouple, leading to $c_s \to 0$ and so $\lambda_J \to 0$. Each wavelength λ longer than $\lambda_J^{\max}$ will have a growing mode as in (228) until the Jeans wavelength becomes greater than λ; it will then stop growing and undergo acoustic oscillations which will last until decoupling when the Jeans length drops towards zero and matter-dominated growth starts according to (200). Growth of small perturbations eventually slows down when the universe becomes curvature dominated at late times (when this happens depends on Ω_0; in a critical density universe, it never occurs).

Thus, the *key times* for any wavelength after the initial perturbations have been seeded[35] are (i) t_J, when they become smaller than the Jeans length (if they are small enough that this occurs), (ii) $t_{\rm equ}$, when the matter dominated era starts (which will be before decoupling of matter and radiation, because $\Omega_0 \geq 0.1$), (iii) $t_{\rm dec}$, when decoupling takes place. The acoustic oscillations have constant amplitude in the radiation dominated era from t_J until $t_{\rm equ}$, and then die away as $t^{-1/6}$ in the matter dominated era until they end at $t_{\rm dec}$ [212]. Baryonic inhomogeneities 'freeze out' at that time; they then start growing by damped gravitational attraction. If they grow large enough, changing to non-linear collapse and ultimately star formation, then local energy generation starts.

Exercise: Establish these behaviours from the equations given above.

By contrast CDM freezes out at $t_{\rm equ}$ and starts growth at that time (Rees [97]). Thus in a CDM dominated universe, as is often supposed, the CDM fluctuations that govern structure formation start

[34] This is early enough that we can ignore the curvature term in the Friedmann equation.

[35] The usual assumption is that perturbations are essentially unaffected by all the strong interactions in the early universe after the end of inflation, including the ending of pair production (when matter ceases to be relativistic), decoupling of neutrinos, and the irreversible interactions during baryosynthesis and nucleosynthesis, and decoupling with Hot Dark Matter ('HDM') or CDM.

gravitational growth earlier than the baryons. They then govern the growth of inhomogeneities, attracting the baryons into their potential wells; a 2-fluid description representing the separate average velocities and their relative motion (see below) is needed to examine this.

This picture has to be modified, however, by allowing for diffusion effects. Kinetic theory is the best way to tackle this. The result is damping of perturbations below diffusion scales which depend on whether or not Hot Dark Matter ('HDM') is present; baryonic fluctuations on small scales are attenuated by photon viscosity and free-streaming of neutrinos [212, 97].

7.6 Other matter

Many other cases have been examined in this GIC formalism. We list them with major references.

7.6.1 Scalar fields

The case of scalar fields is dealt with in a GIC way by Bruni, Ellis and Dunsby [213]. This analysis leads to the usual conserved quantities and theory of growth of inhomogeneities in an inflationary era. A *key element* here is choice of 4-velocity; for small perturbations there is a unique obvious choice, namely, choosing u^a orthogonal to the surfaces on which the scalar field ϕ is constant. The energy-momentum tensor then has the form of a perfect fluid, but with density and pressure depending on both kinetic and potential energy terms for ϕ.

7.6.2 Multi-fluids and imperfect fluids

The physically important case of multi-fluids is dealt with by Dunsby, Bruni and Ellis [214]; for example, enabling modelling of perturbations that include a matter–radiation interaction. The *key element* again is choice of 4-velocity. Each component has a separate 4-velocity $u^a_{(i)}$, and there are various options now for the reference 4-velocity u^a. When linear changes are made in this choice of 4-velocity, the essential effect is to alter the measured momentum in the u^a-frame (Maartens et al [215]). The equations are simplified most by choosing u^a as the centre of mass 4-velocity for the sum of all components, and the 4-velocity $u^a_{(i)}$ of the i-th component as its centre of mass 4-velocity. One must then carefully check the separate momentum and energy equations for each component, as well as for the matter as a whole. These determine the evolution of the $1+3$ covariantly defined *relative velocities*: $V^a_{ij} = u^a_{(i)} - u^a_{(j)}$, and the separate matter densities $\mu_{(i)}$.

Exercise: Establish the equations for the relative velocity and the density inhomogeneities in the 2-fluid case.

The case of imperfect fluids is closely related, and the same issue of choice of 4-velocity arises. As pointed out earlier, it is essential to use realistic equations of state in studying perturbations of an imperfect fluid, such as described by the Müller–Israel–Stewart theory [216, 217] (see Maartens and Triginier [218] for a detailed GIC analysis of such imperfect fluids).

7.6.3 Magnetic fields

These have been examined in a GIC way by Barrow and Tsagas [219], using the $1+3$ covariant splitting of the electromagnetic field and the Maxwell field equations [7].

7.6.4 Newtonian version

A Newtonian version of the analysis can be developed fully in parallel to the relativistic version [220], based on the Newtonian analogue of the approach to cosmology presented in [6], and including derivations of the Newtonian Jeans length and Newtonian formulae for the growth of inhomogeneities.

Exercise: Establish these equations, and hence determine the main differences between the Newtonian and relativistic versions of structure formation.

7.6.5 Alternative gravity

The same GIC approach can be used to analyze higher-derivative gravitational theories; details are in [221].

7.7 Relation to other formalisms

The relation between the GIC approach to perturbations and the very influential Bardeen gauge-invariant formalism [199] has been examined in depth [222]. The essential points are that

* as might be expected, the implications of both approaches for structure formation are the same,

* the implications of the GIC formalism can be worked out in any desired coordinate system, including the Bardeen coordinates (which are incorporated into that approach in an essential way from the start),

* the Bardeen approach is essentially based on the linearised equations, while the GIC starts with the full non-linear equations and linearises them, as explained above. This enables an estimate of the errors involved, and a systematic n-th order approximation scheme,

* the GIC formalism does not use a *non-local* splitting [223] into scalar, vector and tensor modes, and only uses a harmonic splitting (into wavelengths) at a late stage of the analysis; these are both built into the Bardeen approach *ab initio*.

Both approaches have the advantage over gauge approaches that they do not involve gauge modes, and the differential equations are of minimal order needed to characterise the physics of the problem. Many papers on the use of various gauges are rather confused; however, the major paper by Ma and Bertschinger [198] clarifies the relations between important gauges in a clear way. As shown there, the answers obtained for large-scale growth of inhomogeneities is indeed gauge-dependent, and this becomes significant particularly on very large scales. The GIC formalism obviates this problem. However, whatever formalism is used, the issue is how the results relate to observations. The perturbation theory predicts structure growth, gravitational wave emission, gravitational lensing, and background radiation anisotropies. The latter are one of the most important tests of the geometry and physics of perturbed models, and are the topic of the final section.

8 CBR anisotropies

Central to present day cosmology is the study of the information obtainable from measurements of CBR anisotropies. A GIC version of the pioneering Sachs–Wolfe paper [224], based on photon path integration and calculation of the redshift along these paths (cf. the integration in terms of Bardeen's variables by Panek [225]), is given by Dunsby [226] and Challinor and Lasenby [227]. However, a *kinetic theory approach* enables a more in-depth study of the photons' evolution and interactions with the matter inhomogeneities, and so is the dominant way of analyzing CBR anisotropies.

8.1 Covariant kinetic theory

Relativistic kinetic theory (see e.g. [68] and [228]–[230]) provides a self-consistent microscopically based treatment where there is a natural unifying framework in which to deal with a gas of particles in circumstances ranging from hydrodynamical to free-streaming behaviour. The photon gas undergoes a transition from hydrodynamical tight coupling with matter, through the process of decoupling from matter, to non-hydrodynamical free-streaming. This transition is characterised by the evolution of the photon mean free path from effectively zero to effectively infinity. This range of behaviour can appropriately be described by kinetic theory with non-relativistic classical *Thomson scattering* (see e.g. Jackson [231] or Feynman I [232]), and the baryonic matter with which radiation interacts can reasonably be described hydrodynamically during these times.

In this approach, the single-particle *photon distribution function* $f(x^i, p^a)$ over a 7-dimensional phase space [68] represents the number of photons measured in the 3-volume element dV at the event x^i that have momenta in the momentum space volume element π about the momentum p^a through the equation

$$dN = f(x^i, p^a)\,(-p_a u^a)\,\pi\, dV \ , \tag{231}$$

where u^a is the observer's 4-velocity and the redshift factor $(-p_a u^a)$ makes f into a (observer-independent) scalar. The rate of change of f in photon phase space is determined by the *Boltzmann equation*

$$L(f) = C[f] \ , \tag{232}$$

where the *Liouville operator*

$$L(f) = \frac{\partial f}{\partial x^i}\frac{dx^i}{dv} + \frac{\partial f}{\partial p^a}\frac{dp^a}{dv} \tag{233}$$

gives the change of f in parameter distance dv along the geodesics that characterise the particle motions. The *collision term* $C[f]$ determines the rate of change of f due to emission, absorption and scattering processes; it can represent Thomson scattering, binary collisions, etc. Over the period of importance for CBR anisotropies, i.e. considerably after electron–positron annihilation, the average photon energy is much less than the electron rest mass and the electron thermal energy may be neglected so that the quantised Compton interaction between photons and electrons (the dominant interaction between radiation and matter) may reasonably be described in the non-quantised Thomson limit.[36] After decoupling, there is very little interaction between matter and the CBR so we can use the *Liouville equation*:

$$C[f] = 0 \quad \Rightarrow \quad L(f) = 0 \ . \tag{234}$$

The energy-momentum tensor of the photons is

$$T_{R}^{ab} = \int_{T_x} p^a\, p^b\, f\, \pi \ , \tag{235}$$

where π is the volume element in the momentum tangent space T_x. This satisfies the conservation equations (35) and (36), and is part of the total energy-momentum tensor T^{ab} determining the space-time curvature by (1).

Exercise: The same theory applies to particles with non-zero rest-mass. (i) What is the form of the Boltzmann collision term for binary particle collisions? (ii) Show that energy-momentum conservation in individual collisions will lead to conservation of the particle energy-momentum tensor T^{ab}. (iii) Using the appropriate integral definition of entropy, determine an H-theorem for this form of collision. (iv) Under what conditions can equilibrium exist for such a gas of particles, and what is the equilibrium form of the particle distribution function? (See [5].)

[36] The $1+3$ covariant treatment described in this section also neglects polarisation effects.

8.2 Angular harmonic decomposition

In the $1+3$ covariant approach of [233, 234],[37] the photon 4-momentum p^a (where $p_a p^a = 0$) is split relative to an observer moving with 4-velocity u^a as

$$p^a = E\,(u^a + e^a)\ , \qquad e_a e^a = 1\ , \qquad e_a u^a = 0\ , \tag{236}$$

where $E = (-p_a u^a)$ is the photon *energy* and $e^a = E^{-1}\, h^a{}_b\, p^b$ is the photon's *spatial propagation direction*, as measured by a comoving (fundamental) observer. Then the *photon distribution function is decomposed into* $1+3$ *covariant harmonics* via the expansion [233, 238]

$$f(x,p) = f(x,E,e) = \sum_{\ell\geq 0} F_{A_\ell}(x,E)\, e^{A_\ell} = F + F_a e^a + F_{ab} e^a e^b + \cdots = \sum_{\ell \geq 0} F_{A_\ell} e^{\langle A_\ell \rangle}\ , \tag{237}$$

where $e^{A_\ell} \equiv e^{a_1} e^{a_2} \cdots e^{a_\ell}$, and $e^{\langle A_\ell \rangle}$ is the symmetric trace-free part of e^{A_ℓ}. The $1+3$ covariant *distribution function anisotropy multipoles* F_{A_ℓ} are irreducible since they are Projected, Symmetric, and Trace-Free ('PSTF'), i.e.

$$F_{a\cdots b} = F_{\langle a \cdots b\rangle} \quad \Leftrightarrow \quad F_{a\cdots b} = F_{(a\cdots b)}\ ,\ F_{a\cdots b}\, u^b = 0 = F_{a\cdots bc}\, h^{bc} \quad \Rightarrow \quad F_{A_\ell} = F_{\langle A_\ell \rangle}\ .$$

