# Non-Standard Cosmologies

JAYANT V. NARLIKAR and AJIT K. KEMBHAVI
*Tata Institute of Fundamental Research, Bombay 400 005, India*

## Contents

## 1. INTRODUCTION

The purpose of this work is to present in one place some of the important but unconventional attempts made in the last few decades by astronomers, physicists and mathematicians to understand the origin and the large scale structure of the universe. The word "non-standard" in the title indicates the unconventionality, and we clarify it further at the outset.

By "standard" cosmology we mean the hot big bang Friedmann models obtained from Einstein's general relativity. In these models the universe came into existence a finite time ago—at an instant which serves as a starting point for the cosmic time axis. According to these models the universe was initially dominated by radiation. Later its matter component became more and more important in determining the dynamical evolution of the universe. At present although radiation is one of the interesting features of the universe its density



is considerably less ($< 10^{-3}$) than that of observable matter density. The early hot era is believed to have produced helium and deuterium. The presently observed abundances of these nuclei and the microwave background radiation are usually cited as the proofs of this "standard" picture.

Einstein's general theory of relativity certainly represents an intellectual excellence rarely found in the annals of science. The theory beautifully combines the abstract concept of non-Euclidean geometry with the observable phenomena ascribed to gravitation. The standard cosmology described above represents the simplest set of models to follow from the Einstein equations. They can claim the merit of having predicted two very important results: the expansion of the universe and the cosmic microwave background.

The statement "Nothing succeeds like success" is probably best vindicated in the progress of science. The successes of the standard cosmology have inspired considerable important work in theory as well as in observations. Nucleosynthesis, galaxy formation, quasi stellar objects are some of the areas where a great deal of intellectual effort is being spent to-day.

However, this has had negative effects too! By accepting the standard cosmology, the scientist is taking for granted that he has more or less (if not exactly) solved the most important and fundamental problem facing man; viz., the origin of the universe. From the status of "one of the possible pictures" standard cosmology seems to have evolved to "*the* picture" of the universe. A survey of the recent important textbooks in this field will easily confirm this statement. Indeed, there seems to be a general commitment to the standard view which has been aptly summarized by G. R. Burbidge:

> ... In this field at present you find that with very few exceptions if the astronomer is well-known, you know before he speaks what position he will take. Even more disturbing, if he or she is not so well-known but comes from one of the great centers of learning, you also know once you know where the individual comes from, what his or her position will be. ... And the observers tend to get the result they expect. Many of them certainly know what they are looking for, and are not likely to discover anything new. ...
>
> [From the *Proceedings of the* 1976 *IAU/CNRS Colloquium*, Paris, published by CNRS, Paris 1977 (p. 561)]

We find such a state of affairs disturbing for the progress of cosmology as a science. Even Einstein never felt that general relativity was the last word in gravitation and his attempts at a unified field theory testify to his desire to improve on it. To this day general relativity has been verified experimentally only in its weak field approximation (in all due respect to the black hole bandwagon!). None of the cosmological observations enjoy the precision that a laboratory physicist is assustomed to. R. P. Feynman has remarked in this context:

> When a physicist reads a paper by a typical astronomer, he finds an unfamiliar style in the treatment of uncertainties and errors. ... The authors are apparently unwilling to state precisely the

odds that their number is correct, although they have pointed out very carefully the many sources of error, and although it is quite clear that the error is a considerable fraction of the number. The evil is that often other cosmologists or astrophysicists take this number without regard to the possible error, treating it as an astronomical observation as accurate as the period of a planet. . . .

[Lectures on Gravitation, California Institute of Technology]

There are several unsatisfactory features in the standard cosmology. The problem of the origin of the universe is still shrouded in mystery. The isotropy of the microwave background poses problems. The so-called cosmological tests using optical and radio astronomy have become entangled in the parameter fitting exercises in the volutionary properties of galaxies and radio sources. Mach's principle remains an uncomfortable feature which has to be dismissed as based on a pure coincidence. We will discuss some of these aspects in this article.

It is customary in science to subject existing theories to a continued critical scrutiny and to reject, modify or augment them as and when they begin to show inadequacies in explaining observational data. Of late, in astronomy no subject has shown a greater growth in observational output than extragalactic astronomy and cosmology. It is not unreasonable therefore to subject the canonical theoretical framework—the standard cosmology—to the same critical examination as any laboratory theory would be expected to undergo under similar circumstances.

This critical outlook appears to be missing in theoretical cosmology. With a few exceptions, the majority of astronomers and physicists seem to have accepted the standard Einstein–Friedmann hot big bang picture. Here we shall be concerned with the points of view which differ from this band-wagon.

To this end we have collected under the title "non-standard cosmologies" several attempts which depart from the "standard" picture. As will be noticed these attempts range from re-examining the very basic foundations of general relativity to a somewhat superficial patching up of the existing framework. We do not claim to have made an exhaustive search—indeed we may well have inadvertently left out many important ideas. For any such omissions we apologize in advance. As far as possible, we have included ideas which have either generated a certain amount of theoretical discussion or which have inspired fresh observations—or which have shed new light on the existing data. Each section is self-contained and for the sake of comparison we describe the standard picture in the beginning. Those familiar with it may prefer to skip Section 2.

A conventional cosmologist, when faced with a theory purporting to tinker with the basic assumptions of the standard cosmology, counters with a first reaction combining a lack of interest, hostility, suspicion and derision. A common criticism is as follows: "If we want to change the basic framework the number of possibilities is unbounded. We cannot go on testing each and every alternative: if we start doing that we would not know where to begin." M. J. Rees gave an expression to the discomfort of the standard cosmologist when faced with a multitude of non-standard models, in the following words:



In Peebles' well-known textbook, one chapter is entitled "a child's garden of cosmological models". Maybe a "jungle" would better describe the lush diversity of theories expounded at this exceptionally interesting conference. . . .

[From the *Proceedings of the* 1976 *IAU/CNRS Colloquium*, Paris, published 1977 by CNRS, Paris (p. 563)]

To some extent this criticism and this discomfort are justified. Any non-standard cosmology must do at least as well as the standard models. In addition it must make a clear case for the new assumptions that have been introduced. The success of the case is judged either by the conceptual clarity it has brought into the present understanding of physics and cosmology, or by the ease with which it explains any baffling observations.

In presenting the non-standard cosmologies we have tried to emphasize these aspects more than the purely technical details. We hope that our presentation will make the field of non-standard cosmologies look less like a jungle and more like a series of unfamiliar but interesting gardens. To those bored with digging in the same child's garden and who feel venturesome enough to sample new flowers and shrubs we dedicate this article.

## 2. THE GENERAL THEORY OF RELATIVITY AND THE STANDARD COSMOLOGY

The purpose of this chapter is to present a brief outline of the general theory of relativity and the so-called "standard" cosmologies to enable the reader to appreciate the different points of view presented by the non-standard approaches. The treatment here will be elementary but self-contained. For details the reader is referred to standard texts (Eisenhart, 1949; Hawking and Ellis, 1973; Landau and Lifshitz, 1975; Misner, Thorne and Wheeler, 1973; Narlikar, 1978, 1983; Papapetrou, 1974; Wald, 1984; Weinberg, 1972).

### 2.1 The motivation behind general relativity

General relativity took final shape in 1915 and it represents the outcome of Einstein's efforts to resolve certain conceptual difficulties in Newtonian gravitation and in special relativity.

Newtonian gravitation, whether stated as the inverse square law

$$F = G\frac{m_1 m_2}{r^2} \tag{1}$$

or as the Poisson equation

$$\nabla^2 \phi = -4\pi G \varrho \qquad (2)$$

implies an instantaneous propagation of gravitational interaction. As such it is inconsistent with special relativity. If we modify (2) to the wave equation

$$\Box \phi = 4\pi G \varrho \qquad (3)$$

the inconsistency is apparently resolved; but not quite. According to special relativity matter and energy are equivalent. So energy (in the form of radiation) should exert gravitational attraction and $\varrho$ has to be interpreted not only as matter density but also as *energy* density. So, properly speaking it is not a scalar but a (time-time) component of the second ranked energy momentum tensor. This means that Eq. (3) should really form part of a tensor equation

$$\Box \phi_{ik} = 4\pi G T_{ik}. \qquad (4)$$

Thus the original scalar theory has to be further enlarged.

Turning now to special relativity, its basic formulation requires inertial frames. These are the reference frames of observers on whom no forces act. Can such observers be realized in practice? It seems we can eliminate, compensate for or shield against, all natural forces except gravitation. Even if we isolate the experimental system far away from any external gravitating body we cannot eliminate gravitation entirely: for the system's own gravitation is still there. Therefore even the most basic concept of special relativity seems unrealizable in practice!

By proposing the general theory of relativity Einstein sought to combine the ideas of Newtonian gravitation and special relativity in a way that would remove difficulties like those mentioned above. It is worth pointing out that during the pre-general relativity decade 1905–15, there was no compelling *observational* or *experimental* reason for proposing this theory. This situation was in contrast with that prevailing in physics in the post-Michelson–Morley experiment-era prior to special relativity.

## 2.2 Gravity and the curved space-time

Einstein began with the second of the two difficulties mentioned above. If gravitation cannot be eliminated, a physical theory should be so formulated that it takes account of the permanence of gravitation. Any experiment performed on the Earth is being performed in the presence of terrestrial gravity. How to take account of this effect? Einstein's solution was a dramatic one and can be understood with the following example.



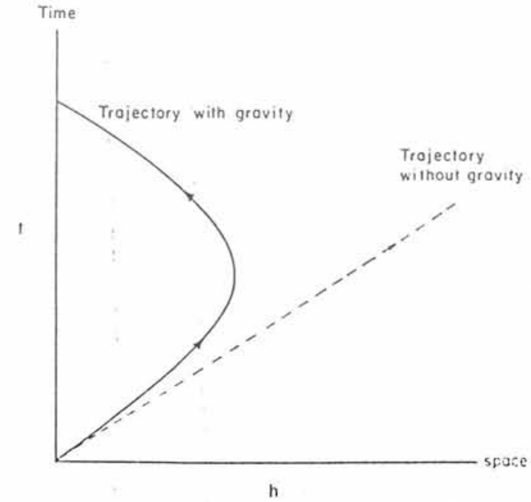



FIGURE 2.1 The dotted trajectory shows how a projectile would move "in a straight line with uniform speed" in the absence of an external force like the Earth's gravity. The continuous trajectory is that of motion under gravity and represents reality.

Imagine a projectile thrown vertically up. If there were no gravity it would continue to move up with uniform speed, as predicted by Newton's first law of motion. Because of gravity the projectile will slow down, come to a halt and begin falling down with increasing speed. The two trajectories, with and without gravity, are shown in Figure 2.1.

According to Einstein the dotted trajectory, being unobservable, is meaningless from the point of view of the terrestrial physicist. Since to him gravity is an irremovable entity, only the continuous trajectory has reality. Einstein therefore sought to rewrite Newton's law in a framework where gravity was eliminated as a *force*; only other forces, e.g., the electromagnetic force, were still separated out as genuine forces. Thus in Figure 2.1 if there are *no* forces acting, the continuous curve represents straight line motion with uniform velocity; and in order to justify this conclusion it is necessary to assume that the background space-time geometry has changed. By the same token planets move in straight trajectories of uniform motion – known as *timelike geodesics* in a curved space-time produced by the Sun's gravity.

In this picture we have therefore the following equivalence:

$$\text{Matter} \sim \text{Gravity} \sim \text{Geometry of curved space-time.} \qquad (5)$$

Einstein gave a quantitative meaning to (5) in such a way as to preserve general co-ordinate invariance.

In special relativity the laws of physics are invariant under the Lorentz transformations. In general relativity they are invariant under the general co-ordinate transformations:

$$x^i \to x'^i, \quad i = 0, 1, 2, 3 \tag{6}$$

where (6) represents a (twice-differentiable) non-singular transformation between two sets of space-time co-ordinates. This transformation has the meaning of describing accelerated frames: an observer at rest in the $x'$-frame is in general accelerated relative to an observer at rest in the $x$-frame.

We will now briefly outline the mathematical tools needed to describe Einstein's ideas in a quantitative manner.

## 2.3 General covariance

Physical laws and the geometrical properties of space-time are not expected to depend on any specific co-ordinate-system. We therefore need quantities which mathematically exhibit this property. Borrowing concepts from physics, we need to define scalars, vectors and tensors in the present scheme.

To fix ideas we start with a space-time manifold $M$ of 3 spacelike plus one timelike dimensions. Let four co-ordinates $x^i (i = 0, 1, 2, 3)$ specify a typical event of $M$. By convention we shall use $x^0$ to denote a timelike co-ordinate and $x^\mu (\mu = 1, 2, 3)$ to denote the three space-like co-ordinates. We shall imply a Latin index to take the four values 0, 1, 2, 3 and a Greek index to take the three values 1, 2, 3. The summation convention will be used wherein a product (or any single expression) with an upper and lower index *identified* is summed over that index. Thus:

$$\begin{aligned} A^i B_i &= A^0 B_0 + A^1 B_1 + A^2 B_2 + A^3 B_3, \\ C_i^{\,i} &= C_0^{\,0} + C_1^{\,1} + C_2^{\,2} + C_3^{\,3}, \\ B_\mu^{\,\mu} &= B_1^{\,1} + B_2^{\,2} + B_3^{\,3}, \qquad \text{etc.} \end{aligned} \tag{7}$$

Expressions like $A^\mu B^\mu$ or $A^i B_i C^i$ are "monsters" which are not expected to arise in a correct calculation.

We will assign a metric $\{g_{ik}\}$ to $M$ through the "line element":

$$ds^2 = g_{ik}\, dx^i dx^k. \tag{8}$$

Here $ds$ represents the "distance" between two neighbouring points with co-ordinates $x^i$ and $x^i + dx^i$. In special relativity with Cartesian co-ordinates the matrix $\|g_{ik}\|$ reduces to the simple form

$$\|\eta_{ik}\| = \text{diagonal}\,(+1, -1, -1, -1). \tag{9}$$



Following Einstein we assume that at any point, by a suitable co-ordinate transformation $g_{ik}$ can be reduced to $\eta_{ik}$. However, this cannot be achieved in general at every point with the same co-ordinate transformation.

How does $g_{ik}$ transform under a co-ordinate transformation? For this we first note that $ds^2$, by its physical meaning, is a *scalar*, i.e., it has the same value in all co-ordinates. Using (6) we therefore get

$$\begin{aligned} g_{ik}\, dx^i dx^k &= g_{ik} \frac{\partial x^i}{\partial x'^m} dx'^m \frac{\partial x^k}{\partial x'^n} dx'^n \\ &= g'_{mn} dx'^m dx'^n \end{aligned}$$

where we have used the standard differential rule:

$$dx^i = \frac{\partial x^i}{\partial x'^m} dx'^m \tag{10}$$

Hence we get

$$g'_{mn} = \frac{\partial x^i}{\partial x'^m} \frac{\partial x^k}{\partial x'^n} g_{ik}. \tag{11}$$

The transformation law (11) is characteristic of the transformation law of *covariant tensors*. In general a covariant tensor or rank $n$ transforms as

$$T'_{i_1 i_2 \ldots i_n} = \frac{\partial x^{k_1}}{\partial x'^{i_1}} \frac{\partial x^{k_2}}{\partial x'^{i_2}} \cdots \frac{\partial x^{k_n}}{\partial x'^{i_n}} T_{k_1 \ldots k_n}. \tag{12}$$

A *covariant vector* transforms as

$$B'_k = \frac{\partial x^i}{\partial x'^k} B_i. \tag{13}$$

Thus if $\phi$ is a scalar function of the co-ordinates $\partial\phi/\partial x^k$ transforms as a covariant vector. On the other hand (10) is an example of the transformation law of a *contravariant vector*. A *contravariant tensor* of rank $n$ transforms as

$$T'^{i_1 \ldots i_n} = \frac{\partial x'^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial x'^{i_n}}{\partial x^{k_n}} T^{k_1 \ldots k_n}. \tag{14}$$

A *mixed tensor* has a few indices *up* (contravariant) and a few indices *down* (covariant). If we identify an upper index with a lower one we *reduce* the rank of the tensor by 2. This is known as *contraction*. Thus if $T_k^{\,i}$ is a mixed tensor of rank 2, $T_i^{\,i}$ is a scalar (i.e., a tensor of zero rank).

The inverse matrix of $\|g_{ik}\|$ is written as $\|g^{ik}\|$, so that

$$g_{ik} g^{kl} = \delta_i^{\,l} \tag{15}$$

where $\delta_i^l$ is the Kronecker delta. It can be verified that $g^{ik}$ is a tensor. It is easy to show that sums of tensors of the same rank are tensors and that products of tensors are tensors. Thus if $A_i$ and $B^k$ are vectors, so are $A_i g^{ik}$ and $B^k g_{ik}$. It is conventional to write

$$A^k = A_i g^{ik}, \quad B_i = B^k g_{ik}. \tag{16}$$

This process is known as the *raising* and *lowering* of index.

The *quotient law* states that if the product of $T_{ik}$ with an arbitrary vector $A^k$:

$$T_{ik} A^k$$

is a vector, then $T_{ik}$ is a tensor. This can be generalized to tensor "quotients" of arbitrary ranks. The proof is straightforward.

*Symmetric* tensors of rank 2 are defined by

$$S_{ik} = S_{ki} \equiv S_{(ik)} = \tfrac{1}{2}(S_{ik} + S_{ki}) \tag{17}$$

while higher rank symmetric tensors are defined by the equality

$$T_{i_1 \ldots i_n} = T_{p_{i_1} \ldots p_{i_n}} \equiv T_{(i_1 \ldots i_n)} \tag{18}$$

$P$ being an arbitrary permutation of $(i_1 \ldots i_n)$. The *antisymmetric* tensors are similarly specified by

$$\left.\begin{aligned} A_{ik} &= A_{ki} \equiv A_{[ik]} = \tfrac{1}{2}(A_{ik} - A_{ki}) \\ B_{i_1 \ldots i_n} &= (-1)^P B_{p_{i_1} \ldots p_{i_n}} \equiv B_{[i_1 \ldots i_n]} \end{aligned}\right\} \tag{19}$$

Thus $g_{ik}$ is a symmetric tensor.

An important antisymmetric tensor is the *Levi-Civita tensor*

$$\varepsilon_{ijkl} = (-g)^{1/2}[i, j, k, l] \tag{20}$$

where

$$g = det \| g_{ik} \| \tag{21}$$

and

$$[i, j, k, l] = \begin{cases} +1 \text{ if } [i, j, k, l] \text{ is an even permutation of } [0, 1, 2, 3] \\ -1 \text{ if } [i, j, k, l] \text{ is an odd permutation of } [0, 1, 2, 3] \\ 0 \text{ otherwise.} \end{cases} \tag{22}$$

However, (20) is really a "pseudo tensor" because it changes sign under reflection. On the other hand a genuinely antisymmetric tensor is the electromagnetic field

$$F_{[ik]} = \frac{\partial A_k}{\partial x^i} - \frac{\partial A_i}{\partial x^k} \tag{23}$$

where $A_k$ is the potential 4-vector.



## 2.4 Covariant differentiation

It is easy to verify that if $A_i$ is a vector function of space-time co-ordinates (i.e., a *vector field*), then $A_{i,k} \equiv \partial A_i / \partial x^k$ is not in general a vector. This is because $A_{i,k}$ represents the rate of change of $A_i$, and hence involves differencing the values of $A_i$ at two neighbouring points before proceeding to the limit:

$$A_{i,k} = \lim_{\delta x^k \to 0} \frac{A_i(x^k + \delta x^k) - A_i(x^k)}{\delta x^k}. \tag{24}$$

However, while the denominator is a vector the numerator is not! $A_i(x^k + \delta x^k)$ transforms with different transformation coefficients than does $A_i(x^k)$. In order to obtain a legitimate difference we should take the difference of vectors at the *same* point. This is achieved by the so-called "parallel transfer" of the vector $A_i(x^k)$ to the point $x^k + \delta x^k$.

In the parallel transfer, we expect the components of $A_i$ to change to say $A_i(x^k) + \delta A_i$, where $\delta A_i$ will be linear in $A_i$ and the displacement $\delta x^k$. The most general way of writing this is

$$\delta A_i = \Gamma^l_{ik} A_l \delta x^k, \tag{25}$$

where the three index symbol $\Gamma^l_{ik}$ is called the Christoffel symbol. Taking the difference in the revised form and proceeding to the limit we get the *covariant derivative* of $A_i$ as

$$A_{i;k} = \frac{\partial A_i}{\partial x^k} - \Gamma^l_{ik} A_l. \tag{26}$$

By definition, $A_{i;k}$ is a second rank tensor field.†

The quantities $\Gamma^i_{kl}$ define the "affine connection" on $M$ by introducing the notion of parallelism between infinitesimally separated directions. We will assume—as is intuitively expected—that a scalar quantity does not change under parallel transfer. This is consistent with the fact that $\partial\phi/\partial x^i$ does transform as a vector. By considering the scalar $A_i B^i$ for two vector fields $A_i$

†Note the difference: a comma is used to denote the ordinary derivative while a semicolan is used for the covariant derivative.

and $B^i$ it is then easy to deduce that the covariant derivative of $B^i$ is given by

$$B^i_{;k} = \frac{\partial B^i}{\partial x^k} + \Gamma^i_{kl} B^l. \tag{27}$$

We will further assume—although it is not so far warranted—that

$$\Gamma^i_{kl} = \Gamma^i_{lk}. \tag{28}$$

The transformation law of $\Gamma^i_{kl}$ is given by

$$\Gamma^i_{kl} = \frac{\partial x^i}{\partial x'^m}\frac{\partial x'^n}{\partial x^k}\frac{\partial x'^p}{\partial x^l}\Gamma'^m_{np} + \frac{\partial^2 x'^p}{\partial x^k \partial x^l}\frac{\partial x^i}{\partial x'^p}. \tag{29}$$

Thus $\Gamma^i_{kl}$ is not a tensor. This is to be expected since it makes up the difference between the covariant derivative (which is a tensor) and the ordinary derivative (which is *not* a tensor).

The form of $\Gamma^i_{kl}$ is determined by the further requirement that

$$g_{ik;l} \equiv 0. \tag{30}$$

This identifies $\Gamma^i_{kl}$ with the *Riemannian* affine connection. A short calculation gives

$$\Gamma^i_{kl} = \tfrac{1}{2} g^{im}\{g_{mk,l} + g_{ml,k} - g_{kl,m}\}, \tag{31}$$

where the comma denotes the ordinary derivative.

The geometry based on these assumptions is called the Riemannian geometry and it forms the basis of the general theory of relativity.

## 2.5 Space-time curvature

Suppose we have two curves $\Gamma$ and $\Gamma'$ joining two points $X_1$ and $X_2$ in $M$. Let $A^i$ be a vector at $X_1$. If we decide to parallel transfer it to $X_2$, will we get a unique answer? In other words, for arbitrary $\Gamma$ and $\Gamma'$ will the parallel transfer along $\Gamma$ be equivalent to parallel transfer along $\Gamma'$?

If $\Gamma$ is given parametrically by the equations

$$x^i = a^i(\lambda), \tag{32}$$

then if $X$ is a general point $(x^i)$ on $\Gamma$ (see Figure 2.2), we get the following differential equation for $A^i$:

$$\frac{dA^i}{d\lambda} = -\Gamma^i_{kl} A^k \frac{da^l}{d\lambda}. \tag{33}$$

Hence the condition for uniqueness and path independence of $A^i$ at $X_2$ is that we are able to generate a vector field $A^i(X)$ which satisfies the relation

$$\frac{\partial A^i}{\partial x^l} = -\Gamma^i_{kl} A^k. \tag{34}$$



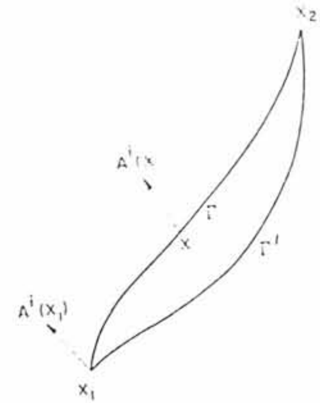


FIGURE 2.2 The parallel transport of a vector from $X_1$ to $X_2$ will give (in general) different answers at $X_2$ along the curves $\Gamma$ and $\Gamma'$.

The necessary and sufficient condition for the existence of $A^i(X)$ is given by the relation

$$R^i_{klm} \equiv \frac{\partial \Gamma^i_{km}}{\partial x^l} - \frac{\partial \Gamma^i_{kl}}{\partial x^m} + \Gamma^i_{ln}\Gamma^n_{km} - \Gamma^i_{mn}\Gamma^n_{kl} = 0. \tag{35}$$

The necessity may be verified by differing (34) once with respect to $x^m$ and using the equality $A^i_{,lm} = A^i_{,ml}$. For sufficiency see Einsenhart (1949) for a detailed proof.

That $R^i_{klm}$ is a tensor can be verified either directly or by using the quotient law in the identity

$$A_{i;ml} - A_{i;lm} \equiv R^i_{klm} A^k. \tag{36}$$

Equation (36) also tells us that unless $R^i_{klm} = 0$, the operation of covariant differentiation is *not* commutative.

The tensor $R_{iklm}$ is called the *Riemann Christoffel tensor* or the *curvature tensor*. It can be shown that if $R^i_{klm} = 0$ the space-time can be given the metric (9) of special relativity. The condition of the vanishing of the curvature tensor implies that $M$ is *flat*. It need not, however, have a Euclidean or a pseudo-Euclidean topology. (The surface of a cylinder has $R^i_{klm} = 0$, although its topology is not that of a plane.)

If $R^i_{klm} \neq 0$ the space is *curved*. In that case we cannot generate a vector field by a parallel transport as attempted in Figure 2.2. We can, however, still choose a special co-ordinate system which has $\Gamma^i_{kl} = 0$ at a specified point $P$. This means that at $P$ and in an infinitesimal neighbourhood of $P$, the $g_{ik}$ can be chosen to be equal to $\eta_{ik}$ with $g_{ik,l} = 0$ at $P$. Such a co-ordinate system is said to be *locally inertial* at $P$. It is easy to see that given non-zero $\Gamma'^{i}_{kl}$, with co-ordinates $x'^{k}$, we can use (29) to obtain $x^k$ such that $\Gamma^i_{kl} = 0$ at $P$.

The tensor $R_{iklm}$ has many symmetries which are described by the relations

$$R_{iklm} = R_{[ik][lm]} = R_{lmik}, \quad R_{i[klm]} = 0. \tag{37}$$

Hence of its $4^4 = 256$ components there are in fact only 20 algebraically independent components.

There are also differential relations connecting $R_{iklm}$:

$$R_{ik[lm;n]} = 0. \tag{38}$$

which are known as the *Bianchi identities*.

Finally we define two important quantities. The first is the *Ricci tensor*:

$$R_{ik} = R^{l}_{ikl}, \tag{39}$$

and the second is the *scalar curvature*:

$$R = R^{k}_{k} \equiv R^{ik}_{ik}. \tag{40}$$

It is easy to verify with (38) that

$$(R^{ik} - \tfrac{1}{2}g^{ik})_{;k} = 0. \tag{41}$$

The tensor

$$G_{ik} = R_{ik} - \tfrac{1}{2}g_{ik}R \tag{42}$$

is called the *Einstein tensor*. Both the Ricci tensor and the Einstein tensor are symmetric. The following expression for the Ricci tensor is often useful for computation:

$$R_{kl} = \frac{\partial^2 \ln\sqrt{(-g)}}{\partial x^k \partial x^l} - \frac{\partial \Gamma^i_{kl}}{\partial x^i} + \Gamma^i_{ln}\Gamma^n_{ki} - \Gamma^n_{kl}\frac{\partial}{x^n}(\ln\sqrt{-g}). \tag{43}$$

## 2.6 Geodesics

In Euclidean geometry a straight line is defined as the line of shortest distance between any two specified points. The word "straight" also implies that the line does not change its direction as we move along it. Both these definitions can be generalized to Riemannian geometry and lead to "geodesics".

Let $x^i(\lambda)$ denote a curve in $M$ for a real parameter $\lambda$. The tangent vector at a typical point is given by $u^i = dx^i/d\lambda$. The condition of "straightness" implies that $u^i$ does not change along the line under parallel transport, i.e.,

$$u^k u^i_{;k} = 0. \tag{44}$$



In terms of $\lambda$ this becomes

$$\frac{d^2x^i}{d\lambda^2} + \Gamma^i_{kl}\frac{dx^k}{d\lambda}\frac{dx^l}{d\lambda} = 0. \tag{45}$$

It is easy to verify that (45) has a first integral

$$g_{ik}\frac{dx^i}{d\lambda}\frac{dx^k}{d\lambda} = C, \qquad C = \text{constant}. \tag{46}$$

Thus the parameter $\lambda$ cannot be arbitrary, since (46) will not hold if we replace $\lambda$ by an arbitrary monotonic function of $\lambda$. The parameter $\lambda$ is called the *affine* parameter.

If $C > 0$ the geodesic is *time-like*, if $C < 0$ it is *space-like* and if $C = 0$ it is *null*. Note that if $C > 0$ we can use $s$ as an affine parameter while for $C < 0$, $|s|$ serves as an affine parameter.

The generalization of shortest distance idea is the concept of stationary distance

$$s(P_2;\, P_1|\Gamma) = \int_{P_1}^{P_2} ds, \qquad \delta s = 0 \tag{47}$$

where $\Gamma$ is a typical curve joining $P_1$ to $P_2$. It can be shown that the same equations (45) result from this procedure.

*Geodesic deviation* This denotes the rate at which two neighbouring geodesics separate. Suppose we take a series of geodesics labelled by a continuous parameter $\mu$ such that each geodesic is parametrically described by an affine parameter $\lambda$. Thus a typical point on a typical geodesic is described by $x^i(\lambda, \mu)$. The vectors

$$u^i = \frac{\partial x^i}{\partial \lambda}, \qquad v^i = \frac{\partial x^i}{\partial \mu} \tag{48}$$

are respectively the tangent vector and the "separation" vector.