They encode the anisotropy structure of the photon distribution function in the same way as the usual spherical harmonic expansion

$$f = \sum_{\ell \geq 0} \sum_{m=-\ell}^{+\ell} f_\ell^m(x,E)\ Y_\ell^m(e)\ ,$$

but with two major advantages: (a) the F_{A_ℓ} are $1+3$ covariant, and thus independent of any choice of coordinates in momentum space, unlike the f_ℓ^m; (b) F_{A_ℓ} is a rank-ℓ tensor field on space-time for each *fixed* E, and directly determines the ℓ-multipole of radiation anisotropy after integration over E. The multipoles can be recovered from the photon distribution function via

$$F_{A_\ell} = \Delta_\ell^{-1} \int f(x,e)\, e_{\langle A_\ell \rangle}\, d\Omega\ , \qquad \text{with} \qquad \Delta_\ell = 4\pi\, \frac{2^\ell\, (\ell!)^2}{(2\ell+1)!}\ , \tag{238}$$

where $d\Omega = d^2 e$ is a solid angle in momentum space. A further useful identity is [233]

$$\int e^{A_\ell}\, d\Omega = \frac{4\pi}{\ell+1} \begin{cases} 0 & \ell \text{ odd}\ , \\ h^{(a_1 a_2}\, h^{a_3 a_4} \cdots h^{a_{\ell-1} a_\ell)} & \ell \text{ even}\ . \end{cases} \tag{239}$$

The first three multipoles determine the radiation energy-momentum tensor, which is, from (235) and (14),

$$T_R^{ab}(x) = \int p^a\, p^b\, f(x,p)\, d^3 p = \mu_R\, (u^a\, u^b + \tfrac{1}{3}\, h^{ab}) + 2\, q_R^{(a}\, u^{b)} + \pi_R^{ab}\ , \tag{240}$$

where $d^3 p = E\, dE\, d\Omega$ is the covariant momentum space volume element on the future null cone at the event x. It follows from (237) and (240) that the dynamical quantities of the radiation (in the u^a-frame) are:

$$\mu_R = 4\pi \int_0^\infty E^3\, F\, dE\ , \qquad q_R^a = \frac{4\pi}{3} \int_0^\infty E^3\, F^a\, dE\ , \qquad \pi_R^{ab} = \frac{8\pi}{15} \int_0^\infty E^3\, F^{ab}\, dE\ . \tag{241}$$

[37] Based on the Ph.D. Thesis of R Treciokas, Cambridge University, 1972, combined with the covariant formalism of F A E Pirani [235] (see also K S Thorne [236]) by G F R Ellis when in Hamburg (1st Institute of Theoretical Physics) in 1972; see also [69, 70]. A similar formalism has been developed by M L Wilson [237].

We extend these dynamical quantities to all multipole orders by defining the 1+3 covariant *brightness anisotropy multipoles*[38] [238]

$$\Pi_{A_\ell} = \int_0^\infty E^3\, F_{A_\ell}\, dE = \Pi_{\langle A_\ell \rangle}\ , \tag{242}$$

so that $\Pi = \mu_{_R}/4\pi$, $\Pi^a = 3\, q_{_R}^a/4\pi$ and $\Pi^{ab} = 15\, \pi_{_R}^{ab}/8\pi$.

Writing the Boltzmann equation (232) in the form

$$\frac{df}{dv} \equiv p^a\, \mathbf{e}_a(f) - \Gamma^a{}_{bc}\, p^b\, p^c\, \frac{\partial f}{\partial p^a} = C[f]\ , \tag{243}$$

the *collision term is also decomposed into* $1+3$ *covariant harmonics*:

$$C[f] = \sum_{\ell \geq 0} b_{A_\ell}(x,E)\, e^{A_\ell} = b + b_a e^a + b_{ab} e^a e^b + \cdots\ , \tag{244}$$

where the $1+3$ covariant *scattering multipoles* $b_{A_\ell} = b_{\langle A_\ell \rangle}$ encode irreducible properties of the particle interactions. Then the Boltzmann equation is equivalent to an *infinite hierarchy of* $1+3$ *covariant multipole equations*

$$L_{A_\ell}(x,E) = b_{A_\ell}[\, F_{A_m}(x,E)\,]\ ,$$

where $L_{A_\ell} = L_{\langle A_\ell \rangle}$ are the anisotropy multipoles of df/dv, and will be given in the next subsection. These multipole equations are tensor field equations on space-time for each value of the photon energy E (but note that energy changes along each photon path). Given the solutions $F_{A_\ell}(x,E)$ of the equations, the relation (237) then determines the full photon distribution function $f(x,E,e)$ as a scalar field over phase space.

8.3 Non-linear $1+3$ covariant multipole equations

The full Boltzmann equation in photon phase space contains *more* information than necessary to analyze radiation anisotropies in an inhomogeneous universe. For that purpose, when the radiation is close to black body, we do not require the full spectral behaviour of the photon distribution function multipoles, but only the *energy-integrated multipoles*. The monopole leads to the average (all-sky) temperature, while the higher-order multipoles determine the temperature anisotropies. The GIC definition of the *average temperature* T is given according to the *Stefan–Boltzmann law* by

$$\mu_{_R}(x) = 4\pi \int E^3\, F(x,E)\, dE = r\, T^4(x)\ , \tag{245}$$

where r is the radiation constant. If f is close to a Planck distribution, then T is the thermal black body average temperature. But note that *no* notion of background temperature is involved in this definition. There is an all-sky average implied in (245). *Fluctuations* across the sky are measured by energy-integrating the higher-order multipoles (a precise definition is given below), i.e. the fluctuations are determined by the Π_{A_ℓ} ($\ell \geq 1$) defined in (242).

The form of $C[f]$ in (244) shows that $1+3$ covariant equations for the temperature fluctuations arise from decomposing the *energy-integrated Boltzmann equation*

$$\int E^2\, \frac{df}{dv}\, dE = \int E^2\, C[f]\, dE \tag{246}$$

[38] Because photons are massless, we do not need the complexity of the moment definitions used in [233]. From now on, all energy integrals will be understood to be over the range $0 \leq E \leq \infty$.

into 1+3 covariant multipoles. We begin with the right-hand side, which requires the 1+3 covariant form of the Thomson scattering term. Defining the $1+3$ covariant *energy-integrated scattering multipoles*

$$K_{A_\ell} = \int E^2\, b_{A_\ell}\, dE = K_{\langle A_\ell \rangle}\ , \tag{247}$$

we find that [215]

$$K = n_{_E}\sigma_{_T}\left[\ \tfrac{4}{3}\,\Pi v_{_B}^2 - \tfrac{1}{3}\,\Pi^a v_{_B\,a}\ \right] + \mathcal{O}[3]\ , \tag{248}$$

$$K^a = -\,n_{_E}\sigma_{_T}\left[\ \Pi^a - 4\,\Pi v_{_B}^a - \tfrac{2}{5}\,\Pi^{ab} v_{_B\,b}\ \right] + \mathcal{O}[3]\ , \tag{249}$$

$$K^{ab} = -\,n_{_E}\sigma_{_T}\left[\ \tfrac{9}{10}\,\Pi^{ab} - \tfrac{1}{2}\,\Pi^{\langle a} v_{_B}^{b\rangle} - \tfrac{3}{7}\,\Pi^{abc} v_{_B\,c}\ - 3\Pi v_{_B}^{\langle a} v_{_B}^{b\rangle}\ \right] + \mathcal{O}[3]\ , \tag{250}$$

$$K^{abc} = -\,n_{_E}\sigma_{_T}\left[\ \Pi^{abc} - \tfrac{3}{2}\,\Pi^{\langle ab} v_{_B}^{c\rangle} - \tfrac{4}{9}\,\Pi^{abcd} v_{_B\,d}\ \right] + \mathcal{O}[3]\ , \tag{251}$$

and, for $\ell > 3$,

$$K^{A_\ell} = -\,n_{_E}\sigma_{_T}\left[\ \Pi^{A_\ell} - \Pi^{\langle A_{\ell-1}} v_{_B}^{a_\ell\rangle} - \left(\frac{\ell+1}{2\ell+3}\right)\Pi^{A_\ell a} v_{_B\,a}\ \right] + \mathcal{O}[3]\ , \tag{252}$$

where the expansion is in terms of the peculiar velocity v_B^a of the baryons relative to the reference frame u^a. Parameters are the *free electron number density* $n_{_E}$ and the *Thomson scattering cross section* $\sigma_{_T}$, the latter being proportional to the square of the classical electron radius [231, 232]. The first three multipoles are affected by Thomson scattering differently than the higher-order multipoles.

Equations (248)–(252), derived in [215], are a non-linear generalisation of the results given by Challinor and Lasenby [239]. They show the *coupling of baryonic bulk velocity to the radiation multipoles, arising from* $1+3$ *covariant non-linear effects in Thomson scattering.* If we linearise fully, i.e. neglect all terms containing $v_{_B}$ except the $\mu_{_R} v_{_B}^a$ term in the dipole K^a, which is first-order, then our equations reduce to those in [239]. The generalised non-linear equations apply to the analysis of $1+3$ covariant second-order effects on an FLRW background, to first-order effects on a spatially homogeneous but anisotropic background, and more generally, to any situation where the baryonic frame is in non-relativistic motion relative to the fundamental u^a-frame.

Next we require the anisotropy multipoles L_{A_ℓ} of df/dv. These can be read off for photons directly from the general expressions in [233], which are exact, $1+3$ covariant, and also include the case of massive particles. For clarity and completeness, we outline an alternative, $1+3$ covariant derivation (the derivation in [233] uses tetrads). We require the rate of change of the photon energy along null geodesics, given by [69, 70]

$$\frac{dE}{dv} = -\ \left[\ \tfrac{1}{3}\,\Theta + (\dot{u}_a e^a) + (\sigma_{ab} e^a e^b)\ \right] E^2\ , \tag{253}$$

which follows directly from $p^b \nabla_b p^a = 0$ with (236) and (11). Then

$$\begin{aligned}
\frac{d}{dv}\left[\ F_{a_1\cdots a_\ell}(x,E)\, e^{a_1}\cdots e^{a_\ell}\ \right] &= \frac{d}{dv}\left[\ E^{-\ell}\, F_{a_1\cdots a_\ell}(x,E)\, p^{a_1}\cdots p^{a_\ell}\ \right] \\
&= E\left\{\ \left[\ \tfrac{1}{3}\,\Theta + \dot{u}_b e^b + \sigma_{bc} e^b e^c\ \right]\left(\ell\, F_{a_1\cdots a_\ell} - E\, F'_{a_1\cdots a_\ell}\right) e^{a_1}\cdots e^{a_\ell} \right. \\
&\quad \left. +\ (u^{a_1} + e^{a_1})\cdots(u^{a_\ell} + e^{a_\ell})\ \left[\ \dot{F}_{a_1\cdots a_\ell} + e^b \nabla_b F_{a_1\cdots a_\ell}\ \right]\ \right\}\ ,
\end{aligned}$$

where a prime denotes $\partial/\partial E$. The first term is readily put into irreducible PSTF form using identities in [233], p470. In the second term, when the round brackets are expanded, only those terms with at most one u^{a_r} survive, and

$$u^a \dot{F}_{a\cdots} = -\,\dot{u}^a F_{a\cdots}\ , \qquad u^b \nabla^a F_{b\cdots} = -\left(\ \tfrac{1}{3}\,\Theta\, h^{ab} + \sigma^{ab} + \eta^{abc}\,\omega_c\ \right) F_{b\cdots}\ .$$

Thus, the $1+3$ covariant *anisotropy multipoles* L_{A_ℓ} of df/dv are

$$\begin{aligned} E^{-1}\,L_{A_\ell} &= \dot{F}_{\langle A_\ell\rangle} - \tfrac{1}{3}\,\Theta\,E\,F'_{A_\ell} + \tilde{\nabla}_{\langle a_\ell}F_{A_{\ell-1}\rangle} + \frac{(\ell+1)}{(2\ell+3)}\,\tilde{\nabla}^b F_{A_\ell b} \\ &\quad - \frac{(\ell+1)}{(2\ell+3)}\,E^{-(\ell+1)}\left[E^{\ell+2}\,F_{A_\ell b}\right]'\dot{u}^b - E^\ell\left[E^{1-\ell}\,F_{\langle A_{\ell-1}}\right]'\dot{u}_{a_\ell\rangle} \\ &\quad + \ell\,\omega^b\,\eta_{bc\langle a_\ell}\,F_{A_{\ell-1}\rangle}{}^c - \frac{(\ell+1)(\ell+2)}{(2\ell+3)(2\ell+5)}\,E^{-(\ell+2)}\left[E^{\ell+3}\,F_{A_\ell bc}\right]'\sigma^{bc} \\ &\quad - \frac{2\ell}{(2\ell+3)}\,E^{-1/2}\left[E^{3/2}\,F_{b\langle A_{\ell-1}}\right]'\sigma_{a_\ell\rangle}{}^b - E^{\ell-1}\left[E^{2-\ell}\,F_{\langle A_{\ell-2}}\right]'\sigma_{a_{\ell-1}a_\ell\rangle} \qquad (254) \\ &= E^{-1}\,b_{A_\ell}\ . \end{aligned}$$

This regains the result of [233] (equation (4.12)) in the massless case. The form given here benefits from the streamlined version of the $1+3$ covariant formalism. We re-iterate that this result is exact and holds for any photon or (massless) neutrino distribution in *any* space-time.