A straightforward calculation comparing the rate of change of separation of two points on neighbouring goedesics gives the so-called equation of geodesic deviation:

$$\frac{d^2v^i}{d\lambda^2} + R^i_{klm}u^k v^l u^m = 0. \tag{49}$$

Thus in flat space the separation vector depends linearly on the affine parameter. In curved space the non-vanishing of $R^i_{klm}$ may lead to "focussing" of geodesics.

## 2.7 The principle of equivalence

The fact that we can make $\Gamma'^{i}_{kl} = 0$ at a point with a suitable choice of co-ordinates was given a physical significance by Einstein. Physically it means that by a choice of a suitably accelerated frame we can eliminate the effect of gravity at any given point $P$ of $M$. This is illustrated by the celebrated example of the falling lift. In a freely falling lift the rider would have a feeling of weightlessness. An astronaut in a space-craft orbiting the Earth has the same experience.

What has been achieved here is really a local elimination of gravity. Unless $R^{i}_{klm} = 0$, we cannot do this globally. Also, the possibility of achieving a limited elimination of gravity is due to the fact that for any body the inertial mass is equal to its passive gravitational mass. (This is nothing else but a restatement of Galileo's law of falling bodies.)

The above limited equivalence of accelerated frames with gravity and its implication that the inertial and gravitational masses are equal are referred to as the *weak principle of equivalence*.

By contrast the *strong principle of equivalence* implies the following: in an inertial co-ordinate system ($\Gamma'^{i}_{kl} = 0$) the physical laws would be seen to behave as they would in a flat space-time. This means that if the laws are expressed in a generally covariant manner, we can observe them *locally* in an inertial co-ordinate system as if we had no gravity present. The word "locally" is important: in general an inertial co-ordinate system can only be achieved locally, not globally.

Consider a Minkowski space-time with the line element

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2. \qquad (50)$$

An observer $0'$ moving along the $x$-axis with uniform acceleration $g$ will have co-ordinates

$$x = \frac{c^2}{g}\left(\cosh\frac{gt'}{c} - 1\right), \qquad t = \frac{c}{g}\sinh\frac{gt'}{c}, \qquad y = z = 0 \qquad (51)$$

where $t'$ is the proper time of $0'$. The velocity of $0'$ at this stage is

$$u = c\tanh\frac{gt'}{c}. \qquad (52)$$

We now consider a frame of reference in which $0'$ is at rest but which is non-inertial. This is obtained by the following transformations:

$$\begin{aligned} x &= \frac{c^2}{g}\left(\cosh\frac{gt'}{c} - 1\right) + x'\cosh\frac{gt'}{c} \\ y &= y', \quad z = z' \\ t &= \frac{c}{g}\sinh\frac{gt'}{c} + \frac{x'}{c}. \end{aligned} \qquad (53)$$



This frame really is made out of a series of inertial frames which momentarily coincide with $0'$. We get for the line element

$$ds^2 = \left(1 + \frac{gx'}{c^2}\right)^2 dt'^2 - dx'^2 - dy'^2 - dz'^2. \qquad (54)$$

In this frame $\Gamma'^{i}_{kl} \neq 0$, and if we seek to replace the uniform acceleration by a pseudo gravitational field, the field will have a potential $\phi \approx -gx'$. Further, for $|gx'| \ll c^2$ we have

$$g_{00} \approx 1 - \frac{2\phi}{c^2}. \qquad (55)$$

Thus $g_{00}$ appears to be related to the gravitational potential.

In the above example we get a hint of the basic idea that led Einstein to the general theory of relativity. We note first that to an accelerated observer, the space-time geometry appears non-Euclidean—in the sense that the coefficient of $dt'^2$ in (54) is not unity but dependent on $x'$. Now, we could produce such an accelerated observer if we had a uniform gravitational field (along the $x'$-axis) in which such an observer were falling freely. And, $\phi$ denotes the potential for such a gravitational field. In other words there seems to be a link between the gravitational field and space-time geometry.

Going over from the above example to actual gravitational fields in nature is like making a transition from the trivial to the non-trivial. In the above example there is no *real* gravitational field—nor is the geometry really non-Euclidean. A change of co-ordinates $(x', y', z', t') \to (x, y, z, t)$ in the above example tells us that the space-time is the simple Minkowski one. Nevertheless the example gives us a hint at to how to proceed in the case of real gravitational fields: in particular we can visualize a connection between $g_{00}$ and $\phi$ even in the general case.

We now turn to the Einstein equations and see how such a connection comes about.

## 2.8 The field equations of general relativity

By a heurustic chain of reasoning (which we will not go into here) Einstein (1915) arrived at the following set of gravitational field equations:

$$R_{ik} - \tfrac{1}{2}g_{ik}R = -\kappa T_{ik}, \qquad (56)$$

where $T_{ik}$ is the energy momentum tensor of all types of matter and radiation present in space (gravity of course excluded), and $\kappa$ is a coupling constant. These equations have the following desirable properties:

(i) They are generally covariant.

(ii) If we take the hint from the preceding section and think of $g_{ik}$ as related to gravitational potentials, with $g_{44}$ given by (55), then (56) are second order differential equations generalizing the concept of the Poisson equation (4).

(iii) In the case of "weak" gravitational fields (56) contain Newtonian dynamics and Newtonian gravity if we take

$$\kappa = \frac{8\pi G}{c^4}. \tag{57}$$

This last statement needs some elaboration. In a "weak field approximation" it is possible to write the space-time metric in a nearly special relativistic form. That is, taking $c = 1$, and the special relativistic metric tensor as $(+1, -1, -1, -1)$, we may write

$$g_{ik} = \eta_{ik} + h_{ik}, |h_{ik}| \ll 1. \tag{58}$$

Also, the matter present is supposed to move slowly compared to the speed of light so that all time derivatives of $h_{ik}$ may be ignored. Then it can be shown that the (4, 4) component of the field equations written in the equivalent form

$$R_{ik} = -\kappa[T_{ik} - \tfrac{1}{2}g_{ik}g^{mn}T_{mn}] \tag{59}$$

reduces to the Poisson equation for Newtonian potential given by

$$\phi = -\tfrac{1}{2}h_{44}. \tag{60}$$

The equations of motion follow from the field equations by making use of the identity (41); i.e., from

$$T^{ik}{}_{;k} = 0. \tag{61}$$

In the Newtonian approximations Eqs. (61) show that the gradient of $\phi$ acts as a force. We have thus seen how the connection of (55) can be made in the case of weak gravitational fields.

As was shown by Hilbert (1915), shortly after Einstein's derivation of these equations, it is possible to obtain them from a variational principle

$$\frac{\delta A}{\delta g_{ik}} = 0 \tag{62}$$

where $A$ is the action defined over a subspace $V$ of $M$:

$$A = \int_V \left(\frac{c^4}{16\pi G} R + L\right) \sqrt{(-g)}d^4x \tag{63}$$

with $L$ standing for the *Lagrangian density* of the matter and radiation in $V$. In this notation $T^{ik}$ is given by

$$\delta \int_V L\sqrt{(-g)}d^4x = -\frac{1}{2}\int_V T^{ik}\delta g_{ik}\sqrt{(-g)}d^4x. \tag{64}$$



The energy-momentum tensor obtained in this way is automatically symmetric in $(i, k)$ and satisfies the conservation law (61). In the same way (41) follows because

$$\delta \int_V R\sqrt{(-g)}d^4x = -\int_V (R^{ik} - \tfrac{1}{2}g^{ik})\delta g_{ik}\sqrt{(-g)}d^4x. \tag{65}$$

Because of the four conservation equations, the number of independent field equations is only six. However, since the ten $g_{ik}$ are subject to four arbitrary co-ordinate transformations $x^i \to x'^i$, the system of equations is not underdetermined.

## 2.9 The Solar-system tests of general relativity

The experimental tests of general relativity have been largely confined to the Solar-system. Here, however, the gravitational field is not strong and so the departure from Newtonian gravity is very small. We briefly describe the tests below. With the exception of the last test, these are based on the Schwarzschild (1916) solution, which describes the curved space-time around a spherically symmetric distribution of mass $M$ by the line element

$$ds^2 = \left(1 - \frac{2GM}{c^2R}\right)dT^2 - \left(1 - \frac{2GM}{c^2R}\right)^{-1} dR^2 - R^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{66}$$

Here $(R, \theta, \phi)$ are the spherical polar co-ordinates with the property that the surface area of a spherical surface of co-ordinate radius $R$ is $4\pi R^2$. They are often referred to as the Schwarzschild co-ordinates. For a detailed discussion of some of the tests see standard texts (e.g., Weinberg, 1972) and Section 11.

(i) *The gravitational redshift* Let the line element

$$ds^2 = g_{ik}dx^i dx^k \tag{67}$$

describe a static space-time. This means that if $x^0$ is a timelike co-ordinate, $g_{ik}$ are independent of $x^0$. If $x^\mu$ = constant = $a^\mu$, $b^\mu$ are the world lines of two typical observers $a$ and $b$, the frequency of light sent by $a$ to $b$ changes from $\nu_a$ to $\nu_b$ according to the formula

$$\frac{\nu_a}{\nu_b} = \sqrt{\left(\frac{g_{44}(b^\mu)}{g_{44}(a^\mu)}\right)}. \tag{68}$$

This has been measured in white dwarf stars Sirius B and 40 Eridani B. The frequency shift $\Delta\nu_a/\nu_b = |\nu_b - \nu_a|/\nu_a \sim 10^{-4}$–$10^{-5}$ is too small to quantitatively confirm (68). However a terrestrial experiment using the Mossbauer effect (Pound and Rebka, 1960) does confirm (68) within the experimental errors. This involves measuring the change in the frequency of a $\gamma$-ray photon emitted by $Fe^{*57}$ (exited state of Iron nucleus) when the photon falls through a height of ~60–70 feet.

(ii) *The perihelion precession of Mercury* The small deviations from Newtonian gravity produce small changes in the planetary orbits. The change is most marked for planet Mercury, whose perihelion is expected to precess at a rate

$$n = \frac{6\pi G M_{\odot}}{l T c^2}$$

where $M_{\odot}$ = mass of Sun, $l$ = semi latus rectum of Mercury's orbit and $T$ = period of Mercury's orbit. This works out at ~43″ of arc per century. Although small, this effect has been measured and found to be in good agreement with the above figure.

(iii) *The bending of light* In the curved space-time a null geodesic represents the track of a light ray. In the weak field approximation, the calculation of the track shows that a light ray grazing the solar limb suffers a "bending" of an angle

$$\alpha = \frac{4GM_{\odot}}{R_{\odot} c^2} \simeq 1.75''$$

where $R_{\odot}$ = radius of the Sun.

In Newtonian gravity if we assume that light (in the form of photons) is attracted by the Sun, the bending produced is half of this value. Recent results (Fomalont and Sramek, 1975) using microwaves clearly favour the relativistic value.

(iv) *Radar echo delay* General relativity predicts that the round trip echo delays, for light signals travelling between the Earth and another planet grazing the solar limb, will be slightly delayed compared to a signal sent when the Sun does not happen to be near the track. Recently Reasenberg *et al.* (1979) have used the data obtained from radio ranging to the Viking landers on Mass to verify the theoretical value predicted by general relativity to within 0.2% accuracy.

(v) *The equality of inertial and gravitational mass* This is discussed in Chapter X in connection with other gravitation theories as well as general relativity.



(vi) *The precession of a gyroscope* A rotating Earth will produce a change in the line element (66) by introducing small terms of the type $g_{0\mu} dT dx^{\mu}$. Such terms show their effect on the precession of the axis of an orbiting gyroscope. The effect is very small (~7″ per year) but is now technically measurable. Hence an experiment has been planned (Everitt, 1974), but at the time of writing this article the results are still to come.

## 2.10 Application to cosmology: the standard models

Apart from the experimental tests discussed in the previous sectiom, general relativity has been applied to the astrophysics of supermassive objects, black holes etc. and to cosmology. Since this review article is devoted to cosmology, we shall only discuss here the cosmological applications of general relativity. In particular we will describe first the so-called "standard" models. For detailes the reader is again referred to standard texts (e.g., Weinberg, 1972; Narlikar, 1978, 1983).

*The standard models* These were first obtained by A. Friedmann (1922), and are based on two simplifying postulates: the Weyl postulate and the cosmological principle.

The *Weyl postulate* identifies the world lines of galaxies (treated as points) with a 3-parameter set of timelike geodesics $\{\Gamma_\mu\}$ orthogonal to a family of spacelike hypersurfaces $\{\Sigma_t\}$. Taking $x^\mu$ as the three parameters specifying $\Gamma_\mu$ and the typical surface $\Sigma_t$ given by $t$ = constant, the line element becomes

$$ds^2 = g_{44} dt^2 + g_{\mu\nu} dx^\mu dx^\nu. \tag{69}$$

Since the lines $x^\mu$ = constant are geodesics, it can be shown that $g_{44}$ depends on $t$ only and hence by a suitable time–time ($t \to t'$) transformation we can make $g_{44} = 1$. Thus $t$ measures the proper time of each observer following a $\Gamma_\mu$-geodesic. Such an observer is called a *fundamental observer*. Note that since we have a system of hypersurfaces $t$ = constant, it is possible to talk of a global simultaneity for all points on $\Sigma_t$. The time $t$ is called the *cosmic time*.

The *cosmological principle* states that the hypersurfaces $\Sigma_t$ are homogeneous and isotropic in $M$. This means, they are maximally-symmetric subspaces of $M$. Physically, all fundamental observers at a given cosmic time notice the same large scale features of the universe and do not see any preferred direction.

Under these assumptions the cosmological line element becomes (with $c = 1$)

$$ds^2 = dt^2 - S^2(t)\left[\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right]. \tag{70}$$

Here $S(t)$ is called the *scale factor* and $k$ is the *curvature parameter* taking the values 1, $-1$ or 0. This is known as the *Robertson-Walker line element* after the two persons who derived it independently in a rigorous form (Robertson, 1935; Walker, 1936).

We will now consider the standard models in two forms depending on whether the universe is largely made of pressure-free matter (dust) or radiation.

(i) *Dust models* Taking the energy momentum tensor for the "smoothed out universe" as for a homogeneous dust distribution

$$T^{ik} = \varrho u^i u^k, \tag{71}$$

the Einstein equations for the line element (70) reduce to only two ordinary differential equations

$$2\frac{\ddot{S}}{S} + \frac{\dot{S}^2 + k}{S^2} = 0, \tag{72}$$

$$3\frac{\dot{S}^2 + k}{S^2} = 8\pi G\varrho. \tag{73}$$

From these, or from the conservation law $T^{ik}{}_{;k} = 0$, we get a first integral

$$\varrho S^3 = \text{constant} = \varrho_0 \quad \text{(say)}. \tag{74}$$

We may, without loss of generality take $S = 1$ at the present epoch $t = t_0$ so that $\varrho_0$ is the present smoothed out density of matter in the universe. Then (73) gives

$$\dot{S}^2 = -k + \frac{8\pi G\varrho_0}{3S}. \tag{75}$$

The solutions of these equations are illustrated in Figure 2.3, for the three values of the parameter $k$ and for a given $\varrho_0$. Note that by fixing $\varrho_0$ we have to place $t_0$ at different positions on the three curves.