We now multiply (254) by E^3 and integrate over all photon energies, using integration by parts and the fact that $E^n\,F_{a\cdots}\to 0$ as $E\to\infty$ for any positive n. We obtain the evolution equations that determine the brightness anisotropies multipoles Π_{A_ℓ}:

$$\begin{aligned} &\dot{\Pi}_{\langle A_\ell\rangle} + \tfrac{4}{3}\,\Theta\,\Pi_{A_\ell} + \tilde{\nabla}_{\langle a_\ell}\Pi_{A_{\ell-1}\rangle} + \frac{(\ell+1)}{(2\ell+3)}\,\tilde{\nabla}^b\Pi_{A_\ell b} \qquad (255) \\ &- \frac{(\ell+1)(\ell-2)}{(2\ell+3)}\,\dot{u}^b\,\Pi_{A_\ell b} + (\ell+3)\,\dot{u}_{\langle a_\ell}\,\Pi_{A_{\ell-1}\rangle} + \ell\,\omega^b\,\eta_{bc\langle a_\ell}\,\Pi_{A_{\ell-1}\rangle}{}^c \\ &- \frac{(\ell-1)(\ell+1)(\ell+2)}{(2\ell+3)(2\ell+5)}\,\sigma^{bc}\,\Pi_{A_\ell bc} + \frac{5\ell}{(2\ell+3)}\,\sigma^b{}_{\langle a_\ell}\,\Pi_{A_{\ell-1}\rangle b} - (\ell+2)\,\sigma_{\langle a_\ell a_{\ell-1}}\,\Pi_{A_{\ell-2}\rangle} = K_{A_\ell}\ . \end{aligned}$$

Once again, this is an exact result, and it holds also for any collision term, i.e. any K_{A_ℓ}. For decoupled neutrinos, we have $K_{_N}^{A_\ell}=0$ in this equation. For photons undergoing Thomson scattering, the right-hand side of (255) is given by (252), which is exact in the kinematical and dynamical quantities, but first-order in the relative baryonic velocity. The equations (252) and (255) thus constitute a non-linear generalisation of the FLRW-linearised case given by Challinor and Lasenby [239].

The monopole and dipole of equation (255) give the evolution equations for the energy and momentum densities:

$$\dot{\Pi} + \tfrac{4}{3}\,\Theta\,\Pi + \tfrac{1}{3}\,\tilde{\nabla}_a\Pi^a + \tfrac{2}{3}\,\dot{u}_a\Pi^a + \tfrac{2}{15}\,\sigma_{ab}\Pi^{ab} = K\ , \qquad (256)$$

$$\begin{aligned} &\dot{\Pi}^{\langle a\rangle} + \tfrac{4}{3}\,\Theta\,\Pi^a + \tilde{\nabla}^a\Pi + \tfrac{2}{5}\,\tilde{\nabla}_b\Pi^{ab} \\ &+ \tfrac{2}{5}\,\dot{u}_b\,\Pi^{ab} + 4\,\Pi\,\dot{u}^a + \eta^{abc}\,\omega_b\,\Pi_c + \sigma^{ab}\,\Pi_b = K^a\ , \qquad (257) \end{aligned}$$

(these are thus the equations giving the divergence of T^{ab} in (235)). For *photons*, the equations are:

$$\dot{\mu}_{_R} + \tfrac{4}{3}\,\Theta\,\mu_{_R} + \tilde{\nabla}_a q^a_{_R} + 2\,\dot{u}_a q^a_{_R} + \sigma_{ab}\pi^{ab}_{_R} = n_{_E}\sigma_{_T}\,(\,\tfrac{4}{3}\,\mu_{_R}\,v^2_{_B} - q^a_{_R}v_{_B\,a}\,) + \mathcal{O}[3]\ , \qquad (258)$$

$$\begin{aligned} &\dot{q}^{\langle a\rangle}_{_R} + \tfrac{4}{3}\,\Theta\,q^a_{_R} + \tfrac{4}{3}\,\mu_{_R}\,\dot{u}^a + \tfrac{1}{3}\,\tilde{\nabla}^a\mu_{_R} + \tilde{\nabla}_b\pi^{ab}_{_R} \\ &+ \sigma^a{}_b\,q^b_{_R} + \eta^{abc}\,\omega_b\,q_{_R c} + \dot{u}_b\,\pi^{ab}_{_R} = n_{_E}\sigma_{_T}\,(\,\tfrac{4}{3}\,\mu_{_R}\,v^a_{_B} - q^a_{_R} + \pi^{ab}_{_R}\,v_{_B b}\,) + \mathcal{O}[3]\ , \qquad (259) \end{aligned}$$

(the present versions of the fluid energy and momentum conservation equations (35) and (36)).

The non-linear dynamical equations are completed by the energy-integrated Boltzmann multipole equations. For *photons*, the quadrupole evolution equation is

$$\dot{\pi}^{\langle ab\rangle}_{_R} + \tfrac{4}{3}\,\Theta\,\pi^{ab}_{_R} + \tfrac{8}{15}\,\mu_{_R}\,\sigma^{ab} + \tfrac{2}{5}\,\tilde{\nabla}^{\langle a}q^{b\rangle}_{_R} + \frac{8\pi}{35}\,\tilde{\nabla}_c\Pi^{abc}$$

$$
\begin{aligned}
&+\ 2\,\dot{u}^{\langle a}\,q_{_R}^{b\rangle}+2\,\omega^c\,\eta_{cd}{}^{\langle a}\,\pi_{_R}^{b\rangle d}+\tfrac{2}{7}\,\sigma_c{}^{\langle a}\,\pi_{_R}^{b\rangle c}-\frac{32\pi}{315}\,\sigma_{cd}\,\Pi^{abcd} \\
&=-\,n_{_E}\sigma_{_T}\,(\,\tfrac{9}{10}\,\pi_{_R}^{ab}-\tfrac{1}{5}\,q_{_R}^{\langle a}\,v_{_B}^{b\rangle}-\frac{8\pi}{35}\,\Pi^{abc}\,v_{_B\,c}-\tfrac{2}{5}\,\rho_{_R}\,v_{_B}^{\langle a}\,v_{_B}^{b\rangle})+\mathcal{O}[3]\ , \qquad (260)
\end{aligned}
$$

(a fluid description gives no analogue of this and the following equations). The higher-order multipoles ($\ell > 3$) evolve according to

$$
\begin{aligned}
&\dot{\Pi}^{\langle A_\ell\rangle}+\tfrac{4}{3}\,\Theta\,\Pi^{A_\ell}+\tilde{\nabla}^{\langle a_\ell}\Pi^{A_{\ell-1}\rangle}+\frac{(\ell+1)}{(2\ell+3)}\,\tilde{\nabla}_b\Pi^{A_\ell b} \\
&-\ \frac{(\ell+1)(\ell-2)}{(2\ell+3)}\,\dot{u}_b\,\Pi^{A_\ell b}+(\ell+3)\,\dot{u}^{\langle a_\ell}\,\Pi^{A_{\ell-1}\rangle}+\ell\,\omega^b\,\eta_{bc}{}^{\langle a_\ell}\,\Pi^{A_{\ell-1}\rangle c} \\
&-\ \frac{(\ell-1)(\ell+1)(\ell+2)}{(2\ell+3)(2\ell+5)}\,\sigma_{bc}\,\Pi^{A_\ell bc}+\frac{5\ell}{(2\ell+3)}\,\sigma_b{}^{\langle a_\ell}\,\Pi^{A_{\ell-1}\rangle b}-(\ell+2)\,\sigma^{\langle a_\ell a_{\ell-1}}\,\Pi^{A_{\ell-2}\rangle} \\
&=-\,n_{_E}\,\sigma_{_T}\left[\,\Pi^{A_\ell}-\Pi^{\langle A_{\ell-1}}\,v_{_B}^{a_\ell\rangle}-\left(\frac{\ell+1}{2\ell+3}\right)\Pi^{A_\ell a}\,v_{_B\,a}\,\right]+\mathcal{O}[3]\ . \qquad (261)
\end{aligned}
$$

For $\ell = 3$, the $\Pi^{\langle A_{\ell-1}}v_{_B}^{a_\ell\rangle}$ term on the right-hand side of equation (261) must be multiplied by $\frac{3}{2}$. For neutrinos, the equations are the same except without the Thomson scattering terms. Note that these equations link angular multipoles of order $\ell-2$, $\ell-1$, ℓ, $\ell+1$, $\ell+2$, i.e they link *five* successive harmonic terms. This is the source of the *harmonic mixing* that occurs as the radiation propagates.

These equations for the radiation (and neutrino) multipoles generalise the equations given by Challinor and Lasenby [239], to which they reduce when we remove all terms $\mathcal{O}(\epsilon v_{_I})$ and $\mathcal{O}(\epsilon^2)$. In this case, on introducing a *FLRW-linearisation*, there is major simplification of the equations:

$$
\begin{aligned}
\dot{\mu}_{_R}+\tfrac{4}{3}\,\Theta\,\mu_{_R}+\tilde{\nabla}_a q_{_R}^a &\approx 0\ , \qquad (262)\\
\dot{q}_{_R}^{\langle a\rangle}+\tfrac{4}{3}\,\Theta\,q_{_R}^a+\tfrac{4}{3}\,\mu_{_R}\,\dot{u}^a+\tfrac{1}{3}\,\tilde{\nabla}^a\mu_{_R}+\tilde{\nabla}_b\pi_{_R}^{ab} &\approx n_{_E}\sigma_{_T}\,(\,\tfrac{4}{3}\,\mu_{_R}\,v_{_B}^a-q_{_R}^a\,)\ , \qquad (263)\\
\dot{\pi}_{_R}^{\langle ab\rangle}+\tfrac{4}{3}\,\Theta\,\pi_{_R}^{ab}+\tfrac{8}{15}\,\mu_{_R}\,\sigma^{ab}+\tfrac{2}{5}\,\tilde{\nabla}^{\langle a}q_{_R}^{b\rangle}+\frac{8\pi}{35}\,\tilde{\nabla}_c\Pi^{abc} &\approx -\,\tfrac{9}{10}\,n_{_E}\,\sigma_{_T}\,\pi_{_R}^{ab}\ , \qquad (264)
\end{aligned}
$$

and, for $\ell \geq 3$,

$$
\dot{\Pi}^{\langle A_\ell\rangle}+\tfrac{4}{3}\,\Theta\,\Pi^{A_\ell}+\tilde{\nabla}^{\langle a_\ell}\Pi^{A_{\ell-1}\rangle}+\frac{(\ell+1)}{(2\ell+3)}\,\tilde{\nabla}_b\Pi^{A_\ell b}\approx-\,n_{_E}\,\sigma_{_T}\,\Pi^{A_\ell}\ . \qquad (265)
$$

Note that these equations now link only angular multipoles of order $\ell-1$, ℓ, $\ell+1$, i.e they link *three* successive terms. This is a major qualitative difference from the full non-linear equations.