In all the models $S(t)$ is shown to increase with $t$ from $S = 0, t = 0$. This is consistent with observations of nebular redshift first found by Hubble (1929). If we are receiving now light waves from a galaxy which sent out these waves at an earlier epoch $t_1 (< t_0)$, we should see an increase in the wavelength of light by a fraction

$$z = \frac{S(t_0)}{S(t_1)} - 1 = \frac{1}{S(t_1)} - 1. \tag{76}$$

Such an increase has been noticed in the spectra of almost all distant galaxies. This strongly suggests that $S(t)$ is an increasing function of time, a result which is often quoted in the form: "the universe is expanding".



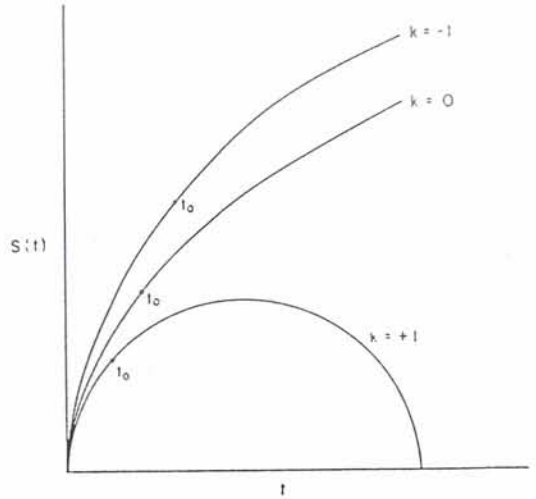



FIGURE 2.3 The three types of Friedmann models for the three values of the space-curvature parameter $k$. The time $t_0$ denotes the present "age" of the universe corresponding to the measured value of Hubble's contant.

For nearby galaxies ($r$ small) we may approximate the distance from $r = 0$ (taken to be the location of our Galaxy) by the formula

$$D \cong rS(t_0) = r,$$

while, from (76) we get by a Taylor expansion near $t = t_0$,

$$z \approx \dot{S}(t_0)(t_0 - t_1).$$

However, light travels along a radial null line which is given by

$$r = \int_{t_1}^{t_0} \frac{dt}{S(t)} \approx t_0 - t_1,$$

which gives us

$$r \simeq \frac{z}{\dot{S}(t_0)} = \frac{z}{\dot{S}(t_0)/S(t_0)} = \frac{z}{H_0}, \tag{77}$$

where in general,

$$H(t) = \frac{\dot{S}(t)}{S(t)}, \quad H(t_0) = H_0. \tag{78}$$

The result (77) was in fact first obtained by Hubble (1929) and its derivation by the Friedmann model went a long way towards establishing the "expanding universe" concept. The present measurements indicate (Tammann, 1977) that

$$H_0^{-1} \cong (1 - 1.8) \times 10^{10} \text{ years.} \tag{79}$$

The wide range of uncertainty reflects the ambiguities in the estimation of distances $r$. Unless otherwise stated we will take for the purpose of calculations

$$H_0^{-1} = \tfrac{4}{3} \times 10^{10} \text{ years.} \tag{80}$$

Knowing $H_0$ we can estimate $t_0$, provided we know which of the three curves of Figure 2.3 describes the actual universe, and the value of $\varrho_0$. For this purpose define the *deceleration parameter*

$$q(t) = -\frac{\ddot{S}S}{\dot{S}^2}, \quad q_0 = q(t_0) = -H_0^{-2}\ddot{S}(t_0). \tag{81}$$

We then get from (72) and (73) at $t = t_0$,

$$-2q_0H_0^2 + H_0^2 + k = 0;$$

$$H_0^2 + k = \frac{8\pi G\varrho_0}{3}.$$

From the first we see that

$$k \gtreqless 0 \Leftrightarrow q_0 \gtreqless \tfrac{1}{2}. \tag{82}$$

we also get

$$\varrho_0 = \frac{3H_0^2}{4\pi G}q_0. \tag{83}$$

Hence if we know $q_0$ we can estimate $\varrho_0$ as well as $k$. In principle $q_0$ is measureable by observations of $r$ and $z$ for remote galaxies. However the ambiguities in the distance determination mount up rapidly so that $q_0$ cannot at present be quoted reliably (Gunn, 1977).

Assuming that $q_0$ is known the "age of the universe" is given by (see Weinberg, 1972)

$$t_0 = \begin{cases} (2/3)H_0^{-1} \text{ for } q_0 = \tfrac{1}{2}, \quad k = 0 \\ H_0^{-1}\left\{\dfrac{q_0}{(2q_0 - 1)^{3/2}}\left(\sin^{-1}\dfrac{q_0 - 1}{q_0} + \dfrac{\pi}{2}\right) - \dfrac{1}{2q_0 - 1}\right\} \\ \qquad\qquad \text{for } q_0 > \tfrac{1}{2}, \quad k = 1 \\ H_0^{-1}\left\{\dfrac{1}{1 - 2q_0} - \dfrac{q_0}{(1 - 2q_0)^{3/2}} \ln\left(\dfrac{1 - q_0}{q_0} + \dfrac{\sqrt{(1 - 2q_0)}}{q_0}\right)\right\} \\ \qquad\qquad \text{for } q_0 < \tfrac{1}{2}, \quad k = -1. \end{cases} \tag{84}$$



For $q_0 = 1$, $t_0 = [(\pi/2) - 1]H_0^{-1}$ and for $q_0 = 0$, $t_0 = H_0^{-1}$. The value

$$\varrho_c = \frac{3H_0^2}{8\pi G} \tag{85}$$

is called the closure density. For $\varrho_0 > \varrho_c$, we must have $k = 1$, i.e., the universe with closed (compact) $\Sigma$-hypersurfaces. The visible density—in the form of galaxies—is only $\sim 3\%$ of $\varrho_c$. Thus if there is no "hidden matter", $q_0 \approx 0.015$. Notice that a large $q_0$ lowers the age of the universe. Since the age of the Galaxy is $(1 - 1.5) \times 10^{10}$ years, a very large $q_0 (q_0 \geqslant 1/2)$ is somewhat embarrassing. We will return to this question in Section 10.

(ii) *The radiation universe* Although matter dominates the $T_{ik}$ in the present epoch of the universe, it is likely that radiation was the dominant part of $T_{ik}$ in the early stages ($S \approx 0$). If $u$ is the radiation density then the conservation of the radiation part of $T_{ik}$ gives

$$S^4u = \text{constant} = A \text{ (say)}. \tag{86}$$

Therefore $\varrho/u \propto S \to 0$ as $S \to 0$. For this reason it is essential to consider the radiation dominated early universe which is now given by the field equations

$$\left.\begin{aligned} 2\frac{\ddot{S}}{S} + \frac{\dot{S}^2 + k}{S^2} &= -\frac{8\pi Gu}{3} \\ 3\frac{\dot{S}^2 + k}{S^2} &= 8\pi Gu. \end{aligned}\right\} \tag{87}$$

The second of these equations together with (86) (which in fact follows from Eqs. (87)) leads to

$$\dot{S}^2 = -k + \frac{8\pi GA}{3S^2}. \tag{88}$$

For $S \to 0$, the $k$-term is not important and can be ignored. We then get

$$S = \left(\frac{32\pi GA}{3}\right)^{1/4} t^{1/2}. \tag{89}$$

The corresponding radiation temperature is given by

$$T = \left(\frac{u}{a}\right)^{1/4} = \left(\frac{A}{aS^4}\right)^{1/4} = \left(\frac{3}{32\pi Ga}\right)^{1/4} t^{-1/2}, \tag{90}$$

where $a$ is the radiation constant. At $t \approx 1$s, $T \sim 1.5 \times 10^{10}\,°$K. However, if pair creation is also taken into account (Hoyle and Taylor, 1964), the temperature at 1s is $\sim 10^{10}\,°$K.

George Gamow and his collaborators (Gamow, 1946; Alpher, Bethe and Gamow, 1948) had suggested that the early universe was hot enough to synthesize elements starting from free neutrons, protons and electrons. Later work (Reeves, Audoze, Fowler and Schramm, 1973) has shown that this idea is workable only up to helium. Thus deuterium (D) and helium (He) are produced soon after the initial epoch $t = 0$. The parameters $q_0$ and $H_0$ determine the ratios D/H and He/H of these light nuclei by weight. If there is subsequently no more production of D and He in significant amounts, the present abundances carry the signature of the early epochs, provided they are universal.

The data (Reeves, Audoze, Fowler and Schramm, 1973) on abundances is still too scanty to justify the adjective "universal". If it is so assumed, it becomes difficult to provide alternative scenarios for production of D and He of this order. Stars, which are able to account for abundances of heavier nuclei fail to produce D and He in quantities of this order:

$$\frac{\text{He}}{\text{H}} \sim 0.33 \qquad \frac{\text{D}}{\text{H}} \sim 10^{-6}\text{–}10^{-5}. \tag{91}$$

If such observed abundances are considered universal a "hot early universe" gains more credence. We will return to this point in Section 11.

*The microwave background* This is perhaps the most striking (and the only!) evidence directly relating the origin of the universe to the present epoch. If one adopts the above picture, the universe started in a *big bang* with a hot radiation dominated phase. Soon after the creation H, D, He were formed. The expansion cooled the temperature of the universe and eventually matter and radiation decoupled. It is argued (but by no means clearly established) that the decoupling took place and galaxies began to be formed at redshifts $\sim 1000$. The matter dominated phase has continued to this day. This is the so-called standard canonical picture of the universe much in vogue today. And one of its predictions—which was originally made by Gamow (1948) —is that there should be a radiation background—a cooled relic of the hot era—even today.

Such a background was first observed unexpectedly by Penzias and Wilson (1965) at 7 cm. The results of many subsequent observations at higher and lower wavelengths are shown in Figure 2.4, as an intensity temperature curve (see Melchiorri and Maiani (1985) for reviews and Johnson and Wilkinson for a recent accurate measurement of the background temperature and for other references to recent work).



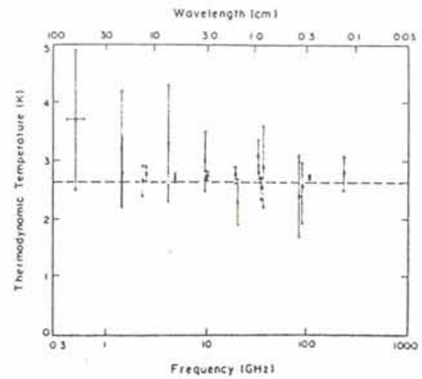



FIGURE 2.4 A composite spectrum of the microwave background combining the data at various frequencies. The dashed curve corresponds to the black body spectrum at $T = 2.7$ K.

*The very early universe* The successes of the early radiation dominated phase of the standard model inspired cosmologists to be more daring and to attempt studies of the physics of the universe at epochs even closer to the $t = 0$ instant than the period 1-200 sec discussed by Gamow *et al.* These studies necessarily require the physics of matter at energies far exceeding the rest mass energies of hadrons. For example, if we express particle energies $E$ as equivalent thermodynamic temperature $T$ by the relation $E = kT$, then Eq. (90) leads us progressively to $E \propto t^{-1/2} \to \infty$ as we approach the initial epoch.

The particle physicists in the meanwhile have independently felt the need to study physics at such high energies. For, guided by the successes of the electroweak theory of S. Weinberg (1967) and A. Salam (1968), the physicists today are looking for a unified theory of all interactions. Already, the attempts to unify the strong, the weak and the electromagnetic interactions suggest that one must study particle energies as high as $\sim 10^{15}$ times the hadronic rest mass energy. No man made accelerator can in the foreseeable future generate particle energies of this order. Thus experimental high energy physics appears to have reached a dead end so far as the unification programme is concerned unless one resorts to cosmology.

For, the extension of (90) to $t \lesssim 10^{-35}$ sec leads us to particle energies of such high order. This very early phase in the standard model is probably the only situation wherein the full unified theory of physics (if such a theory exists!) would have had a role to play. Not surprisingly therefore cosmologists and particle physicists have joined hands to explore the consequences of the very early universe.

Since as yet there is no accepted unified theory, work on the very early universe cannot be considered 'standard' in the sense of this section. In a subject which is rapidly changing with time (for fashions at CERN are more transient than those in Paris) it will be hard to do justice to this field. However, we will review some highlights in a later section (cf. Section 10).

For details of the present work on the very early universe see article by J. D. Barrow in this volume.

## 2.11 The $\lambda$-cosmologies

In his heuristic derivation of the field equations Einstein had arrived at the possibility of including an extra-term $\lambda g_{ik}$ ($\lambda$ = constant) on the left hand side of the equations:

$$R_{ik} - \tfrac{1}{2} g_{ik} R + \lambda g_{ik} = -\kappa T_{ik}. \tag{92}$$

Since $g_{ik;l} \equiv 0$, the divergence of $T_{ik}$ still follows from the field equations. In Hilbert's derivation the change in the action integrand

$$R \to R + 2\lambda \tag{93}$$

gives the same answer. In this case we may look upon $\lambda$ as an undetermined Lagrange multiplier introduced for a subsidiary condition that the four-volume of $V$

$$\int_V \sqrt{(-g)}\, d^4x = \text{constant}. \tag{94}$$

What is the physical significance of $\lambda$? For $\lambda > 0$ (which turns out to be the more interesting case) the $\lambda$-term has the meaning of a repulsive force which, in the Newtonian approximation increases in proportion to distance. This force is therefore important for large distances, i.e., for cosmology.

In the Robertson–Walker space-times the field equations (92) become

$$\left.\begin{aligned} 2\frac{\ddot{S}}{S} + \frac{\dot{S}^2 + k}{S^2} - \lambda &= 0, \\ 3\frac{\dot{S}^2 + k}{S^2} - \lambda &= 8\pi G\varrho. \end{aligned}\right\} \tag{95}$$

It is easy to check that these equations permit a static solution of the form

$$k = +1, \quad \varrho = \text{constant} = \varrho_0, \quad S = \text{constant} = S_0 \tag{96}$$

where

$$\lambda = \frac{1}{S_0{}^2} = 4\pi G\varrho_0. \tag{97}$$



This model, now known as the *Einstein universe*, was introduced by Einstein (1917) long before the Hubble redshift data became available. Indeed the primary motivation of Einstein in introducing the $\lambda$-term in the equations was to obtain a *static* model of the universe (which the equations (72) and (73) clearly do not permit). If $\varrho_0$ is as small as $\sim 10^{-29}\,\mathrm{g\,cm^{-3}}$, the value of $\lambda$ is so small as to leave all the solar system tests practically unaffected.

Einstein had hoped that this model of the universe would turn out to be unique. However, shortly after this work, de Sitter (1917) produced another solution of (95) in which

$$k = 0, \quad \varrho = 0, \quad S = \exp Ht, \quad \lambda = 3/H^2. \tag{98}$$

This was the model of an empty expanding universe (motion without matter) in contrast to Einstein's non-empty static universe (matter without motion).

However, once expanding universes are admitted a whole range of models is possible. Some of these are illustrated in Figure 2.5 schematically. For $\lambda$ close to the critical value which gives the Einstein universe (shown by $E$) we can have a "nearly static" universe. Such a universe may have originated from a big bang and will subsequently expand. This model was suggested by Eddington (1930) and Lemaitre (1927) and is shown in the figure by $EL$. The expanding phase asymptotically resembles the de Sitter model (*ds* in the

THE λ - COSMOLOGIES

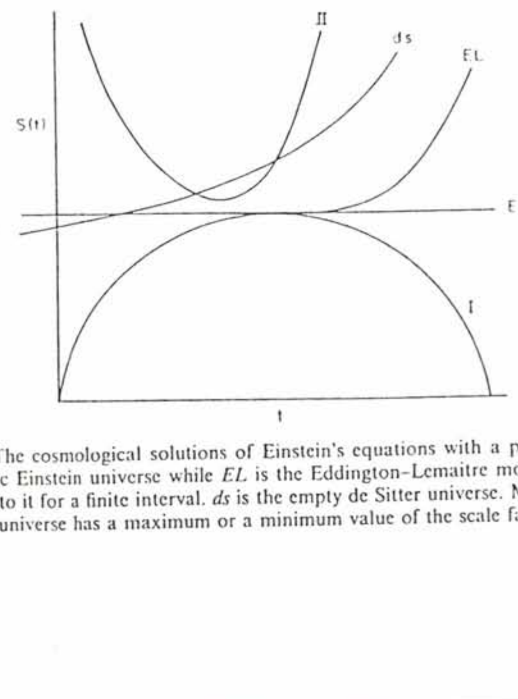



FIGURE 2.5 The cosmological solutions of Einstein's equations with a positive $\lambda$-term. $E$ denotes the static Einstein universe while $EL$ is the Eddington–Lemaitre model which comes arbitrarily close to it for a finite interval. *ds* is the empty de Sitter universe. Models I or II are cases where the universe has a maximum or a minimum value of the scale factor.

figure). Models I and II are cases of expansion followed by contraction and vice versa. It should be mentioned that the Einstein universe is unstable. On perturbation it either contracts to $S = 0$ or expands to $S = \infty$.

The $\lambda$-cosmologies today have the main advantage that they contribute an additional parameter ($\lambda$) to the cosmological problem. If a pooling of all available cosmological data (Gunn and Tinsley, 1975) indicates that the standard models are not viable, then the introduction of $\lambda$ may be taken seriously. Apart from this however, the case for it rests on a somewhat shaky ground and the use of $\lambda$ today would be considered a "non-standard" procedure by most cosmologists.

### 2.12 Anisotropic cosmologies

Although the present observational evidence for anisotropy in the large scale structure of the universe is not such as to necessitate abandoning the Robertson–Walker models, there have been many theoretical investigations of anisotropic cosmologies. We do not wish to go into details here since excellent and comprehensive discussions are available in the literature (Ellis, 1971; McCallum, 1973). We will outline some of the main features of these investigations which would be referred to elsewhere in this work.

As in the Weyl postulate it is assumed that there exists a unique normalized time like vector field $u^i$ representing the typical velocity of a *fundamental observer* in the cosmological space-time. It is *not* however required that there is a system of hypersurfaces to which $u^i$ is orthogonal. (Indeed this is where the concept of spin comes in as will be seen later.)

It is convenient to define a *projection tensor*

$$h_{ik} = g_{ik} - u_i u_k \tag{99}$$

with the properties

$$h_{ik} u^k = 0, \qquad h_i^{\ k} h_k^{\ l} = h_i^{\ l}, \qquad h_i^{\ i} = 3. \tag{100}$$

It is clear that for a fundamental observer with co-ordinates $x^i$ the space-like separation of a nearby point $x^i + \delta x^i$ is $(-h_{ik}\delta x^i \delta x^k)^{1/2}$ while the timelike separation is $(u_i \delta x^i)^{1/2}$.

By comparing the velocity vectors at two neighbouring fundamental observers, it is possible to introduce the concept of spin, shear and volume expansion in an obvious generalization of the Newtonian concepts:

$$u_{i;k} = \omega_{ik} + \sigma_{ik} + \tfrac{1}{3}\theta\, h_{ik} - u_{i;l}u^l u_k. \tag{101}$$

Here $\omega_{ik}$ is the spin tensor, $\sigma_{ik}$ the shear tensor and $\theta$ the volume expansion factor. Explicitly we have

$$\omega_{ik} = v_{[ik]},\ \sigma_{ik} = v_{(ik)} - \tfrac{1}{3}\theta h_{ik}, \tag{102}$$



where

$$v_{ik} = h_i^{\ l} h_k^{\ m} u_{l;m}, \qquad \theta = u^l_{\ ;l}. \tag{103}$$

Suppose we examine the displacement of a fundamental observer $G$ relative to a fundamental observer O. Assuming **OG** is small we express it as $n^i \delta l$ where $n^i$ is a normalized *space-like* vector. Then we get the anisotropic Hubble law:

$$(\delta l)^{\cdot} = (\sigma_{ik} n^i n^k - \tfrac{1}{3}\theta)\delta l. \tag{104}$$

Here ($^{\cdot}$) denotes differentiation in a timelike sense measured by O. The rate of change of direction is

$$h_i^{\ k} \dot{n}_k = \{\omega_i^{\ k} + \sigma_i^{\ k} - (\sigma_{lm} n^l n^m) h_i^{\ k}\} n_k. \tag{105}$$

We may identify $\theta/3$ with the direction-averaged Hubble constant. The shear tensor satisfies $\sigma_{ik} u^k = 0$ and has the magnitude

$$\sigma = |\sqrt{(\tfrac{1}{2}\sigma^{ik}\sigma_{ik})}|. \tag{106}$$

We may similarly define the local spin $\omega^i$ with the help of the Levi–Civita tensor:

$$\omega^i = \tfrac{1}{2}\varepsilon^{iklm} u_k \omega_{lm}, \tag{107}$$

and denote the magnitude of spin by $\omega$.

For a general discussion of anisotropic models within the framework of Einstein's equations the following energy momentum tensor is useful:

$$T_{ik} = \mu u_i u_k + (q_i u_k + q_k u_i) + p h_{ik} + \pi_{ik}. \tag{108}$$

Here $\mu$ = energy density of matter, $p$ = isotropic pressure, $\pi_{ik}$ = anisotropic stresses (e.g., viscous stresses) and $q_i$ = the energy flux relative to the mean fluid velocity $u_i$. The identity $T^{ik}_{\ \ ;k} \equiv 0$ gives two equations

$$\left.\begin{aligned} &\dot{\mu} + (\mu + p)\theta + \pi^{ik}\sigma_{ik} + q^k_{\ ;k} + q^i \dot{u}_i = 0; \\ &(\mu + p)\dot{u}_i + h_i^{\ k}(p_{,k} + \pi^l_{\ k;l} + \dot{q}_k) + (\omega_i^{\ k} + \sigma_i^{\ k} + \tfrac{4}{3}\theta h_i^{\ k})q_k = 0. \end{aligned}\right\} \tag{109}$$

The general restrictions on the equations are of the type which require the *positivity* of energy density. The following types are normally imposed from physical considerations:

$$\mu + p > 0, \quad \mu + 3p > 0. \tag{110}$$

For a perfect fluid we have

$$\pi_{ik} = 0, \quad q_i = 0. \tag{111}$$

*Raychaudhuri's equation* One of the main motivations for the study of anisotropic cosmologies was to see whether the space-time singularity inherent in the standard Friedmann models could be avoided by relaxing the symmetry requirements. For instance, if the universe were homogeneous but not isotropic, e.g., if it had spin, would the resulting centrifugal force prevent the collapse of the universe into a singularity?

A clue to this problem came from an important equation derived by A. K. Raychaudhuri (1955) from Einstein's field equations. For the energy tensor given by (108) the equation may be written in the form

$$3\frac{\ddot{R}}{R} = 2(\omega^2 - \sigma^2) + \dot{u}^k{}_{;k} - \tfrac{1}{2}(\mu + 3p) + \lambda, \qquad (112)$$

where we have written

$$R^3 = \exp\int^t \theta\, dt. \qquad (113)$$

Thus $R$ is an average linear scale factor.

Notice that for a non-singular solution $R$ must have a positive minimum, i.e., $\ddot{R} > 0$ at $\dot{R} = 0, R > 0$. Although $\lambda > 0$, it gets dominated by the density term which is large and negative. The shear term is also negative. The rotation term however, seems to hold out the hope of achieving a nonsingular solution.

There is a similar equation to (112) for spin rate of change:

$$h_k{}^i(R^2\omega^k)^{\cdot} = \sigma_k{}^i(R^2\omega^k) + \frac{R^2}{2}\varepsilon^{ijk}u_j\dot{u}_k \qquad (114)$$

Thus if we had $\sigma_{ik} = 0$, $\dot{u}_k = 0$ then for $\omega \neq 0, \theta \neq 0$ we get

$$(R^2\omega^k)^{\cdot} = 0, \quad \text{i.e.,} \quad \omega \propto 1/R^2. \qquad (115)$$

Also, for dust ($p = 0$ in addition to perfect fluid restrictions) we have $\mu \propto R^{-3}$. Thus the Raychaudhuri equation integrates to the form

$$3\dot{R}^2 + \frac{2\Omega^2}{R^2} - \frac{A}{R} - \lambda R^2 = B \qquad (116)$$

where $A$, $B$, $\Omega^2$ are constants. Thus $R$ is prevented from going to zero.

Unfortunately, it can be shown (Ellis, 1967) that general relativity *does not* permit such solutions ($\sigma_{ik} = 0$, $\dot{u}_k = 0$, $\theta \neq 0$) for $p = 0$ or for $\partial p/\partial\mu \neq 0$. In fact the general rule seems to be that you can have shear without rotation but not *vice versa*. Thus it seems that a proof that general relativity does permit non-singular cosmological models cannot proceed along these lines. In fact the Penrose–Hawking theorems (Hawking and Ellis, 1973) seem to rule out non-singular models altogether unless some very esoteric conditions are imposed. Below we describe a few specific models.



*Godel's model* The interest in anisotropic spinning universes was initiated by a model of Kurt Godel (1949) with the following line-element for a dust filled universe:

$$ds^2 = dt^2 - 2\exp(x^1)\,dt\,dx^2 - (dx^1)^2 + \tfrac{1}{2}\exp[2x^1](dx^2)^2 - (dx^3)^2. \qquad (117)$$

In this model $\mu = 2\omega^2$, $\lambda = -\omega^2$, $\sigma = 0$, $\theta = 0$ and $\dot{u}_a = 0$. This model is unrealistic because (i) it has no expansion and hence no nebular redshift and (ii) it has closed time like lines.

Godel proposed this model as an anti-Machian model, i.e., it was a model in which the local inertial frame is such that in it the distant stars are seen to be rotating. (See Section 4 for a further discussion of this). Later work of others (Ozsvath and Schucking, 1962) showed that such anti-Machian models without the above two "defects" of Godel's models and also without the $\lambda$-term can be obtained from general relativity.

*Homogeneous models* Spatially homogeneous but anisotropic models have been systematically classified and studied during the last ten years. The basis for classification is the underlying group of motion. For homogeneity the group must act transitively on the whole of a typical space section $t$ = constant (cf. Weyl postulate). Thus the group must have at least 3 dimensions and it must act transitively on spacelike orbits. It turns out that barring an exceptional case of a four dimensional group all possible cases reduce to three dimensional transitive groups. A classification on the basis of all such groups was done by Bianchi (1918) and the different cosmological models are known under Bianchi types I–IX. Behr and others have introduced a modified scheme. For a detailed discussion of this classification see MacCallum (1973).

The Bianchi type-I model is given by the line element

$$ds^2 = dt^2 - X_1{}^2(t)(dx^1)^2 - X_2{}^2(t)(dx^2)^2 - X_3{}^2(t)(dx^3)^2. \qquad (118)$$

For a dust filled universe, the functions $X_1$, $X_2$, $X_3$ are given by

$$X_\mu(t) = S\left(\frac{t^{2/3}}{S}\right)^{2\sin a_\mu} \qquad (119)$$

where

$$S^3(t) = X_1X_2X_3 = \tfrac{9}{2}Mt(t + \Sigma), \ (M, \Sigma \text{ constants}) \qquad (120)$$

and for an arbitrary $a$

$$a_\mu = a + \left(\frac{\mu - 1}{3}\right)\cdot 2\pi. \qquad (121)$$

For different values of the constants the model ha a pancake singularity or a line singularity.

We will not discuss here the details of the various Bianchi types; cf. MacCallum (1973) for such a discussion. These models have been useful in trying to understand how and why the space-time singularity exists in the past, to what extent the presently observed isotropy is a matter of initial conditions, and in relating the presently observed abundances of helium and deuterium to the early history of the universe.

## 3. NEWTONIAN COSMOLOGY

By Newtonian cosmology is implied cosmology based on Newtonian concepts of space and time, on Newtonian mechanics and on the Newtonian inverse square law of gravitation. It might be asked: "why go back to outdated ideas when we *know* then to be incorrect?" Special relativity has supplanted Newtonian mechanics while gneral relativity is accepted by many to be superior to Newtonian gravitation both conceptually and observationally. Nevertheless Newtonian cosmology has played an interesting role in clarifying the limitations of Newtonian concepts. The similarity between Newtonian and relativistic models is striking and the relative simplicity of the former has been helpful in the study of cosmological models.

Historically speaking, no significant progress seems to have been made in Newtonian cosmology in the pre-relativity days. The reason seems to be that Newtonian equations failed to describe a static homogeneous universe. Even if such a universe existed in an initiall static state, the mutual gravitational pull of its constituents would pull them together so that a contraction would result. Today we know that the universe is *not* static and hence a failure to produce static models need not be a drawback to a cosmological theory. Even general relativity had the same limitation and in the pre-Hubble days Einstein felt inclined to introduce a $\lambda$-term to generate static universe.

Ironically, *after* the discovery of the expanding universe and the established success of Friedmann models of general relativity, Milne and McCrea (1934) demonstrated that with suitable assumptions Newtonian cosmology can generate identical models! We begin by a discussion of this important work.

### 3.1 Homogeneous and isotropic cosmologies

Following Milne and McCrea, we assume the universe to be infinite, with a Euclidean space-geometry and a uniformly flowing *absolute* time $t$, as conceived of by Newton. Suppose a typical observer O sets up Cartesian co-ordinates $x\mu$, $\mu = 1, 2, 3$, with himself at the origin. The co-ordinates of a typical point $P$ will be denoted by $x\mu$ and its position vector by $\mathbf{r}$.