8.4 Temperature anisotropy multipoles

Finally, we return to the definition of temperature anisotropies. As noted above, these are determined by the Π_{A_ℓ}. We define the *temperature fluctuation* $\tau(x,e)$ via the directional temperature which is determined by the directional bolometric brightness:

$$
T(x,e)=T(x)\,[\,1+\tau(x,e)\,]=\left[\,\frac{4\pi}{r}\int E^3\,f(x,E,e)\,dE\,\right]^{1/4}\ . \qquad (266)
$$

This is a GIC definition which is also exact. We can rewrite it explicitly in terms of the Π_{A_ℓ}:

$$
\tau(x,e)=\left[\,1+\frac{4\pi}{\mu_{_R}}\sum_{\ell\geq1}\Pi_{A_\ell}e^{A_\ell}\,\right]^{1/4}-1=\tau_a e^a+\tau_{ab}e^a e^b+\cdots=\sum_{\ell\geq1}\tau_{A_\ell}(x)\,e^{A_\ell}\ . \qquad (267)
$$

In principle, we can extract the $1+3$ covariant irreducible PSTF *temperature anisotropy multipoles* $\tau_{A_\ell} = \tau_{\langle A_\ell \rangle}$ by using the inversion (238):

$$\tau_{A_\ell}(x) = \Delta_\ell^{-1} \int \tau(x,e)\, e_{\langle A_\ell \rangle}\, d\Omega \ . \tag{268}$$

In the almost-FLRW case, when τ is $\mathcal{O}[1]$, we regain from (267) the linearised definition given in [240]:

$$\tau_{A_\ell} \approx \left(\frac{\pi}{\mu_{_R}}\right) \Pi_{A_\ell} \ , \tag{269}$$

where $\ell \geq 1$. In particular, the dipole and quadrupole are

$$\tau^a \approx \frac{3\, q^a_{_R}}{4\, \mu_{_R}} \quad \text{and} \quad \tau^{ab} \approx \frac{15\, \pi^{ab}_{_R}}{2\, \mu_{_R}} \ . \tag{270}$$

We can normalise the dynamical brightness anisotropy multipoles Π_{A_ℓ} of the radiation to define the dimensionless $1+3$ covariant *brightness temperature anisotropy multipoles* ($\ell \geq 1$)

$$\mathcal{T}_{A_\ell} = \left(\frac{\pi}{r\, T^4}\right) \Pi_{A_\ell} \approx \tau_{A_\ell} \ . \tag{271}$$

Thus, the $\mathcal{T}_{A_\ell} = \mathcal{T}_{\langle A_\ell \rangle}$ are equal to the temperature anisotropy multipoles plus non-linear corrections. In terms of these quantities, the *hierarchy of radiation multipoles becomes*:

$$\begin{aligned}
\frac{\dot{T}}{T} &= -\tfrac{1}{3}\,\Theta - \tfrac{1}{3}\,\tilde{\nabla}_a \mathcal{T}^a - \tfrac{4}{3}\,\mathcal{T}^a\, \frac{\tilde{\nabla}_a T}{T} - \tfrac{2}{3}\,\dot{u}_a\, \mathcal{T}^a - \tfrac{2}{15}\,\sigma_{ab}\, \mathcal{T}^{ab} \\
&\quad + \tfrac{1}{3}\, n_{_E}\sigma_{_T}\, v_{_B\,a}\,(\, v^a_{_B} - \mathcal{T}^a\,) + \mathcal{O}[3] \ , \qquad (272) \\
\dot{\mathcal{T}}^{\langle a \rangle} &= -4\left(\frac{\dot{T}}{T} + \tfrac{1}{3}\,\Theta\right) \mathcal{T}^a - \frac{\tilde{\nabla}^a T}{T} - \dot{u}^a - \tfrac{2}{5}\,\tilde{\nabla}_b \mathcal{T}^{ab} + n_{_E}\sigma_{_T}\,(\, v^a_{_B} - \mathcal{T}^a\,) \\
&\quad + \tfrac{2}{5}\, n_{_E}\sigma_{_T}\, \mathcal{T}^{ab}\, v_{_B\,b} - \sigma^a{}_b\, \mathcal{T}^b - \tfrac{2}{5}\,\dot{u}_b\, \mathcal{T}^{ab} - \eta^{abc}\,\omega_b\, \mathcal{T}_c - \tfrac{8}{5}\,\mathcal{T}^{ab}\, \frac{\tilde{\nabla}_b T}{T} + \mathcal{O}[3] \ , \qquad (273) \\
\dot{\mathcal{T}}^{\langle ab \rangle} &= -4\left(\frac{\dot{T}}{T} + \tfrac{1}{3}\,\Theta\right) \mathcal{T}^{ab} - \sigma^{ab} - \tilde{\nabla}^{\langle a}\, \mathcal{T}^{b\rangle} - \tfrac{3}{7}\,\tilde{\nabla}_c \mathcal{T}^{abc} - \tfrac{9}{10}\, n_{_E}\sigma_{_T}\, \mathcal{T}^{ab} \\
&\quad + n_{_E}\sigma_{_T}\,(\, \tfrac{1}{2}\,\mathcal{T}^{\langle a}\, v^{b\rangle}_{_B} + \tfrac{3}{7}\,\mathcal{T}^{abc}\, v_{_B\,c} + \tfrac{3}{4} v^{\langle a}_{_B} v^{b\rangle}_{_B}) - 5\,\dot{u}^{\langle a}\, \mathcal{T}^{b\rangle} - \tfrac{4}{21}\,\sigma_{cd}\, \mathcal{T}^{abcd} \\
&\quad - 2\,\omega^c\, \eta_{cd}{}^{\langle a}\, \mathcal{T}^{b\rangle d} - \tfrac{10}{7}\,\sigma_c{}^{\langle a}\, \mathcal{T}^{b\rangle c} - \tfrac{12}{7}\,\mathcal{T}^{abc}\, \frac{\tilde{\nabla}_c T}{T} + \mathcal{O}[3] \ , \qquad (274)
\end{aligned}$$

and, for $\ell > 3$:

$$\begin{aligned}
\dot{\mathcal{T}}^{\langle A_\ell \rangle} &= -4\left(\frac{\dot{T}}{T} + \tfrac{1}{3}\,\Theta\right) \mathcal{T}^{A_\ell} - \tilde{\nabla}^{\langle a_\ell} \mathcal{T}^{A_{\ell-1}\rangle} - \frac{(\ell+1)}{(2\ell+3)}\, \tilde{\nabla}_b \mathcal{T}^{A_\ell b} - n_{_E}\sigma_{_T}\, \mathcal{T}^{A_\ell} \\
&\quad + n_{_E}\sigma_{_T}\left[\, \mathcal{T}^{\langle A_{\ell-1}}\, v^{a_\ell\rangle}_{_B} + \left(\frac{\ell+1}{2\ell+3}\right) \mathcal{T}^{A_\ell b}\, v_{_B\,b}\,\right] + \frac{(\ell+1)\,(\ell-2)}{(2\ell+3)}\, \dot{u}_b\, \mathcal{T}^{A_\ell b} \\
&\quad - (\ell+3)\,\dot{u}^{\langle a_\ell}\, \mathcal{T}^{A_{\ell-1}\rangle} - \ell\,\omega^b\, \eta_{bc}{}^{\langle a_\ell}\, \mathcal{T}^{A_{\ell-1}\rangle c} + (\ell+2)\,\sigma^{\langle a_\ell a_{\ell-1}}\, \mathcal{T}^{A_{\ell-2}\rangle} \\
&\quad + \frac{(\ell-1)\,(\ell+1)\,(\ell+2)}{(2\ell+3)\,(2\ell+5)}\, \sigma_{bc}\, \mathcal{T}^{A_\ell bc} - \frac{5\ell}{(2\ell+3)}\, \sigma_b{}^{\langle a_\ell}\, \mathcal{T}^{A_{\ell-1}\rangle b} \\
&\quad - 4\,\frac{(\ell+1)}{(2\ell+3)}\, \mathcal{T}^{A_\ell b}\, \frac{\tilde{\nabla}_b T}{T} + \mathcal{O}[3] \ . \qquad (275)
\end{aligned}$$

For $\ell = 3$, the Thomson scattering term $\mathcal{T}^{\langle A_{\ell-1}}v_{_B}^{a_\ell\rangle}$ must be multiplied by a factor $\frac{3}{2}$.

The $1+3$ covariant non-linear multipole equations given in this form show more clearly the evolution of temperature anisotropies (including the monopole, i.e. the average temperature T), in general linking five successive harmonics. Although the $\mathcal{T}_{A_\ell}$ only determine the actual temperature fluctuations τ_{A_ℓ} to linear order, they are a useful dimensionless measure of anisotropy. Furthermore, equations (272)–(275) apply as the evolution equations for the temperature anisotropy multipoles when the radiation anisotropy is small (i.e. $\mathcal{T}_{A_\ell} \approx \tau_{A_\ell}$), but the space-time inhomogeneity and anisotropy are not restricted. This includes the particular case of small CBR anisotropies in general Bianchi universes, or in perturbed Bianchi universes.

FLRW-linearisation, i.e. the case when only first-order effects relative to the FLRW limit are considered, reduces the above equations to the linearised form, generically linking three successive harmonics:

$$\frac{\dot{T}}{T} \approx -\tfrac{1}{3}\,\Theta - \tfrac{1}{3}\,\tilde{\nabla}_a\tau^a \;, \tag{276}$$

$$\dot{\tau}^{\langle a\rangle} \approx -\frac{\tilde{\nabla}^a T}{T} - \dot{u}^a - \tfrac{2}{5}\,\tilde{\nabla}_b\tau^{ab} + n_{_E}\sigma_{_T}\,(\,v_{_B}^a - \tau^a\,)\;, \tag{277}$$

$$\dot{\tau}^{\langle ab\rangle} \approx -\sigma^{ab} - \tilde{\nabla}^{\langle a}\tau^{b\rangle} - \tfrac{3}{7}\,\tilde{\nabla}_c\tau^{abc} - \tfrac{9}{10}\,n_{_E}\sigma_{_T}\,\tau^{ab}\;, \tag{278}$$

and, for $\ell \geq 3$:

$$\dot{\tau}^{\langle A_\ell\rangle} \approx -\tilde{\nabla}^{\langle a_\ell}\tau^{A_{\ell-1}\rangle} - \frac{(\ell+1)}{(2\ell+3)}\,\tilde{\nabla}_b\tau^{A_\ell b} - n_{_E}\sigma_{_T}\,\tau^{A_\ell}\;. \tag{279}$$

These are GIC multipole equations leading to the Fourier mode formulation of the energy-integrated Boltzmann equations used in the standard literature (see e.g. [241] and the references therein) when they are decomposed into spatial harmonics and associated wavelengths.[39] This then allows examination of diffusion effects, which are wavelength-dependent.

These linearised equations, together with the linearised equations governing the kinematical and gravitational quantities, may be $1+3$ *covariantly split* into scalar, vector and tensor modes, as described in [222, 239]. The modes can then be expanded in $1+3$ covariant eigentensors of the comoving Laplacian, and the Fourier coefficients obey ordinary differential equations.

Exercise: Determine the resulting hierarchy of mode equations. Show from these equations that there is a wavelength λ_S such that for shorter wavelengths the perturbations are heavily damped (the physical reason is photon diffusion).

Numerical integration of these equations are performed for scalar modes by Challinor and Lasenby [239], with further analytic results given in [242]. These are the Sachs–Wolfe family of integrations, of fundamental importance in determining CBR anisotropies, as discussed in many places; see e.g. Hu and Sugiyama [241], Hu and White [243] (the GIC version will be given in the series of papers by Gebbie, Maartens, Dunsby and Ellis). However, they assume an almost-FLRW geometry. We turn now to justifying that assumption.

8.5 Almost-EGS-Theorem and its applications

One of our most important understandings of the nature of the universe is that at recent times it is well-represented by the standard spatially homogeneous and isotropic FLRW models. The basic reason for this belief is the observed high degree of isotropy of the CBR, together with a fundamental result of Ehlers, Geren and Sachs [69] (hereafter 'EGS'), taken nowadays (see e.g. [3]) to establish

[39] They contain the free-streaming subcase for $n_{_E} = 0$.

that the universe is almost-FLRW at recent times (i.e. since decoupling of matter and radiation).

The EGS programme can be summarized as follows: Using (a) the measured high isotropy of the CBR at our space-time position, and (b) the Copernican assumption that we are not at a privileged position in the universe, the aim is to deduce that the universe is accurately FLRW. EGS gave an exact theorem of this kind: if a family of freely-falling observers measure self-gravitating background radiation to be everywhere *exactly* isotropic in the case of non-interacting matter and radiation, then the universe is *exactly* FLRW. This is taken to establish the desired conclusion, in view of the measured near-isotropy of the radiation at our space-time location. However, of course the CBR is not exactly isotropic. Generally, we want to show stability of arguments we use [244]; in this case, we wish to do so by showing the EGS result remains *nearly* true if the radiation is *nearly* isotropic, thus providing the foundation on which further analysis, such as that in the famous Sachs–Wolfe paper [224], are based. More precisely, we aim to prove the following theorem [245].

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

This description is intended to represent the situation since decoupling to the present day. The pressure-free matter represents the galaxies on which fundamental observers live, who measure the radiation to be almost isotropic.

8.5.1 Assumptions

In detail, we consider matter and radiation in a space-time region U. In our application, we consider U to be the region within and near our past light cone from decoupling to the present day (this is the observable space-time region where we would like to prove the universe is almost-FLRW; before decoupling a different analysis is needed, for collisions dominate there, and we do not have sufficient data to comment on the situation far from our light cone). Our assumptions are that, in the region considered,

(1) The EFE are satisfied, with the total energy-momentum tensor T_{ab} composed of *non-interacting* matter and radiation components:

$$T^{ab} = T^{ab}_{M} + T^{ab}_{R} \ , \quad \nabla_b T^{ab}_{M} = 0 \ , \quad \nabla_b T^{ab}_{R} = 0 \ . \tag{280}$$

The independent conservation equations express the decoupling of matter from radiation; the matter energy-momentum tensor is $T^{ab}_{M} = \rho\, u^a\, u^b$, ($\mu_{M} = \rho$ is the matter density) and, at each point, the radiation energy-momentum tensor is $T^{ab}_{R} = \int f\, p^a\, p^b\, \pi$, where π is the momentum space volume element. The only non-zero energy-momentum tensor contribution from the matter, relative to the 4-velocity u^a, is the energy density; so the total energy density is

$$\mu = \mu_{R} + \rho = \mathcal{O}[0] \ , \tag{281}$$

while all other energy-momentum tensor components are simply equal to the radiation contributions. Without confusion we will *omit* a subscript 'R' on these terms.