At each point $P$ we will place a "fundamental observer" as in relativistic cosmology (cf. Weyl postulate) and ascribe to him a definite 3-velocity $\mathbf{v} = \mathbf{v}(\mathbf{r}, t)$. The existence of the velocity field implies an orderliness in the motion of the fundamental observers. We also define $\varrho(\mathbf{r}, t)$ and $p(\mathbf{r}, t)$ as the density and pressure at $P$ at time $t$.

We now introduce the cosmological principle by requiring that the functions $\mathbf{v}$, $\varrho$ and $p$ as determined by any fundamental observer should be the same at any given time. To fix ideas suppose another fundamental observer O′ located at $a$ relative to O makes the same measurements of these quantities at $P$. Then he should find the velocity of the fundamental observer at $P$ to be $\mathbf{v}(\mathbf{r}, t)$ and the density and pressure to be respectively $\varrho(\mathbf{r}, t)$ and $p(\mathbf{r}, t)$, where

$$\mathbf{r}' = \mathbf{r} - \mathbf{a}. \qquad (1)$$

We therefore have

$$\varrho(\mathbf{r} - \mathbf{a}, t) = \varrho(\mathbf{r}, t) \equiv \varrho(t), \qquad (2)$$

$$p(\mathbf{r} - \mathbf{a}, t) = p(\mathbf{r}, t) \equiv p(t). \qquad (3)$$

As for the velocity, the usual law of relative velocity gives

$$\mathbf{v}(\mathbf{r} - \mathbf{a}) = \mathbf{v}(\mathbf{r}) - \mathbf{v}(\mathbf{a}). \qquad (4)$$

The most general solution of (4) is expressible in the tensor form

$$v_\mu(\mathbf{r}, t) = H_{\mu\nu}(t) \cdot x_\nu \qquad (5)$$

where $H_{\mu\nu}(t)$ is a second rank Cartesian tensor function of time only.

The equations (2, 3, 5) express the "homogeneity" property of the cosmological principle. If in addition we also demand isotropy at every point (as seen by the fundamental observer) then we must have

$$H_{\mu\nu}(t) = H(t) \cdot \delta_{\mu\nu} \qquad (6)$$

and (5) reduces to the *Hubble law*:

$$\mathbf{v} = H(t) \cdot \mathbf{r}. \qquad (7)$$

This velocity field can be integrated in the form

$$\mathbf{r} = S(t)\mathbf{r}_0 \qquad (8)$$

where $S(t)$ is given by

$$\frac{1}{S}\frac{dS}{dt} = \frac{\dot{S}}{S} = H(t), \quad S(t_0) = 1. \qquad (9)$$

Thus the observer at $\mathbf{r}_0$ was at $\mathbf{r}$ at time $t$.

Next we use Euler's equations of motion and continuity for the "cosmological fluid". The latter gives

$$0 = \frac{\partial \varrho}{\partial t} + \operatorname{div}(\varrho \mathbf{v}) = \dot{\varrho} + 3\frac{\dot{S}}{S}\varrho,$$

$$\text{i.e.,} \quad \varrho S^3 = \text{constant} = \varrho(t_0) = \varrho_0 \text{ (say).} \tag{10}$$

The equation of motion gives, for an external force $F$ per unit mass

$$\dot{\mathbf{v}} + (\mathbf{v} \cdot \nabla)\mathbf{v} = \mathbf{F}. \tag{11}$$

Since F is the gravitational force, we get by Gauss's theorem

$$\operatorname{div} \mathbf{F} = -4\pi G\varrho. \tag{12}$$

Taking divergence of (11) then gives

$$\dot{H}(t) + H^2(t) = -\frac{4\pi G\varrho}{3},$$

$$\text{i.e.,} \quad \ddot{S} = -\frac{4\pi G\varrho}{3} S. \tag{13}$$

Note that the equations (2.72, 73) of relativistic cosmology give exactly the same answer!

However, if we substitute (13) in (11) we get

$$\mathbf{F} = -\frac{4\pi G\varrho}{3}\mathbf{r}. \tag{14}$$

How is this to be interpreted? It is the force on the observer at $P$ due to the mass contained in the sphere of radius $OP$, centred at $O$. If we draw spherical shells concentric with $O$ but of varying radii, the inverse square law of gravitation tells us that the shells with radii exceeding $OP$ (see Figure 3.1) make no contribution to the gravitational force of the observer at $P$. Hence the above result.

However, looked at from the point of view of the observer at $P$, the universe being isotropic, it is absurd to image a well directed force F at $O$! By symmetry there can be no force at $P$. This ambiguity arises from our attempts to apply the Newtonian laws to an infinite system and is one of the conceptual drawbacks of Newtonian cosmology. See discussions of this by Layzer (1954) and McCrea (1954).

Returning to (13) we see why a static universe is impossible. This equation only admits the trivial case of the empty universe if we insist on the solution $S$ = constant. However, as mentioned earlier the equations (10 and (13) are



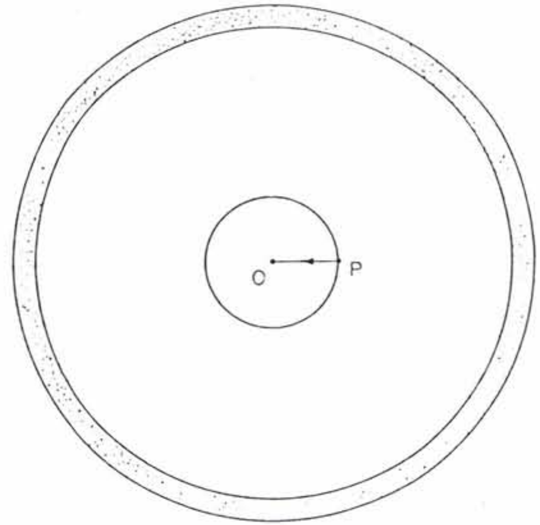



FIGURE 3.1 A homogeneous spherical shell of matter centred at O does not produce any gravitational field at a typical interior point $P$. The net field at $P$ is therefore limited to shells of radius *smaller* than $OP$. This field is directed along $PO$.

exactly the same as in relativistic cosmology. The first integral of these is

$$\dot{S}^2 = \frac{8\pi G\varrho_0}{3S} - k, \qquad k = \text{constant}. \tag{15}$$

This equation is analogous to (73) of Section 2. Notice, however, that the constant $k$ has only a dynamical significance in Newtonian cosmology whereas in relativistic cosmology it also has a geometrical significance (the sign of curvature of hypersurfaces $t$ = constant).

How does redshift arise in this model? Consider a ray of light emitted by a fundamental observer at $P_1(\mathbf{r}_1)$ at $t = t_1$ which is received by the observer at $O(\mathbf{r} = \mathbf{a})$ at $t = t_0$. This ray has a velocity $c$ relative the absolute standard of rest as judged by an fundamental observer. When it passes a typical point $P(\mathbf{r})$ en route to $O$ (see Figure 3.2) it has a velocity $c$ towards $O$ as seen by $P$. However, $P$ is moving relative to $O$ with a velocity $H\mathbf{r}$. Hence the speed of light determined in the Newtonian way along $PO$, as "seen" by $O$ is $c$-$Hr$. Thus the equation of motion of the light ray is

$$\frac{dr}{dt} = Hr - c = r\frac{\dot{S}}{S} - c.$$

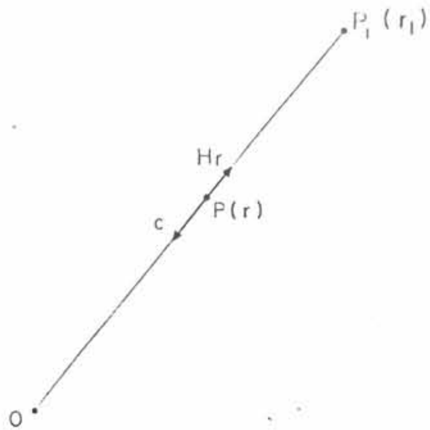



FIGURE 3.2 The light signal from $P_1$ to $O$, when at an intermediate point $P$, has two velocity components within the Newtonian framework. The component $c$ is directed towards $O$ while the component $Hr$ due to the expansion at $P$ is directed away from $O$.

This integrates to

$$\frac{r_1}{S(t_1)} = c\int_{t_1}^{t_0} \frac{dt}{S(t)}. \qquad (16)$$

Hence if two signals are emitted successively at $t_1$ and $t_1 + \Delta t_1$ and they are received respectively at $t_0$ and $t_0 + \Delta t_0$ we have from (16)

$$\frac{\Delta t_1}{S(t_1)} = \frac{\Delta t_0}{S(t_0)}.$$

This leads to a redshift identical to the relativistic formula (2.76):

$$1 + z = \frac{S(t_0)}{S(t_1)}. \qquad (17)$$

As discussed by Bondi (1961) extensively this approach essentially reproduces (with suitable interpretation) all the general properties of standard cosmologies. Even the $\lambda$-term can be incorporated as a force of repulsion proportional to distance and the $\lambda$-cosmologies can be reproduced as in Section 2.

## 3.2 Homogeneous and anisotropic cosmologies

There have been several investigations generalizing the above picture developed by Milne and McCrea to cosmologies with shear and rotation. Broadly speaking there have been two different types of approaches. Of these the approach discussed by Heckmann and Schucking (1955, 1956) and Raychaudhuri (1955) will be discussed first.



3.2.1 *The potential function approach* Heckmann and Schucking (1955, 1956) wrote the gravitational force in the form

$$F_\mu = \phi_{;\mu}, \qquad (18)$$

where $\phi$ is the gravitational potential satisfying the Poisson equation

$$\nabla^2\phi = 4\pi G\varrho. \qquad (19)$$

Equation (19) could also be written as div F $= 4\pi G\varrho$, provided we recognize that F is the gradient of a scalar, so that the integrability conditions

$$\text{curl F} = O \qquad (20)$$

must be satisfied.

Thus we have the following three equations to describe the general behaviour of a homogeneous but anisotropic universe:

$$\text{Eq. of continuity:} \quad \dot{\varrho}/\varrho + \text{div } \mathbf{v} = 0 \qquad (21)$$

$$\text{Eq. of motion:} \quad \dot{\mathbf{v}} + (\mathbf{v}\cdot\nabla)\mathbf{v} = \mathbf{F}, \qquad (22)$$

$$\text{Eq. of gravitation:} \quad \nabla^2\phi + \lambda = 4\pi G\varrho. \qquad (23)$$

It should be remembered in (21) and (22) that $\varrho$ and $p$ are functions of $t$ only while in (23) the $\lambda$-term has also been included. The condition (20) can be rewritten in the form

$$\phi_{,\mu\nu} = \phi_{,\nu\mu}. \qquad (24)$$

In this approach therefore of the nine components of the tensor there are really only 6 algebraically independent components by virtue of (24). Equation (23) reduces the number further by 1. Hence Heckmann and Schücking had effectively five unknown quantities still to be determined. Physically we may link the arbitrariness of the problem to the arbitrariness of the choice of accelerated frames at each fundamental observer. Such frames lead to fictitious gravitational forces *without sources*.

As in relativity (cf. Section 2) the velocity vector can be used to define tensors for shear $q_{\lambda\mu}$ and spin $\omega_{\lambda\mu}$ and the spin vector $\omega_\lambda$. Writing the velocity-distance relation as in (5) we get

$$\left.\begin{aligned} \omega_{\mu\nu} &= \tfrac{1}{2}(H_{\mu\nu} - H_{\nu\mu}), \\ q_{\mu\nu} &= \tfrac{1}{2}(H_{\mu\nu} + H_{\nu\mu}) - \tfrac{1}{3}\delta_{\mu\nu}H_{\lambda\lambda}, \\ \omega_\lambda &= \tfrac{1}{2}\varepsilon_{\lambda\mu\nu}\omega_{\mu\nu}. \end{aligned}\right\} \qquad (25)$$

From these definitions and the equations (21)–(23), Heckmann and Schücking deduced the following relations:

$$\begin{aligned} R(t) &= \exp \int \tfrac{1}{3} H_{\lambda\lambda}\, dt, \\ \tfrac{4}{3}\pi\varrho R^3 &= \mathcal{M}, \\ \frac{d}{dt}(R^2\omega_\lambda) &= R^2\omega_\mu q_{\mu\lambda}, \\ \ddot{R} - \frac{R}{3}(\lambda - q_{\mu\nu}q_{\mu\nu}) - \tfrac{2}{3}R\omega_\mu\omega_\mu + \frac{G\mathcal{M}}{R^2} &= 0. \end{aligned} \tag{26}$$

Now utilizing the indeterminacy of the potential function, Heckmann and Schücking looked for a special class of solutions which are *non-singular*, i.e., in which $R \not\to O$. Their hope was that if such solutions exist in the Newtonian framework we may look for their analogues in relativistic cosmology.

The situation in relativity as mentioned in Section 2.12 is that the shear tensor *does not* help in this project. In Newtonian framework, however, it is possible (in the above approach) to set

$$q_{\mu\nu} = 0. \tag{27}$$

Then we have 5 additional equations to complete the solution of the problem. Following this approach, Heckmann and Schücking gave the solution

$$R^2\omega_\lambda = \text{constant}, \quad R\omega_\nu\omega_\nu \propto \frac{1}{R^3} \tag{28}$$

At small $R$ the rotation term dominates and we get a minimum of $R$ for $R > O$. It is possible to obtain for $\lambda \neq O$ the analogue of Godel's model (cf., Section 2) in the Newtonian framework.

We will return to this approach after discussing the second approach which is somewhat more restrictive and permits fewer models than in the present case.

3.2.2 *The gravitational force approach* This approach was initiated by Narlikar (1963) and explored in detail by Davidson and Evans (1973) and is essentially based on a generalization of the scope of Eq. (14).