(2) The matter 4-velocity field u^a is geodesic and expanding:

$$\dot{u}^a = 0 \ , \quad \Theta = \mathcal{O}[0] > 0 \ , \tag{282}$$

(the first requirement in fact follows from momentum conservation for the pressure-free matter; see (38)) . The assumption of an expanding universe is essential to what follows, for otherwise there are counter-examples to the result [246].

(3) The radiation obeys the Liouville equation (234): $L(f) = 0$, where L is the Liouville operator. This means that there is *no* entropy production, so that q^a and π_{ab} are *not* dissipative quantities, but measure the extent to which f deviates from isotropy.

(4) Relative to u^a — i.e. for all matter-comoving observers — the photon distribution function f is almost isotropic everywhere in the region U. Formally, in this region, $F = F(x, E)$ and its time derivatives are zero-order, while $F_{A_\ell} = F_{A_\ell}(x, E)$ plus their time and space derivatives are at most first-order for $\ell > 0$. In brief:

$$F,\, \dot{F} = \mathcal{O}[0]\ , \qquad F_{A_\ell},\, \dot{F}_{\langle A_\ell \rangle},\, \tilde{\nabla}_a F_{A_\ell} = \mathcal{O}[1]\ , \tag{283}$$

(and we assume all higher derivatives of F_{A_ℓ} are also $\mathcal{O}[1]$ for $\ell > 0$).

It immediately follows, under reasonable assumptions about the phase space integrals of the harmonic components, that the scalar moments are zero-order but the tensor moments and their derivatives are first-order. We assume the conditions required to ensure this are true, i.e. we additionally suppose

$$\mu_{\scriptscriptstyle R},\, p_{\scriptscriptstyle R} = \mathcal{O}[0]\ , \qquad q_a,\, \dot{q}_{\langle a \rangle},\, \tilde{\nabla}_a q_b = \mathcal{O}[1]\ , \qquad \pi_{ab},\, \dot{\pi}_{\langle ab \rangle},\, \tilde{\nabla}_a \pi_{bc} = \mathcal{O}[1] \tag{284}$$

are satisfied as a consequence (283), and that the same holds for the higher time and space derivatives of q^a, π_{ab} and for all higher-order moments (specifically, those defined in (292) below). The *first aim* now is to show that

(a) the kinematical quantities and the Weyl tensor are almost-FLRW, i.e. σ_{ab} and ω^a are first-order, which then implies that E_{ab} and H_{ab} are also first-order. The *second aim* is to show that then,

(b) there are coordinates such that the metric tensor takes a perturbed RW form.

8.5.2 Proving almost-FLRW kinematics

Through an appropriate integration over momentum space, the zeroth harmonic of the Liouville equation (234) gives the energy conservation equation for the radiation (cf. (258)),

$$\dot{\mu}_{\scriptscriptstyle R} + \tfrac{4}{3}\,\Theta\,\mu_{\scriptscriptstyle R} + \tilde{\nabla}_a q^a + \sigma_{ab}\pi^{ab} = 0\ , \tag{285}$$

while the first harmonic gives the momentum conservation equation (cf. (259)),

$$\dot{q}^{\langle a \rangle} + \tfrac{4}{3}\,\Theta\, q^a + \tfrac{1}{3}\,\tilde{\nabla}^a \mu_{\scriptscriptstyle R} + \tilde{\nabla}_b \pi^{ab} + \sigma^a{}_b\, q^b + \eta^{abc}\,\omega_b\, q_c = 0\ , \tag{286}$$

which implies by (284),

$$\tilde{\nabla}_a \mu_{\scriptscriptstyle R} = \mathcal{O}[1]\ . \tag{287}$$

Taking spatial derivatives of (286), the same result follows for the higher spatial derivatives of μ_R; in particular, its second derivatives are at most first-order.

Now the definition of the $\tilde{\nabla}$-derivative leads to the identity (27), giving

$$(\tilde{\nabla}_a \tilde{\nabla}_b - \tilde{\nabla}_b \tilde{\nabla}_a)\mu_{\scriptscriptstyle R} = 2\,\eta_{abc}\,\omega^c\,\dot{\mu}_{\scriptscriptstyle R}\ . \tag{288}$$

In an expanding universe, by (284), the energy conservation equation (285) for the radiation shows

$$\dot{\mu}_{\scriptscriptstyle R} + \tfrac{4}{3}\,\Theta\,\mu_{\scriptscriptstyle R} = \mathcal{O}[1]\ . \tag{289}$$

Thus, because the universe is expanding, $\dot{\mu}_R$ is zero-order. However, the left-hand side of (288) is first-order; consequently, by the higher-derivative version of (287),

$$\omega^a = \mathcal{O}[1] \ . \tag{290}$$

The second harmonic of the Liouville equation (234) leads, after an appropriate integration over momentum space, to an evolution equation for the anisotropic stress tensor π_{ab} (which is the present version of (260)):

$$\dot{\pi}^{\langle ab\rangle} + \tfrac{4}{3}\,\Theta\,\pi^{ab} + \tfrac{8}{15}\,\mu_R\,\sigma^{ab} + \tfrac{2}{5}\,\tilde{\nabla}^{\langle a}q^{b\rangle} + J^{ab} + 2\,\omega^c\,\eta_{cd}{}^{\langle a}\,\pi^{b\rangle d} + \tfrac{2}{7}\,\sigma_c{}^{\langle a}\,\pi^{b\rangle c} - \frac{32\pi}{315}\,\sigma_{cd}\,\Pi^{abcd} = 0 \ , \tag{291}$$

where (cf. (242))

$$J^{ab} = \frac{8\pi}{35}\int_0^\infty E^3\,\tilde{\nabla}_c F^{abc}\,dE \ , \qquad \Pi^{abcd} = \int_0^\infty E^3\,F^{abcd}\,dE \ . \tag{292}$$

Consequently, by (290), (283) and (284),

$$\sigma_{ab} = \mathcal{O}[1] \ , \tag{293}$$

and (on taking derivatives of the above equation) the same is true for its time and space derivatives. Then, to first order, the evolution equations become

$$\dot{\mu}_R + \tfrac{4}{3}\,\Theta\,\mu_R + \tilde{\nabla}_a q^a \approx 0 \ , \tag{294}$$

$$\dot{q}^{\langle a\rangle} + \tfrac{4}{3}\,\Theta\,q^a + \tfrac{1}{3}\,\tilde{\nabla}^a\mu_R + \tilde{\nabla}_b\pi^{ab} \approx 0 \ , \tag{295}$$

$$\dot{\pi}^{\langle ab\rangle} + \tfrac{4}{3}\,\Theta\,\pi^{ab} + \tfrac{8}{15}\,\mu_R\,\sigma^{ab} + \tfrac{2}{5}\,\tilde{\nabla}^{\langle a}q^{b\rangle} + J^{ab} \approx 0 \ , \tag{296}$$

showing the equations can *only* close at first order *if* we can argue that $J^{ab} = \mathcal{O}[2]$. *Any* such approximation must be done with utmost care because of the *radiation multipole truncation theorem* (see below).

It now follows from the shear propagation equation (31) and the H_{ab}-equation (34) that all the Weyl tensor components are also at most first-order:

$$E_{ab} = \mathcal{O}[1] \ , \qquad H_{ab} = \mathcal{O}[1] \ . \tag{297}$$

Consequently, the $(\operatorname{div} E)$-equation (45) shows that

$$X_a \equiv \tilde{\nabla}_a\mu = \mathcal{O}[1] \quad \Rightarrow \quad \tilde{\nabla}_a\rho = \mathcal{O}[1] \ , \tag{298}$$

by (281). Finally, the (0α)-equation (32), or spatial derivatives of (289), show that

$$Z_a \equiv \tilde{\nabla}_a\Theta = \mathcal{O}[1] \ . \tag{299}$$

This establishes that the kinematical quantities for the matter flow and the Weyl tensor are almost-FLRW: all the quantities that vanish in the FLRW case are at most first-order here. Thus, the only zero-order $1+3$ covariantly defined quantities are those that are non-vanishing in FLRW universes.

8.5.3 Proving almost-FLRW dynamics

It follows that the zero-order equations governing the dynamics are just those of a FLRW universe. Because the kinematical quantities and the Weyl tensor are precisely those we expect in a perturbed FLRW universe, we can linearise the $1+3$ covariant equations about the FLRW values in the usual way, in which the background model will — as we have just seen — obey the usual FLRW equations. Hence, the usual linearised $1+3$ covariant FLRW perturbation analyses can be applied [197, 222], leading to the usual results for growth of inhomogeneities in almost-FLRW universes.

8.5.4 Finding an almost-RW metric

Given that the kinematical quantities and the Weyl tensor take an almost-FLRW form, the *key issue* in proving existence of an almost-RW metric is choice of a time function to use as a cosmic time (in the realistic, inhomogeneous universe model). The problem is that although we have shown the vorticity will be small, it will in general not be zero; hence, there will be no time surfaces orthogonal to the matter flow lines [5, 6]. The problem can equivalently be viewed as the need to find a vorticity-free (that is, hypersurface-orthogonal) congruence of curves to use as a kinematical reference frame, which, if possible, one would like to also be geodesic. As we cannot assume the matter 4-velocity fulfills this condition, we need to introduce another congruence of curves, say $\hat{u}^a$, that *is* hypersurface-orthogonal and that does not differ too greatly from u^a, so the matter is moving slowly (non-relativistically) relative to that frame. Then we have to change to that frame, using a 'hat' to denote its kinematical quantities. (See [245].)

8.5.5 Result

This gives the result we want; we have shown that freely propagating almost-isotropic background radiation everywhere in a region U implies the universe is an almost-FLRW universe in that region. Thus we have proved the stability of the Ehlers, Geren and Sachs [69] result. This result is the foundation for the important analysis of Sachs and Wolfe [224] and all related analyses, determining the effect of inhomogeneities on the CBR by integration from last scattering till today in an almost-FLRW universe, for these papers *start off* with the assumption that the universe *is* almost-FLRW since decoupling, and build on that basis.

It should also be noted that the assumption that radiation is isotropic about distant observers can be partially checked by testing how close the CBR spectrum is to black body in those directions where we detect the Sunyaev–Zel'dovich effect [104] (see also Goodman [247]). Since that effect mixes and scatters incoming radiation, substantial radiation anisotropies relative to those clusters of galaxies inducing the effect would result in a significant distortion of the outgoing spectrum.

Exercise: How good are the limits we might obtain on the CBR anisotropy at distance points on our past null cone by this method? Are there any other ways one can obtain such limits?

8.6 Other CBR calculations

One can extend the above analysis to obtain model-independent limits on the CBR anisotropy (Maartens et al [240]). However, there are also a whole series of model-dependent analyses available.

8.6.1 Sachs–Wolfe and related effects

We will not pursue here the issue of Sachs–Wolfe [224] type calculations, integrating up the red-shift from the surface of last scattering to the observer and so determining the CBR anisotropy, and, hence, cosmological parameters [248], because they are so extensively covered in the literature. The GIC version is covered by Challinor and Lasenby [239] and developed in depth in the series of papers by Gebbie, Dunsby, Maartens and Ellis [215, 238, 242]. However, a few comments are in order.

One can approach this topic $1+3$ covariantly by a direct generalisation of the photon type of calculation (see [226, 227]), or by use of the kinetic theory formalism developed above, as in the papers just cited. Two things need to be very carefully considered. These are,

(i) The issue of putting *correct limits* on the Sachs–Wolfe integral. This integral needs to be taken to the *real* surface of last scattering, not the background surface of last scattering (which is fictitious). If this is not done, the results may be gauge-dependent. One should place the limits on this integral properly, which can be done quite easily by calculating when the optical depth due to Thomson scattering is unity [225, 249].

(ii) In order to obtain *solutions*, one has somehow to cut off the harmonic series, for otherwise there is an infinite regress whereby higher-order harmonics determine the evolution of lower-order ones, as is clear from the multipole equations above. However, truncation of harmonics in a kinetic theory description needs to be approached with great caution, because of the following theorem for collision-free radiation [234]:

> **Radiation Multipole Truncation Theorem**: In the exact Einstein–Liouville theory, truncation of the angular harmonics seen by a family of geodesic observers *at any order whatever* leads to the vanishing of the shear of the family of observers; hence, this can only occur in highly restricted spaces.

The proof is given in [234]. This is an exact result of the full theory. However, *it remains true in the linear theory* if the linearisation is done carefully, *even though the term responsible for the conclusion is second-order* and so is omitted in most linearised calculations.

The point is that in that equation, at the relevant order, there are *only* second-order terms in the equation, so one cannot drop the terms responsible for this conclusion (one can only drop them in an equation where there are also linear terms, so the second-order terms are negligible compared with the linear terms; that is not the case here). Thus, a proper justification for the acceptability of some effective truncation procedure relies on showing how a specific approximation procedure gives an acceptable approximation to the results of the exact theory, despite this disjuncture.

Exercise: Give such a justification.

8.6.2 Other models

As mentioned previously in these lectures, there is an extensive literature on the CBR anisotropy:

- in Bianchi (exact spatially homogeneous) models; see e.g. [190]–[192],
- in Swiss-Cheese (exact inhomogeneous) models; see e.g. [161] and [142],
- in small universes (FLRW models with compact spatial topology that closes up on a small scale, so that we have seen round the universe already since decoupling [89]); see e.g. [250] and [92].

These provide important parametrized sets of models that one can use to test and exploit the restrictions the CBR observations place on alternative universe models that are not necessarily close to the standard models at early or late times, but are still potentially compatible with observations, and deserve full exploration.

In each case, whether carrying out Sachs–Wolfe type or more specific model calculations, we get tighter limits than in the almost-EGS case, but they are also more model-dependent. In each case they can be carried out using the $1+3$ covariantly defined harmonic variables and can be used to put limits on the $1+3$ covariantly defined quantities that are the theme of this article.

The overall conclusion is that — given the Copernican assumption underlying use of the almost-FLRW models — the high degree of observed CBR isotropy confirms use of these models for the observable universe and puts limits on the amplitude of anisotropic and inhomogeneous modes in this region (i.e. within the horizon and since decoupling). Other models remain viable as representations of the universe at earlier and later times and outside the horizon.

9 Conclusion and open issues

The sections above have presented the $1+3$ covariant variables and equations, together with the tetrad equations, a series of interesting exact cosmological models, and a systematic procedure for obtaining approximate almost-FLRW models and examining observations in these models. It has been shown how a number of anisotropic and inhomogeneous models are useful in studying observational limits on the geometry of the real universe, and how interesting dynamical and observational issues arise in considering these models; indeed issues such as whether inflation in fact succeeds in making the universe isotropic or not cannot be tackled without examining such models [251].

9.1 Conclusion

The standard model of cosmology is vindicated in the following sense (cf. [94, 41]):

- The universe is expanding and evolving, as evidenced by the (*magnitude, redshift*)–relation for galaxies and other sources, together with number count observations and a broad compatibility of age estimates;
- It started from a hot big bang early stage which evolved to the presently observed state, this early era being evidenced by the CBR spectrum and concordance with nucleosynthesis observations;
- Given a[40] Copernican assumption, the high degree of isotropy of the CBR and other observations support an almost-FLRW (nearly homogeneous and isotropic) model within the observable region (inside our past light cone) since decoupling and probably back to nucleosynthesis times;
- There are globally inhomogeneous spherically symmetric models that are also compatible with the observations if we do not introduce the Copernican assumption [144], and there are inhomogeneous and anisotropic modes that suggest the universe is not almost-FLRW at very early and very late times [51, 196], and on very large scales (outside the horizon) [66].

The issue of cosmological parameters may be resolved in the next decade due to the flood of new data coming in — from deep space number counts and redshift measurements, observations of supernovae in distant galaxies, measurement of strong and weak gravitational lensing, and CBR anisotropy measurements (see Coles and Ellis [41] for a discussion).

The further extensions of this model proposed from the particle physics side are at present mainly compatible but not yet as compelling from an observational viewpoint, namely:

- An early inflationary era helps resolve some philosophical puzzles about the structure of the universe related specifically to the homogeneity and isotropy of the universe [64]–[66], but the link to particle physics will not be compelling until a specific inflaton candidate is identified;
- Structure formation initiated by quantum fluctuations at very early stages in an inflationary model are a very attractive idea, with the proposal of a CDM-dominated late universe and associated predictions for the CBR anisotropy giving a strong link to observations that may be confirmed in the coming decade; however, theoretical details of this scheme, and particularly the associated issues of biasing and the normalisation of matter to radiation, are still to be resolved;
- The density of matter in the universe is probably below the critical value predicted by the majority of inflationary models [41]; compatibility with inflation may be preserved by either moving to low density inflationary models[41] or by confirmation of a cosmological constant that

[40] Necessarily philosophically based [150].
[41] For an early proposal, see [252] and [73].

is dynamically dominant at the present time;[42] the latter proposal needs testing relative to number counts, lensing observations, and structure formation scenarios;

- The vibrant variety of pre-inflationary proposals involve a huge extrapolation of presently known physics to way beyond the testable domain and, given their variety, do not as yet give a compelling unique view of that era.

Hence, it is suggested [41] that these proposals should not be regarded as part of the standard model but rather as interesting avenues under investigation.

9.2 Open issues

Considered from a broader viewpoint, substantial issues remain unresolved. Among them are,

(1) The Newtonian theory of cosmology is not yet adequately resolved. Newtonian theory is only a good theory of gravitation when it is a good approximation to General Relativity; obtaining this limit in non-linear cosmological situations raises a series of questions and issues that still need clarifying [24], particularly relating to boundary conditions in realistic Newtonian cosmological models.

(2) We have some understanding of how the evolution of families of inhomogeneous models relates to that of families of higher symmetry models. It has been indicated that a skeleton of higher symmetry models seems to guide the evolution of lower symmetry models in the state space (the space of space-times) [51]. This relation needs further elucidation. Also anisotropic and inhomogeneous inflationary models are relatively little explored and problems remain [184, 254];

(3) We need to find a suitable measure of probability in the full space of cosmological space-times, and in its involutive subspaces. The requirement is a natural measure that is plausible. Some progress has been made in the FLRW subcase [255]–[258], but even here it is not definitive. Closely related to this is the issue of the stability of the results we derive from cosmological modelling [244];

(4) We need to be able to relate descriptions of the same space-time on different scales of description. This leads to the issue of averaging and the resulting effective (polarization) contributions to the energy-momentum tensor, arising because averaging does not commute with calculating the field equations for a given metric [88], and we do not have a good procedure for fitting a FLRW model to a lumpy realistic universe model [259]. It includes the issue (discussed briefly above) 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.

(5) Related to this is the question of definition of entropy for gravitating systems in general,[43] and cosmological models in particular. This may be expected to imply a coarse-graining in general, and so is strongly related to the averaging question. 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 arrow of time problem [260], which in turn relates to a series of further issues concerned with the effect on local physics of boundary conditions at the beginning of the universe (see e.g. [28]).

(6) One can approach relating a model cosmology to astronomical observations by a strictly observational approach [98, 101], as opposed to the more usual model-based approach as envisaged in the main sections above and indeed in most texts on cosmology. Intermediate between them is a best-fitting approach [259]. Use of such a fitting approach is probably the best way to tackle

[42] But way below that predicted by present field theory [253].

[43] In the case of black holes, there is a highly developed theory; but there is no definition for a general gravitational field; see e.g. [260].

modelling the real universe [261], but it is not yet well-developed.

(7) Finally, underlying all these issues is the series of problems arising because of the uniqueness of the universe, which is what particularly gives cosmology its unique character and underlies the special problems in cosmological modelling and application of probability theory to cosmology [262]. Proposals to deal with this by considering an ensemble of universes realized in one or other of a number of possible ways are in fact untestable and, hence, of a metaphysical rather than physical nature; but this needs further exploration.

There is interesting work still to be done in all these areas. It will be important in tackling these issues to as far as possible use gauge-invariant and 1 + 3 covariant methods, because coordinate methods can be misleading.

Acknowledgements

GFRE and HvE thank Roy Maartens, John Wainwright, Peter Dunsby, Tim Gebbie, and Claes Uggla for work done together, some of which is presented in these lectures, Charles Hellaby for useful comments and references, Tim Gebbie for helpful comments, and Anthony Challinor and Roy Maartens for corrections to some equations. We are grateful to the Foundation for Research and Development (South Africa) and the University of Cape Town for financial support. HvE acknowledges the support by a grant from the Deutsche Forschungsgemeinschaft (Germany).

References

[1] R d'Inverno, *Introducing Einstein's Relativity* (Oxford Univerity Press, Oxford, 1992).

[2] R M Wald, *General Relativity* (University of Chicago Press, Chicago, 1984).

[3] S W Hawking and G F R Ellis, *The Large Scale Structure of Space-Time* (Cambridge University Press, Cambridge, 1973).

[4] H Stephani, *General Relativity* (Cambridge University Press, Cambridge, 1990).

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

[6] G F R Ellis, in *General Relativity and Cosmology, Proceedings of the XLVII Enrico Fermi Summer School*, Ed. R K Sachs (Academic Press, New York, 1971).

[7] G F R Ellis, in *Cargèse Lectures in Physics, Vol. 6*, Ed. E Schatzman (Gordon and Breach, New York, 1973).

[8] A Krasiński, *Physics in an Inhomogeneous Universe* (Cambridge University Press, Cambridge, 1996).

[9] J A Wheeler, *Einsteins Vision* (Springer, Berlin, 1968).

[10] W Kundt and M Trümper, *Akad. Wiss. Lit. Mainz, Abhandl. Math.-Nat. Kl.* **12**, 1 (1962).
M Trümper, *J. Math. Phys.* **6**, 584 (1965).

[11] R Maartens, *Phys. Rev.* D **55**, 463 (1997).

[12] M B Ribeiro and A Y Miguelote, *Braz. Journ. Phys.* **28**, 132 (1998).

[13] R Maartens, G F R Ellis and S T C Siklos, *Class. Quantum Grav.* **14**, 1927 (1997).

[14] T Levi-Civita, *Math. Ann.* **97**, 291 (1926).

[15] F A E Pirani, *Acta Phys. Polon.* **15**, 389 (1956).

[16] F A E Pirani, *Phys. Rev.* **105**, 1089 (1957).

[17] P Szekeres, *J. Math. Phys.* **6**, 1387 (1965).

[18] P Szekeres, *J. Math. Phys.* **7**, 751 (1966).

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

[20] K Gödel, *Proc. Int. Cong. Math.* (Am. Math. Soc.) **175**, (1952).

[21] A Raychaudhuri, *Phys. Rev.* **98**, 1123 (1955).

[22] S W Hawking, *Astrophys. J.* **145**, 544 (1966).

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

[24] H van Elst and G F R Ellis, *Class. Quantum Grav.* **15**, 3545 (1998).

[25] H Friedrich, *Phys. Rev.* D **57**, 2317 (1998).

[26] H van Elst and G F R Ellis, *Phys. Rev.* D **59**, 024013 (1999).

[27] T Velden, *Diplomarbeit*, Universität Bielefeld/Albert–Einstein–Institut, Potsdam, (1997).

[28] G F R Ellis and D W Sciama, in *General Relativity* (Synge Festschrift), Ed. L O'Raifeartaigh (Oxford University Press, Oxford, 1972).

[29] G F R Ellis, in *Gravitation and Cosmology* (Proceedings of ICGC95), Eds. S Dhurandhar and T Padmanabhan (Kluwer, Dordrecht, 1997).

[30] G F R Ellis, *J. Math. Phys.* **8**, 1171 (1967).

[31] C B Collins, *J. Math. Phys.* **26**, 2009 (1985).

[32] J M Senovilla, C Sopuerta and P Szekeres, *Gen. Rel. Grav.* **30**, 389 (1998).

[33] S Matarrese, O Pantano and D Saez, *Phys. Rev. Lett.* **72**, 320 (1994).

[34] H van Elst, C Uggla, W M Lesame, G F R Ellis and R Maartens, *Class. Quantum Grav.* **14**, 1151 (1997).

[35] C F Sopuerta, R Maartens, G F R Ellis and W M Lesame, Nonperturbative gravito-magnetic fields are not transverse. Preprint *gr-qc/9809085*, (1998).

[36] F J Dyson, *Sci. Am.*, September (1971).

[37] See C W Misner, K S Thorne and J A Wheeler, *Gravitation* (Freeman and Co., New York, 1973), Exercise 15.2 (p. 382), and references given there.

[38] M S Madsen and G F R Ellis, *Mon. Not. Roy. Astr. Soc.* **234**, 67 (1988).