The central assumption of this approach is that in an infinite homogeneous universe we can still adopt (14) on the basis of the inverse square law of Newton. The universe at any given time $t$ is uniformly dense and can be divided into concentric spherical shells as discussed earlier. Thus we begin by substituting (14) into the Euler equation of motion but take for $v$ the general



anisotropic Hubble relation (5) derived earlier. Differentiating (5) with respect to $t$ and using $v_\mu = \dot{x}_\mu$ we get

$$\ddot{x}_\mu = A_{\mu\nu}x_\nu \tag{29}$$

where

$$A_{\mu\nu} = \dot{H}_{\mu\nu} + H_{\mu\lambda}H_{\lambda\nu}. \tag{30}$$

The Euler equation of motion and the equation of continuity give under the present assumptions the following two relations respectively:

$$\left.\begin{aligned} A_{\mu\nu} \equiv \dot{H}_{\mu\nu} + H_{\mu\lambda}H_{\lambda\nu} &= -\frac{4\pi G\varrho}{3}\delta_{\mu\nu}, \\ \dot{\varrho} + \varrho H_{\lambda\lambda} &= 0. \end{aligned}\right\} \tag{31}$$

To integrate these we follow a procedure similar to that of Milne and McCrea. Write the solution of (5) in the form

$$x_\mu = a_{\mu\nu}(t)x_\nu^0 \tag{32}$$

where at some specified time $t = t_0$, $x_\mu = x_\mu^0$, i.e.,

$$a_{\mu\nu}(t_0) = \delta_{\mu\nu}. \tag{33}$$

We then have

$$\dot{a}_{\mu\nu} = H_{\mu\lambda}a_{\lambda\nu} \tag{34}$$

Writing

$$\Delta = \det \|a_{\mu\nu}\| \tag{35}$$

we get $\dot{\Delta}/\Delta \leqslant = H_{\lambda\lambda}$, so that (31) integrates to

$$\varrho\Delta = \text{constant} = \varrho_0 \tag{36}$$

with $\varrho_0 = \varrho(t_0)$. Finally, the equation of motion reduces to

$$\Delta\ddot{a}_{\mu\nu} = -\frac{4\pi G}{3}\varrho_0 a_{\mu\nu}. \tag{37}$$

The general Newtonian problem is therefore contained in the solution of (36) and (37). Narlikar (1963) showed that it is possible to eliminate *all* $a_{\mu\nu}$ and obtain a fourth order non-linear differential equation for $\Delta$. Writing a dimensionless time co-ordinate

$$\tau = \left(\frac{4\pi G}{3}\varrho_0\right)^{1/2} t \tag{38}$$

and denoting $d/d\tau$ by a dash, this equation is

$$\Delta^2\Delta'''' + 7\Delta\Delta'' - 4\Delta'^2 + 9\Delta = 0. \tag{39}$$

Narlikar used this equation to investigate whether there exists a Newtonian singularity, i.e., an epoch when $\varrho \to \infty$, $\Delta \to 0$. With the help of a series of transformations

$$\Delta = F^2, \quad \left(\frac{dF}{d\tau}\right)^2 = X(F), \quad F = e^U, \quad \frac{dX}{dU} = Y(X),$$

$$\text{i.e.,} \quad X = (F')^2, \quad Y = 2FF'', \tag{40}$$

the fourth order equation (39) can be reduced to a second order equation:

$$XY^2\frac{d^2Y}{dX^2} + XY\left(\frac{dY}{dX}\right)^2 + \left(X + \frac{Y}{2}\right)Y\frac{dY}{dX}$$
$$+ Y^2 - 2XY + 7Y - 2X + 9 = 0. \tag{41}$$

Since (39) has the property that if $\Delta(t)$ is a solution, so is $A^{-2}\Delta(A\tau + B)$ for arbitrary constants $A$ and $B$, it turns out that all solutions of this family are characterized by one curve in the $(X, Y)$ plane. Along a typical curve $C$ joining points $P_1$ and $P_2$ the $\Delta$-values at $P_1$ and $P_2$ are related by

$$\frac{\Delta_2}{\Delta_1} = \exp\int_{P_1}^{P_2}\frac{2dX}{Y}. \tag{42}$$

In the curves investigated by Narlikar (1963) asymptotically $Y = -2X$, so that the integral in (42) is negatively infinite as $P_2$ recedes to infinity: i.e., $\Delta_2 \to 0$ and a singularity results. It is also interesting to note that there is an exceptional curve which is the straight line $\zeta$:

$$Y = \tfrac{2}{3}X - 3 \tag{43}$$

and an exceptional point $(E)$, $(9/2, 0)$ which are singular solutions. The point corresponds to the Einstein-de Sitter solution (see Figure 3.3).

Davidson and Evans (1973) have pointed out that there exist solutions of (41) which Narlikar had missed, in which $\Delta$ moves from a minimum $[X = 0, Y > 0]$ to a maximum $[X = 0, Y < 0]$ before eventually moving to a singularity. In Figure 3.3 such a curve is shown by D. Since the curves could be continued to the past ($\tau < 0$) in a symmetrical manner, $D$ would represent a minimum with a maximum and a singularity of $\Delta$ on either side. The curve $C$ on the other hand has only a maximum and two singularities.

Davidson and Evans (1973) considered the solutions of (37) in detail rather than analyse the fourth order differential equation for $\Delta$. To this end they



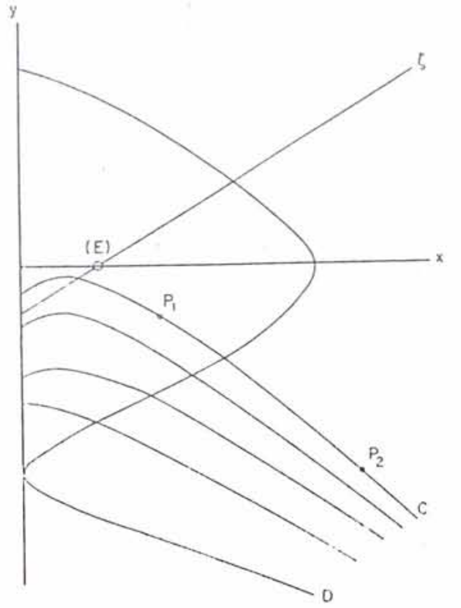



FIGURE 3.3 The solutions of the homogeneous models in the Newtonian framework as studied by Narlikar, Davidson and Evans. The point $E$ denotes the isotropic Einstein-de Sitter model. The curve $C$ denotes the case where as the point $P_2$ moves further away from $P_1$, it approaches singularity. In curve $D$, the approach to singularity is preceded by an epoch of minimum density.

defined the shear and spin tensors by

$$q_{\mu\nu} = H_{(\mu\nu)} - \tfrac{1}{3}H_{\lambda\lambda}\delta_{\mu\nu},$$
$$\omega_{\mu\nu} = H_{[\mu\nu]} \tag{44}$$

so that

$$H_{\mu\nu} = q_{\mu\nu} + \omega_{\mu\nu} + \frac{1}{3}\frac{\dot\Delta}{\Delta}\delta_{\mu\nu}, \tag{45}$$

and

$$H_{\mu\nu}H_{\mu\nu} = q^2 - 2\omega^2 + \frac{1}{3}\frac{\dot\Delta^2}{\Delta^2}, \tag{46}$$

where $q^2 = q_{\mu\nu}q_{\mu\nu}$ and $2\omega^2 = \omega_{\mu\nu}\omega_{\mu\nu}$. (Thus $\omega$ is the magnitude of the spin vector $\omega_\lambda = \frac{1}{2}\varepsilon_{\lambda\mu\nu}\omega_{\mu\nu}$.) The equation (30) in contracted form then leads to

$$3\frac{\ddot R}{R} = 2\omega^2 - q^2 - \frac{4\pi G}{R^3}\varrho_0 \tag{47}$$

with $R^3 = \Delta$. This is the same equation as derived by Heckmann and Schucking (1955) from their approach and is the Newtonian analogue of the Raychaudhuri equation (2.112). A little manipulation with these equations also gives

$$\frac{d}{dt}(R^2\omega_\lambda) = R^2 q_{\lambda\mu}\omega_\mu. \tag{48}$$

It is now possible to contrast the two approaches described so far.

In the Heckmann and Schucking approach the Poisson equation replaces the force equation (37). In this there is an extra freedom available, corresponding to the solutions of $\nabla^2\phi = 0$. The forces $\nabla\phi$ due to the sourceless potential equation may be interpreted as arising from arbitrarily chosen accelerated frames at different fundamental observers. To fix this part Heckmann and Schucking assumed that the five independent components of $q_{\mu\nu}$ could be assigned arbitrarily. They therefore looked for solutions with $q_{\mu\nu} = 0$, although $\omega_{\mu\nu} \neq 0$, $\dot{\Delta} \neq 0$. From (47) we see that such a situation may result in preventing the singular situation $R \to 0$. And with this analogy they conjectured that non singular solutions might exist in relativistic cosmology.

In the Narlikar–Davidson–Evans approach the force is fixed and $q_{\mu\nu}$ is determined by the dynamical equations (37) with given initial conditions. It is not therefore possible to have $q_{\mu\nu} = 0$; indeed even if $q_{\mu\nu} = 0$ at $t = t_0$, say, subsequently $q_{\mu\nu} \neq 0$ and shear develops. Thus whereas in the Heckmann–Schucking case $q_{\mu\nu} = 0$ in (48) leads to $\omega_\nu \propto R^{-2}$ and $\omega^2 \propto R^{-4}$ so that as $R \to O$ spin dominates and prevents a singularity (vis-a-vis the singularity-conducing gravitational term $\propto R^{-3}$). In the second approach $q_{\mu\nu} \neq 0$ does not permit the solution $\omega_\lambda \propto R^{-2}$. The second approach is therefore more analogous to the theory of relativity (cf. Section 2) where a similar conclusion was arrived at: *shearfree spinning universes are not possible.*

Davidson and Evans (1973) have analyzed several specific solutions of (37) and shown how they are analogous to the solutions in relativistic cosmology. For example a model spinning about the $x_3$ axis and axisymmetric about it will have

$$a_{\mu\nu} = \begin{bmatrix} x & u & 0 \\ u & x & 0 \\ 0 & 0 & z \end{bmatrix}, \quad H_{\mu\nu} = \begin{bmatrix} H & -\omega & 0 \\ \omega & H & 0 \\ 0 & 0 & K \end{bmatrix}. \tag{49}$$

The behaviour of $x, z, u, H, \omega, K$ can be calculated in terms of the initial conditions. In all cases $\Delta \to 0$ eventually, i.e., a singularity is reached. For example, a curve like $D$ of Figure 3.3 results from the initial condition $H_0 = 0$, $K_0 = 0$, $\omega_0 \neq 0$. Although shear is initially zero, it develops soon and eventually dominates over the spin-term as the singularity is reached.



With the help of this solution Davidson and Evans (1973) have given the following argument as to why $\Delta \to 0$ in all cases of Narlikar's equation (39). In this equation if at any given instant $\tau = \tau_0$, the values $\Delta_0$, $\Delta_0'$, $\Delta_0''$ and $\Delta_0'''$ are given then the function $\Delta(\tau)$ for $\tau \geqslant \tau_0$ is uniquely determined. Now suppose that $\Delta$ does not reach a zero, i.e., there is no singularity. Then there must exist a non-zero lower bound of $\Delta$. Without loss of generality set $\Delta = \Delta_0 > 0$ at this minimum and let $\tau = \tau_0$ at this instant. Then for a minimum at $\tau_0$, we need $\Delta_0' = 0$, $\Delta_0'' > 0$. However, if $\Delta_0'' = 0$ then a minimum requires $\Delta_0''' = 0$, $\Delta_0'''' > 0$, which from (39) is impossible. Hence the minimum can only have the following range of parameters:

$$\infty > \Delta_0 > 0,\ \Delta' = 0,\ \infty > \Delta_0'' > 0,\ \infty > \Delta_0''' > -\infty. \tag{50}$$

Now in (49), a particular case of axisymmetric solution, we have *three*, general parameters $H_0$, $\omega_0$ and $K_0$, apart from the scaling parameter $\Delta_0$ (which may be put equal to 1 for convenience). Given any $\Delta_0'$, $\Delta_0''$ and $\Delta_0'''$ we can always find an axisymmetric solution with the values equal to them. Thus for (50) we need

$$\Delta_0' = 2H_0 + K_0 = 0,\quad \Delta_0'' = 2\omega_0^2 - \tfrac{3}{2}K_0^2 - 3 > 0,$$
$$\text{and}\quad \Delta_0'''' = \cdot 6K_0\left(\omega_0^2 + \frac{K_0^2}{4}\right). \tag{51}$$

However, for this axisymmetric solution it can be explicitly shown that $\Delta \to 0$ at some epoch $\tau > \tau_0$. By the uniqueness argument the general case will also have this property ($\Delta \to 0$) even though the individual functions $a_{\mu\nu}$ may behave differently from the axisymmetric case. Thus the original assumption of no singularity leads to a contradiction.

Thus the second approach gives results more analogous to the results from relativity than the first approach. Davidson and Evans (1973) have calculated specific cases of models with shear and no rotation and have shown similarities with the Bianchi type I models of Section 2 with pancake singularities. These exist many close parallels between Newtonian models and the relativistic models of Bianchi types I, II, V and IX.

Davidson and Evans (1973) also discuss ages of the various axisymmetric models from observational considerations of Hubble's constant, mean density of matter in the universe, observed limits on anisotropy etc. They find that a pancake singularity in the past $\sim 1.5 \times 10^{10}$ years can be compatible with the observed data.

### 3.3 The cosmological principle

In the preceding discussions it has been tacitly assumed that the equivalent reference frames of fundamental observers are all similarly oriented. On this basis the linear velocity distance relation (5) was obtained. However, translations are not the only possible transformations between such reference frames in the Euclidean space. Evans and Davidson (1973) have generalized the scope of the cosmological principle by going over to the most general possible transformations between equivalent fundamental frames (EFF). Their approach is briefly described below.

Suppose $S$ and $S'$ are two EFFs and let $P$ and $P'$ be points which have the same position vectors respectively in $S$ and $S'$. Let $\mathbf{x}$ be the position vector of $P$ relative to $S$ and $\mathbf{x}^+$ be the position vector of $P'$ relative to $S$. Under a mapping

$$\mathbf{x} \rightarrow \mathbf{x}^+ = f(\mathbf{x}) \tag{52}$$

the point $P$ of $S$ is carried over to $P'$ of $S'$. It is assumed that $S$ and $S'$ are cartesian frames. Then we may regard the mapping $f$ as a translation of origin $O$ of $S$ to origin $O'$ of $S'$ followed by a rotation of axes. Let $\mathbf{a}$ be the position vector of $O'$ relative to $O$. Then at any given epoch the rotation of axes may be described by an orthonormal matrix $T(\mathbf{a})$ which depends on $\mathbf{a}$. Thus if a vector is measured as $\mathbf{u}$ in $S$ and $\mathbf{u}'$ in $S'$, we have

$$\mathbf{u} = T(\mathbf{a})\mathbf{u}'. \tag{53}$$

Then it is clear that the translation and rotation implied by (52) is given by

$$\mathbf{x}^+ = T(\mathbf{a})\mathbf{x} + \mathbf{a}. \tag{54}$$

Since $\mathbf{a} = \mathbf{O}$ should give $S' = S$, $\mathbf{x}^+ = \mathbf{x}$, we have

$$T(\mathbf{O}) = I(= \text{identity}). \tag{55}$$

Consider now *three* EFFs $S$, $S'$ and $S''$. Let $\mathbf{OO'}$ be the vector $\mathbf{a}$ as before, $\mathbf{O'O''}$ be $\mathbf{b}$ as measured in $S'$ and let a typical point $P$ have co-ordinates $\mathbf{c}$ in $S''$. Then as measured in $S'$,

$$\mathbf{O''P} = T(\mathbf{b})\mathbf{c}$$

and as measured in $S$

$$\mathbf{O''P} = T(\mathbf{a})T(\mathbf{b})\mathbf{c}.$$

However, we can do this calculation by another method. As measured in $S$ we have

$$\mathbf{O'O''} = T(\mathbf{a})\mathbf{b}.$$

Hence in $S$ we have

$$\mathbf{O'O''} = \mathbf{a} + T(\mathbf{a})\mathbf{b},$$



and hence, again in $S$

$$\mathbf{O''P} = T[\mathbf{a} + T(\mathbf{a})\mathbf{b}]\mathbf{c}.$$

Therefore

$$T[\mathbf{a} + T(\mathbf{a})\mathbf{b}] = T(\mathbf{a})T(\mathbf{b}). \tag{56}$$

Under what conditions do the transformations (54) form a continuous group in $V_3$? These conditions have been discussed by Petrov (1969) in the general transformations of the type

$$f(\mathbf{x}, \mathbf{a}) = T(\mathbf{a})\mathbf{x} + \mathbf{a}, \tag{57}$$

and it is easily verified by Evans and Davidson (1973) that these conditions are indeed satisfied because of (56). Thus (54) form a simply transitive group of transformations. It turns out that there are essentially only two types of transformations implied by this group in the Euclidean space.