[39] S W Hawking and R Penrose, *Proc. R. Soc. London A* **314**, 529 (1970).

[40] F J Tipler, C J S Clarke and G F R Ellis, in *General Relativity and Gravitation: One Hundred Years after the Birth of Albert Einstein, Vol. 2*, Ed. A Held (Plenum Press, New York, 1980).

[41] P Coles and G F R Ellis, *The Density of Matter in the Universe* (Cambridge University Press, Cambridge, 1997).

[42] A Raychaudhuri, *Zeits. f. Astroph.* **43**, 161 (1957).

[43] O Heckmann, *Astronom. Journ.* **66**, 205 (1961).

[44] O Heckmann and E Schücking, *Zeits. f. Astrophys.* **38**, 95 (1955).

[45] O Heckmann and E Schücking, *Zeits. f. Astrophys.* **40**, 81 (1956).

[46] J Ehlers and T Buchert, On the Newtonian limit of the Weyl-tensor. Preprint AEI/MPI für Gravitationsphysik, (1996).

[47] H van Elst and C Uggla, *Class. Quantum Grav.* **14**, 2673 (1997).

[48] M A H MacCallum, in *Cargèse Lectures in Physics, Vol. 6*, Ed. E Schatzman (Gordon and Breach, New York, 1973).

[49] G F R Ellis and M A H MacCallum, *Commun. Math. Phys.* **12**, 108 (1969).

[50] J M Stewart and G F R Ellis, *J. Math. Phys.* **9**, 1072 (1968).

[51] J Wainwright and G F R Ellis (Eds.), *Dynamical Systems in Cosmology* (Cambridge University Press, Cambridge, 1997).

[52] G F R Ellis, in *Vth Brazilian School on Cosmology and Gravitation*, Ed. M Novello (World Scientific, Singapore, 1987).

[53] G F R Ellis and H van Elst, in *On Einstein's Path — Essays in Honor of Engelbert Schücking*, Ed. A Harvey, (Springer, New York, 1999). Available as preprint *gr-qc/9709060.*

[54] G F R Ellis, *Gen. Rel. Grav.* **2**, 7 (1971).

[55] G F R Ellis et al, in *Dahlem Workshop Report ES19, The Evolution of the Universe*, Eds. S Gottlöber and G Börner (Wiley, New York, 1996).

[56] W E Harris, R P Durrell, M J Pierce, and J Seckers, *Nature* **395**, 45 (1998).
B F Madore et al, *Nature* **395**, 47 (1998).

[57] S Perlmutter et al, *Nature* **391**, 51 (1998).

[58] H P Robertson, *Rev. Mod. Phys.* **5**, 62 (1933).

[59] R Stabell and S Refsdal, *Mon. Not. Roy. Astr. Soc.* **132**, 379 (1966).

[60] E Schrödinger, *Expanding Universes* (Cambridge University Press, Cambridge, 1956).

[61] S Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).

[62] R M Barnett et al, *Rev. Mod. Phys.* **68**, 708 (1996).

[63] D N Schramm and M S Turner, *Rev. Mod. Phys.* **70**, 303 (1998).

[64] A H Guth, *Phys. Rev.* D **23**, 347 (1981).

[65] E W Kolb and M S Turner, *The Early Universe* (Wiley, New York, 1990).

[66] A D Linde, *Particle Physics and Inflationary Cosmology* (Harwood, Chur, 1990).

[67] G F R Ellis and M Madsen, *Class. Quantum Grav.* **8**, 667 (1991).

[68] J Ehlers, in *General Relativity and Cosmology, Proceedings of the XLVII Enrico Fermi Summer School*, Ed. R K Sachs (Academic Press, New York, 1971).

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

[70] R Treciokas and G F R Ellis, *Commun. Math. Phys.* **23**, 1 (1971).

[71] G F R Ellis, D R Matravers and R Treciokas, *Gen. Rel. Grav.* **15**, 931 (1983).

[72] J Ehlers and W Rindler, *Mon. Not. Roy. Astr. Soc.* **238**, 503 (1989).

[73] G F R Ellis, in *Gravitation, Proceedings of the Banff Summer Research Institute on Gravitation*, Eds. R Mann and P Wesson (World Scientific, Singapore, 1991).

[74] G B Field, H Arp and J N Bahcall, *The Redshift Controversy* (Benjamin, Reading, MA, 1973).

[75] D R Matravers and A M Aziz, *Mon. Not. Astr. Soc. S.A.* **47**, 124 (1988).

[76] W Mattig, *Astr. Nach.* **284**, 109 (1958).

[77] A Sandage, *Astrophys. J.* **133**, 355 (1961).

[78] G F R Ellis and T Rothman, *Am. J. Phys.* **61**, 883 (1993).

[79] G F R Ellis and G Tivon, *Observatory* **105**, 189 (1985).

[80] G F R Ellis and J J Perry, *Mon. Not. Roy. Astr. Soc.* **187**, 357 (1979).

[81] I M H Etherington, *Phil. Mag.* **15**, 761 (1933).

[82] M J Disney, *Nature* **263**, 573 (1976).

[83] G F R Ellis, J J Perry and A Sievers, *Astronom. Journ.* **89**, 1124 (1984).

[84] T Rothman and G F R Ellis, *Observatory* **107**, 24 (1987).

[85] W Rindler, *Mon. Not. Roy. Astr. Soc.* **116**, 662 (1956).

[86] R Penrose, in *Relativity Groups and Topology*, Eds. C M DeWitt and B S DeWitt (Gordon and Breach, New York, 1963).

[87] G F R Ellis and W R Stoeger, *Class. Quantum Grav.* **5**, 207 (1988).

[88] G F R Ellis, in *General Relativity and Gravitation*, Eds. B Bertotti, F de Felice and A Pascolini (Reidel, Dordrecht, 1984).

[89] G F R Ellis and G Schreiber, *Phys. Lett.* **115A**, 97 (1986).

[90] B F Roukema, *Mon. Not. Roy. Astr. Soc.* **283**, 1147 (1996).

[91] B Roukema and A C Edge, *Mon. Not. Roy. Astr. Soc.* **292**, 105 (1997).

[92] N J Cornish, D N Spergel and G D Starkman, *Class. Quantum Grav.* **15**, 2657 (1998).

[93] G Ellis and R Tavakol, *Class. Quantum Grav.* **11**, 675 (1994).

[94] P J E Peebles, D N Schramm, E L Turner and R G Kron, *Nature* **352**, 769 (1991).

[95] C W Misner, *Astrophys. J.* **151**, 431 (1968).

[96] C W Misner, *Phys. Rev. Lett.* **22**, 1071 (1969).

[97] M J Rees, *Perspectives in Astrophysical Cosmology* (Cambridge University Press, Cambridge, 1995).

[98] J Kristian and R K Sachs, *Astrophys. J.* **143**, 379 (1966).

[99] P Schneider, J Ehlers and E Falco, *Gravitational Lenses* (Springer, Berlin, 1992).

[100] D E Holz and R M Wald, *Phys. Rev.* D **58**, 063501 (1998).

[101] G F R Ellis, S D Nel, W Stoeger, R Maartens and A P Whitman, *Phys. Rep.* **124**, 315 (1985).

[102] G F R Ellis, in *Galaxies and the Young Universe*, Eds. H von Hippelein, K Meisenheimer and J H Roser (Springer, Berlin, 1995).

[103] W Stoeger (Ed.), *Theory and Observational Limits in Cosmology*, Specola Vaticana (The Vatican Observatory), (1987).

[104] R A Sunyaev and Ya B Zel'dovich, *Astrophys. Space Sci.* **9**, 368 (1970).

[105] L P Eisenhart, *Continuous Groups of Transformations* (Princeton University Press, Princeton, 1933). Reprinted: (Dover, New York, 1961).

[106] P M Cohn, *Lie Algebras* (Cambride University Press, Cambridge, 1961).

[107] H van Elst and G F R Ellis, *Class. Quantum Grav.* **13**, 1099 (1996).

[108] K Gödel, *Rev. Mod. Phys.* **21**, 447 (1949).

[109] G F R Ellis, in *Gödel 96*, Lecture Notes in Logic 6, Ed. P Hajek (Springer, Berlin, 1996).

[110] I Oszvath, *J. Math. Phys.* **6**, 590 (1965); *J. Math. Phys.* **9**, 2871 (1970).

[111] I Oszvath and E Schücking, *Nature* **193**, 1168 (1962).

[112] H Bondi, *Cosmology* (Cambridge University Press, Cambridge, 1960).

[113] A S Kompaneets and A S Chernov, *Sov. Phys. JETP* **20**, 1303 (1965).

[114] R Kantowski and R K Sachs, *J. Math. Phys.* **7**, 443 (1966).

[115] C B Collins, *J. Math. Phys.* **18**, 2116 (1977).

[116] A R King and G F R Ellis, *Commun. Math. Phys.* **31**, 209 (1973).

[117] G Lemaître, *Ann. Soc. Sci. Bruxelles I* **A 53**, 51 (1933).
Translation: G Lemaître, *Gen. Rel. Grav.* **29**, 641 (1997).

[118] R C Tolman, *Proc. Nat. Acad. Sci. U.S.* **20**, 69 (1934).

[119] H Bondi, *Mon. Not. Roy. Astr. Soc.* **107**, 410 (1947).

[120] P Szekeres, *Commun. Math. Phys.* **41**, 55 (1975).

[121] P Szekeres, *Phys. Rev.* D **12**, 2941 (1975).

[122] S W Goode and J Wainwright, *Phys. Rev.* D **26**, 3315 (1982).

[123] H Stephani, *Class. Quantum Grav.* **4**, 125 (1987).

[124] A Krasiński, *Gen. Rel. Grav.* **15**, 673 (1983).

[125] D Kramer, H Stephani, M A H MacCallum and E Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge University Press, Cambridge, 1980).

[126] O Heckmann and E Schücking, in *Gravitation*, Ed. L Witten (Wiley, New York, 1962).

[127] K Tomita, *Prog. Theor. Phys.* **40**, 264 (1968).

[128] M A H MacCallum and G F R Ellis, *Commun. Math. Phys.* **19**, 31 (1970).

[129] K S Thorne, *Astrophys. J.* **148**, 51 (1967).

[130] C W Misner, *Phys. Rev. Lett.* **19**, 533 (1967).

[131] M J Rees, *Astrophys. J.* **153**, L1 (1968).

[132] C W Misner, K S Thorne and J A Wheeler, *Gravitation* (Freeman and Co., New York, 1973).

[133] W B Bonnor, *Mon. Not. Roy. Astr. Soc.* **159**, 261 (1972).

[134] W B Bonnor, *Mon. Not. Roy. Astr. Soc.* **167**, 55 (1974).

[135] Ya B Zel'dovich and L P Grishchuk, *Mon. Not. Roy. Astr. Soc.* **207**, 23P (1984).

[136] F C Mena and R Tavakol, *Class. Quantum Grav.* **16**, 435 (1999).

[137] J Silk, *Astron. Astrophys.* **59**, 53 (1977).

[138] R Kantowski, *Astrophys. J.* **155**, 1023 (1969).

[139] C Hellaby and K Lake, *Astrophys. J.* **282**, 1 (1984).

[140] D J Raine and E G Thomas, *Mon. Not. Roy. Astr. Soc.* **195**, 649 (1981).

[141] B Paczynski and T Piran, *Astrophys. J.* **364**, 341 (1990).

[142] M Panek, *Astrophys. J.* **388**, 225 (1992).

[143] W B Bonnor and G F R Ellis, *Mon. Not. Roy. Astr. Soc.* **218**, 605 (1986).

[144] N Mustapha, C Hellaby and G F R Ellis, *Mon. Not. Roy. Astr. Soc.* **292**, 817 (1998).

[145] N Mustapha, B A C C Bassett, C Hellaby and G F R Ellis, *Class. Quantum Grav.* **15**, 2363 (1998).

[146] R Maartens, N P Humphreys, D R Matravers and W Stoeger, *Class. Quantum Grav.* **13**, 253 (1996).

[147] N P Humphreys, R Maartens and D R Matravers, Regular spherical dust spacetimes, Preprint *gr-qc/9804023*, (1998).

[148] C Hellaby, *Gen. Rel. Grav.* **20**, 1203 (1988).

[149] C Hellaby and K Lake, *Astrophys. J.* **290**, 381 (1985); *Astrophys. J.* **300**, 461 (1986).

[150] G F R Ellis, *Qu. Journ. Roy. Astr. Soc.* **16**, 245 (1975).

[151] W Israel, *Nuovo Cimento* **44**, 1 (1966); Corrections: **48**B, N2 (1967).

[152] G F R Ellis and M Jaklitsch, *Astrophys. J.* **346**, 601 (1989).

[153] J Traschen, *Phys. Rev.* D **29**, 1563 (1984);
Phys. Rev. D **31**, 283 (1985).

[154] A Einstein and E G Strauss, *Rev. Mod. Phys.* **17**, 120 (1945);
Rev. Mod. Phys. **18**, 148 (1945).

[155] E Schücking, *Z. Physik* **137**, 595 (1954).

[156] K Lake, *Astrophys. J.* **240**, 744 (1980);
Astrophys. J. **242**, 1238 (1980).

[157] C Hellaby and K Lake, *Astrophys. J.* **251**, 429 (1981).

[158] C Hellaby and K Lake, *Astrophys. Lett.* **23**, 81 (1983).

[159] R Kantowski, *Astrophys. J.* **155**, 89 (1969).

[160] M J Rees and D W Sciama, *Nature* **217**, 511 (1968).

[161] C C Dyer, *Mon. Not. Roy. Astr. Soc.* **175**, 429 (1976).

[162] A Meszaros and Z Molner, *Astrophys. J.* **470**, 49 (1996).

[163] W B Bonnor and A Chamorro, *Astrophys. J.* **361**, 21 (1990);
Astrophys. J. **378**, 461 (1991).

[164] A Chamorro, *Astrophys. J.* **383**, 51 (1991).

[165] K Lake, in *Vth Brazilian School of Cosmology and Gravitation*, Ed. M Novello (World Scientific, Singapore, 1987).

[166] M Harwit, *Astrophys. J.* **392**, 394 (1992).

[167] W B Bonnor, *Commun. Math. Phys.* **51**, 191 (1976).

[168] C C Dyer, S Landry and E G Shaver, *Phys. Rev.* D **47**, 1404 (1993).

[169] S Landry and C C Dyer, *Phys. Rev.* D **56**, 3307 (1997).

[170] C B Collins and G F R Ellis, *Phys Rep.* **56**, 63 (1979).

[171] A H Taub, *Ann. Math.* **53**, 472 (1951).

[172] C B Collins and S W Hawking, *Astrophys. J.* **180**, 317 (1973).

[173] R T Jantzen, *Commun. Math. Phys.* **64**, 211 (1979).

[174] R T Jantzen, in *Cosmology of the Early Universe*, Ed. L Z Fang and R Ruffini (World Scientific, Singapore, 1984).

[175] M A H MacCallum, in *General Relativity, An Einstein Centenary Survey*, Eds. S W Hawking and W Israel (Cambridge University Press, Cambridge, 1979).

[176] D W Hobill, A Burd and A A Coley (Eds.), *Deterministic Chaos in General Relativity* (Plenum Press, New York, 1994).

[177] N J Cornish and J J Levin, *Phys. Rev. Lett.* **78**, 998 (1997).

[178] V G LeBlanc, *Class. Quantum Grav.* **14**, 2281 (1997).

[179] V G LeBlanc, D Kerr and J Wainwright, *Class. Quantum Grav.* **12**, 513 (1995).

[180] G F R Ellis and A R King, *Commun. Math. Phys.* **38**, 119 (1974).

[181] R M Wald, *Phys. Rev.* D **28**, 2118 (1983).

[182] M Goliath and G F R Ellis, Homogeneous cosmologies with cosmological constant. Preprint, (1998). Accepted by *Phys. Rev.* D . Available at *gr-qc/9811068*.

[183] A Rothman and G F R Ellis, *Phys. Lett.* **180B**, 19 (1986).

[184] A Raychaudhuri and B Modak, *Class. Quantum Grav.* **5**, 225 (1988).

[185] G F R Ellis and J Baldwin, *Mon. Not. Roy. Astr. Soc.* **206**, 377 (1984).

[186] J D Barrow, *Mon. Not. Roy. Astr. Soc.* **175**, 359 (1976); *Mon. Not. Roy. Astr. Soc.* **211**, 221 (1984).

[187] A Rothman and R Matzner, *Phys. Rev.* D **30**, 1649 (1984).

[188] R Matzner, A Rothman and G F R Ellis, *Phys. Rev.* D **34**, 2926 (1986).

[189] C B Collins and S W Hawking, *Mon. Not. Roy. Astr. Soc.* **162**, 307 (1973).

[190] J D Barrow, R Juszkiewicz and D H Sonoda, *Nature* **305**, 397 (1983); *Mon. Not. Roy. Astr. Soc.* **213**, 917 (1985).

[191] S Bajtlik, R Juszkiewicz, M Prószyński and P Amsterdamski, *Astrophys. J.* **300**, 463 (1986).

[192] E F Bunn, P Ferreira and J Silk, *Phys. Rev. Lett.* **77**, 2883 (1996).

[193] C B Collins, *Commun. Math. Phys.* **23**, 137 (1971).

[194] J Wainwright, in *Relativity Today*, Ed. Z Perjés (World Scientific, Singapore, 1988).

[195] L Hsu and J Wainwright, *Class. Quantum Grav.* **3**, 1105 (1986).

[196] J Wainright, A A Coley, G F R Ellis and M Hancock, *Class. Quantum Grav.* **15**, 331 (1998).

[197] G F R Ellis and M Bruni, *Phys. Rev.* D **40**, 1804 (1989).

[198] C P Ma and E Bertschinger, *Astrophys. J.* **455**, 7 (1996).

[199] J Bardeen, *Phys. Rev.* D **22**, 1882 (1980).

[200] J Bardeen, P Steinhardt and M S Turner, *Phys. Rev.* D **28**, 679 (1983).

[201] J M Stewart and M Walker, *Proc. R. Soc. London A* **341**, 49 (1974).

[202] P Hogan and G F R Ellis, *Class. Quantum Grav.* **14**, A171 (1997).

[203] B Bertotti, *Proc. R. Soc. London A* **294**, 195 (1966).

[204] R Kantowski, *Astrophys. J.* **155**, 89 (1969).

[205] C C Dyer and R C Roeder, *Astrophys. J. Lett.* **180**, L31 (1973); *Gen. Rel. Grav.* **13**, 1157 (1981).

[206] J P Boersma, *Phys. Rev.* D **57**, 798 (1998).

[207] S. Weinberg, *Astrophys. J.* **208**, L1 (1976).

[208] G F R Ellis, B A C C Bassett and P K S Dunsby, *Class. Quantum Grav.* **15**, 2345 (1998).

[209] G F R Ellis and D M Solomons, *Class. Quantum Grav.* **15**, 2381 (1998).

[210] G F R Ellis, J Hwang and M Bruni (1989), *Phys. Rev.* D **40**, 1819 (1989).

[211] J H Jeans, *Phil. Trans.* **129**, 587 (1902).

[212] M J Rees, in *General Relativity and Cosmology, Proceedings of the XLVII Enrico Fermi Summer School*, Ed. R K Sachs (Academic Press, New York, 1971).

[213] M Bruni, G F R Ellis and P K S Dunsby, *Class. Quantum Grav.* **9**, 921 (1992).

[214] P K S Dunsby, M Bruni and G F R Ellis, *Astrophys. J.* **395**, 54 (1992).

[215] R Maartens, T Gebbie and G F R Ellis, *Phys. Rev.* D **59**, 083506 (1999).

[216] I Müller, *Z. Physik* **198**, 329 (1967).

[217] W Israel and J M Stewart, *Ann. Phys. (N.Y.)* **118**, 341 (1979).

[218] R Maartens and J Triginier, *Phys. Rev.* D **56**, 4640 (1997);
Phys. Rev. D **58**, 123507 (1998).

[219] C G Tsagas and J D Barrow, *Class. Quantum Grav.* **14**, 2539 (1997);
Class. Quantum Grav. **15**, 3523 (1998).

[220] G F R Ellis, *Mon. Not. Roy. Astr. Soc.* **243**, 509 (1990).

[221] T Hirai and K Maeda, *Astrophys. J.* **431**, 6 (1994).

[222] M Bruni, P K S Dunsby and G F R Ellis, *Astrophys. J.* **395**, 34 (1992).

[223] J M Stewart, *Class. Quantum Grav.* **7**, 1169 (1990).

[224] R K Sachs and A M Wolfe, *Astrophys. J.* **147**, 73 (1967).

[225] M Panek, *Phys. Rev.* D **34**, 416 (1986).

[226] P K S Dunsby, *Class. Quantum Grav.* **14**, 3391 (1997).

[227] A Challinor and A Lasenby, *Phys. Rev.* D **58**, 023001 (1998).

[228] R W Lindquist, *Ann. Phys.* **37**, 487 (1966).

[229] W Israel, in *General Relativity*, Ed. O'Raifeartaigh (Oxford University Press, Oxford, 1972).

[230] J M Stewart, *Non Equilibrium Relativistic Kinetic Theory*, Springer Lecture Notes in Physics 10 (Springer, Berlin, 1971).

[231] J D Jackson, *Classical Electrodynamics* (Wiley, New York, 1962).

[232] R P Feynman, R B Leighton and M Sands, *The Feynman Lectures on Physics*, Vol. I (Addison–Wesley, Reading, MA, 1963).

[233] G F R Ellis, D R Matravers and R Treciokas, *Ann. Phys.* **150**, 455 (1983).

[234] G F R Ellis, R Treciokas and D R Matravers, *Ann. Phys.* **150**, 487 (1983).

[235] F A E Pirani, in *Lectures in General Relativity*, Eds. S Deser and K W Ford (Prentice–Hall, Englewood Cliffs, NJ, 1964).

[236] K S Thorne, *Rev. Mod. Phys.* **52**, 299 (1980).

[237] M L Wilson, *Astrophys. J.* **273**, 2 (1983).

[238] T Gebbie and G F R Ellis, Covariant cosmic microwave background anisotropies. I: Algebraic relations for mode and multipole representations. Preprint *astro-ph/9804316*, (1998).

[239] A D Challinor and A N Lasenby, Cosmic microwave background anisotropies in the CDM model: A covariant and gauge-invariant approach. Preprint *astro-ph/9804301*, (1998).

[240] R Maartens, G F R Ellis and W J Stoeger, *Phys. Rev.* D **51**, 1525 (1995); *Phys. Rev.* D **51**, 5942 (1995).

[241] W Hu and N Sugiyama, *Phys. Rev.* D **51**, 2599 (1995).

[242] T Gebbie and G F R Ellis: Covariant cosmic microwave background anisotropies. II: Almost-FLRW standard model. Preprint, (1998).

[243] W Hu and M White, *Astrophys. J.* **471**, 30 (1996).

[244] R K Tavakol and G F R Ellis, *Phys. Lett.* **130A**, 217 (1988).

[245] W Stoeger, R Maartens and G F R Ellis, *Astrophys. J.* **443**, 1 (1995).

[246] G F R Ellis, R Maartens and S D Nel, *Mon. Not. Roy. Astr. Soc.* **184**, 439 (1978).

[247] J Goodman, *Phys. Rev.* D **52**, 1821 (1995).

[248] G Jungman, M Kamionkowski, A Kosowsky and D N Spergel, *Phys. Rev.* D **54**, 1332 (1996).

[249] W R Stoeger, C Xu, G F R Ellis and M Katz, *Astrophys. J.* **445**, 17 (1995).

[250] D Stevens, D Scott and J Silk, *Phys. Rev. Lett.* **71**, 20 (1993).

[251] P Anninos, R A Matzner, T Rothman and M P Ryan, *Phys. Rev.* D **43**, 3821 (1991).

[252] G F R Ellis, D H Lyth and M B Mijić, *Phys. Lett.* **271B**, 52 (1991).

[253] S Weinberg, *Rev. Mod. Phys.* **61**, 1 (1989).

[254] R Penrose, in *Proc. 14th Texas Symp. on Relativistic Astrophysics*, Ed. E J Fergus, (New York Academy of Sciences, New York, 1989).

[255] G W Gibbons, S W Hawking and J M Stewart, *Nucl. Phys.* B **281**, 736 (1987).

[256] D N Page, *Phys. Rev.* D **36**, 1607 (1987).

[257] G Evrard and P Coles, *Class. Quantum Grav.* **12**, L93 (1995).

[258] D H Coule, *Class. Quantum Grav.* **12**, 455 (1995).

[259] G F R Ellis and W R Stoeger, *Class. Quantum Grav.* **4**, 1679 (1987).

[260] R Penrose, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics* (Oxford University Press, Oxford, 1989).

[261] D R Matravers, G F R Ellis and W R Stoeger, *Qu. J. Roy. Astr. Soc.* **36**, 29 (1995).

[262] G F R Ellis, *Mem. Ital. Astr. Soc.* **62**, 553 (1991).

* * *