Of these the simpler case implied by

$$T(\mathbf{a}) = I \tag{58}$$

gives the translational group of transformations which led to (5). The second type involves translations coupled with rotations. Without loss of generality we may take $x_3$ as the axis of rotation. Then we have

$$T(\mathbf{a}, t) = \begin{bmatrix} \cos[m(t)a_3] & -\sin[m(t)a_3] & 0 \\ \sin[m(t)a_3] & \cos[m(t)a_3] & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{59}$$

where $a_3$ = the $x_3$-component of the displacement vector $\mathbf{a}$ and $m(t)$ is an arbitrary function of time.

Given $T(\mathbf{a})$ we can compute the angular velocity of EFF : $S'$ relative to EFF : $S$. This is given by $\mathbf{n}(\mathbf{a})$ where

$$\eta_\lambda = \tfrac{1}{2}\varepsilon_{\lambda\mu\nu}\dot{T}_{\mu\sigma}(\mathbf{a}, t)\dot{T}_{\nu\sigma}(\mathbf{a}, t). \tag{60}$$

If a particle of the cosmological fluid has the position vector $\mathbf{b}$ in $S'$ and a velocity vector $\mathbf{v}(\mathbf{b})$ in $S'$ then a reasoning similar to the derivation (56) tells us that the velocity of this particle as measured in $S$ will be

$$\mathbf{v}[\mathbf{a} + T(\mathbf{a})\mathbf{b}] = \mathbf{v}(\mathbf{a}) + T(\mathbf{a})\mathbf{v}(\mathbf{b}) + \mathbf{n}(\mathbf{a}) \times T(\mathbf{a})\mathbf{b}. \tag{61}$$

As before, the density and pressure functions depend on $t$ only, with

$$\operatorname{div} \mathbf{v} = -\frac{\dot{\varrho}}{\varrho}. \tag{62}$$

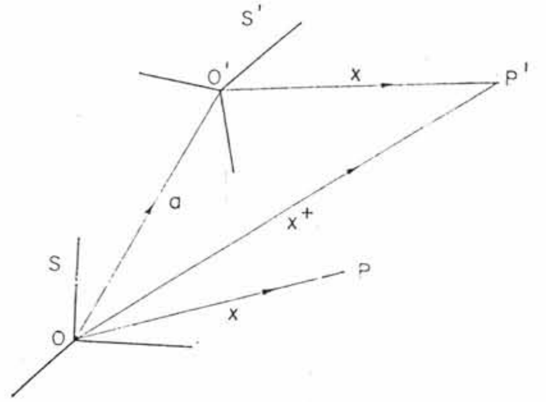



FIGURE 3.4 Diagram showing a generalized description of the cosmological principle in the Newtonian framework (see text).

However, explicit models of the second type show non-linear velocity functions. Although the dynamical behaviour of these models does not introduce any qualitatively new features, the above groups of motion analysis puts the Newtonian cosmological principle on the same footing as the cosmological principle of relativistic cosmology.

### 3.4 Concluding remarks

Although there are conceptual problems of instantaneous action at a distance, velocities exceeding the light barrier, the boundary conditions at infinity, Newtonian cosmology can be made to give a consistent and wide range of cosmological models, in close analogy with relativistic cosmology. This lends confidence to the expansion that in the absence of explicit solutions of Einstein's equation in many cases involving shear and rotation, the Newtonian approach may give a useful insight.

## 4. THE STEADY STATE THEORY

Judging by the extent of controversy it has raised in the past the steady state cosmology is perhaps the best known and the most misunderstood non-standard cosmology. Today, on observational grounds it has receded into the background, although its conceptual elegance cannot be discounted. Nor can its bitterest opponent deny the useful role this theory has played in promoting cosmology *as a science*. In this chapter we shall be



concerned with the theoretical aspects of the steady state theory. The observational assessment will be deferred to Section 10 when the standard as well as the non-standard cosmologies will be discussed together.

### 4.1 The motivation

The steady state cosmology arose in 1948 out of discussions between three British astonomers Herman Bondi, Thomas Gold and Fred Hoyle. Although the end product of these discussions was the same cosmological model. the motivation and approach of Hoyle (1948) differed from that of Bondi and Gold (1948). Before describing these alternative points of view it is essential first to review briefly the reasons for the need to discuss new cosmologies as they were felt three decades ago.

In 1948, the value of Hubble constant was estimated to be $\sim 530\,\mathrm{km\,s^{-1}}$ $\mathrm{Mpc^{-1}}$ so that $H_0^{-1}$ was $\sim 1.8 \times 10^9$ years. Consequently the age of the universe estimated by standard cosmologies was no higher than this period; for a closed universe the age was less than 2/3 of this value. By contrast the age of the Earth as estimated by geological and geophysical arguments was certainly higher than this (Jeffereys, 1948). The crude estimates (Hoyle, 1947) of the stellar ages on the basis of the theory of stellar evolution (which was then in a very rudimentary stage) was also higher than the age of the universe. Thus it was plausible to view the standard models with some degree of skepticism.

The standard models do not provide any clues as to why and how the universe came into existence. The singular epoch $t = 0$ when the big bang is supposed to have taken place seemed therefore somewhat arbitrary. The space-time singularity at $t = 0$ seemed to point to a defect of the theory of standard models. Could one not look instead for models *without* such singular epochs?

Today the "age" problem is not as severe as it was in 1948, thanks largely to a *downward* revision of $H_0$, while problem of space-time singularity still remains pertinent. If at all, it has become more sharply focussed with the realization that space-time singularity is most probably an inherent feature of cosmological models based on general relativity (cf. Section 2.12). We will now review how the two approaches to the steady state cosmology dealt with these problems.

### 4.2 The perfect cosmological principle

Bondi and Gold began by pointing out an important gap in the reasoning behind the standard cosmologies. The usual way of testing a cosmological theory is by a comparison of the distant parts of the universe with the immediate neighbourhood of the observer. Since observations are via the electromagnetic waves these distant parts lie in the remote past; and thus the cosmologist compares the state of the universe here and now with what it was far away a long time ago. Now the basis of the comparison is that the laws of physics are the same in the two regions of space-time. For example the Hubble redshift is deduced on the assumption that the remote galaxy in the past gave rise to similar spectra as found in our neighbourhood galaxies; that the laws of atomic physics which operated then have not changed over billions of years of light travel time.

What is the basis of such an assumption? The ordinary *cosmological principle* (cf. Section 2) ensures that so far as variation in distance is concerned, on any surface of cosmic simultaneity the laws of physics are the same. However, the same cannot be claimed for invariance with respect to time. By definition the universe is the totality of everything, including the laws of science. Can we convince ourselves that the laws of physics have not changed even though the universe has? Clearly there is no mathematical proof for this conviction. All we have is the Occam's razor: make a theory with as few assumptions as possible. A theory in which the laws of physics change with epoch presents many more possibilities than a theory where there is no change. Hence the cosmologist makes this assumption of no change.

Bondi and Gold questioned the propriety of making such a sweeping assumption when the resulting model of the universe was so different in the past as to include a singular epoch. Instead, they argued that the only legitimate way of making this assumption was to presuppose that the universe in the large is itself unchanging in time. Thus the cosmological principle is enlarged to include homogeneity with respect to time. Bondi and Gold called this enlarged cosmological principle the *Perfect Cosmological Principle* (PCP in brief).

Mathematically the PCP implies the existence of a timelike Killing vector (Eisenhart, 1949) over and above those required for defining spatial homogeneity and isotropy. If we consider the Robertson–Walker line-element and impose the above condition on it we find that the function $S(t)$ is determined immediately:

$$S(t) = \begin{cases} \exp \pm Ht & k = 0 \\ H^{-1} \cosh \mathrm{Ht} & k = +1 \\ H^{-1} \sinh \mathrm{Ht} & k = -1 \end{cases} \tag{1}$$

where $H$ = constant > 0.

Of these Bondi and Gold rejected the cases $k = \pm 1$ on the grounds that the surfaces $t$ = constant have time dependent curvature which could in principle be measured. For example, the $k = -1$ case at $t = 0$ gives infinite curvature. The case $k = 0$, on the other hand, does not have this possibility. The world line of the fundamental observer has the timelike Killing vector for a tangent so that such an observer *sees* the universe unchanging. Bondi and Gold gave the name "steady state" to such a universe.



Thus the line element of the steady state universe is given by

$$ds^2 = dt^2 - \exp(2Ht)\{dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)\}. \tag{2}$$

Note that we have committed ourselves to the positive exponential in the scale factor. This deduction can be made on the grounds of local thermodynamic disequilibrium. If $S$ = constant or $v/\exp(Ht)$, the line element would describe a static or a contracting universe with a consequential infinite sky brightness (for details, see Bondi and Gold, 1948). The observed darkness of the night sky and the fact that the process in our neighbourhood appear to be operating as if the universe is far from the state of thermodynamic equilibrium led Bondi and Gold to the conclusion that only the line element (2) is viable. Thus the steady state universe must *expand*.

This deductive approach typifies the Bondi–Gold approach to cosmology. Once we accept the PCP, all we have to do is survey our neighbourhood. The state of the universe anywhere at any time must be the same. Of course the underlying symmetry of PCP is not to be taken to apply on a very small scale. On distance scales up to ~30 mpc and time scales up to ~$10^9$ yrs there is evidence for inhomogeneity (in space) and evolution (in time). The word steady state is supposed to apply over distance and time scales large compared to these.

### 4.3 The continuous creation of matter

Suppose we take a region of the steady state universe bounded by co-ordinates $(r, r + dr)$, $(\theta, \theta + d\theta)$ and $(\phi, \phi + d\phi)$. The proper volume of the 3-dimensional space in this region at a cosmic epoch $t$ is given by

$$dV = r^2 \sin\theta d\theta\, d\phi\, dr \cdot \exp(3Ht). \tag{3}$$

As $t$ increases $dV$ increases, its rate of increase being given by

$$\frac{d}{dt}(dV) = 3H(dV). \tag{4}$$

Now if there were no creation of matter in the universe at any time, the existing density of matter will rapidly fall to zero. This would be contrary to the PCP. If $\varrho$ is the density at time $t$, the same density must be found at *any* other time $t$. In the above instance, during a short time interval $\delta t$, $dV$ will have increased by $3H\,dV\,\delta t$, and in order to maintain the density at $\varrho$, an additional amount of matter must somehow be created to fill up this extra space. The amount needed is $3H\,dV\,\delta t\varrho$ so that the rate of creation per unit time per unit volume is

$$Q = 3H\varrho. \tag{5}$$

Estimating the observed density (in the form of galaxies etc.) to be $\sim 10^{-30}\,\mathrm{g\,cm^{-3}}$ and taking $H \sim 2.5 \times 10^{-18}\,\mathrm{s^{-1}}$ (corresponding to $75\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$) we get $Q \sim 7.5 \times 10^{-48}\,\mathrm{g\,cm^{-3}\,s^{-1}}$.

This is an extremely small quantity—too small to be detected in the laboratory. Nevertheless it represents an apparent violation of the law of conservation of matter. For this reason the steady state theory has been often castigated. The critics apparently *ignore* that in the standard model a much more dramatic violation of the conservation law took place when the whole universe was created in a point singularity. The steady state theory at least brings creation of matter within the purview of the physicist by making it, in principle, an observable and testable phenomenon.

Unfortunately, Bondi and Gold did not take this any further. Granted that continuous creation of matter does take place at the rate (5), what is the mechanism of it? What is created primarily? In what way would it affect or control the expansion of the universe? Is the process stable?

Questions like these cannot be answered by PCP along, however powerful a deductive principle it may be. The conventional physicist's approach to such problems is via a field theory. This was the approach of Hoyle.

## 4.4 The creation field

Hoyle's argument was to point out first that the standard relativistic models fail to explain the one most important problem of cosmology: the creation of matter. By adopting the standard Einstein equations

$$R_{ik} - \tfrac{1}{2}g_{ik}R = -\kappa \underset{(m)}{T_{ik}}, \tag{6}$$

one is forced to deduce the conservation of the matter tensor:

$$\underset{(m)}{T^{ik}}{}_{;k} = 0. \tag{7}$$

Thus, except at a singularity (where (6) breaks down) there is *no* possibility of creating matter. The alternative is to argue that the universe exists for ever with no matter being created (or destroyed). However, relativity does not provide such models. As the well-known singularity theorems have shown, under reasonable physical assumptions all relativistic cosmological solutions must have singularities (Hawking and Ellis, 1973). Thus there is a logical anomaly in the conventional approach: by following Eqs. (6) and (7) we deduce singularity and at a singularity we are forced to introduce matter creation which violates (6) and (7)! This has been discussed by Narlikar (1973).



Hoyle's resolution of the matter creation difficulty was through the addition of another tensor to the right hand side of (6):

$$R_{ik} - \tfrac{1}{2}g_{ik}R = -\kappa[\underset{(m)}{T_{ik}} + \underset{(c)}{T_{ik}}] \tag{8}$$

where $T_{ik}/(c)$ is a tensor representing the so-called creation terms. The conservation law now gives

$$\underset{(m)}{T^{ik}}{}_{;k} \equiv -\underset{(c)}{T^{ik}}{}_{;k}. \tag{9}$$

and the trick is to ensure that both divergencies are non-zero. Hoyle tried several different forms for $T_{ik}/(c)$ but finally settled on one suggested by M. H. L. Pryce.

In the Pryce-formulation† the starting point is the action principle. To the action leading to (4) (see Section 2) are added the extra terms representing creation. These terms arise from a scalar field $C$ in interaction with matter. The action as a whole is given by (with the velocity of light $c = 1$),

$$A = \frac{1}{16\pi G}\int R\sqrt{(-g)}d^4x - \sum_a \int m_a ds_a$$
$$- \tfrac{1}{2}f\int C_i C^i \sqrt{(-g)}d^4x + \sum_a \int C_i da^i. \tag{10}$$

Here $C_i = \partial C/\partial x^i$ and $f$ is a coupling constant.

The variation of $C$ gives the field equations of the $C$-field:

$$\Box C = \frac{n}{f} \tag{11}$$

where $n$ is the net number of *creation events* per unit proper $-4$ volume. What do we mean by a creation event? Assuming as in Figure 4.1 that the world line of a typical particle a begins at the world point $A_+$ and ends at the world point $A_-$, the last integral of (10) gives

$$\int C_i da^i = C(A_-) - C(A_+). \tag{12}$$

In the $C$-variation therefore we get opposite contributions from the end-points $A_\pm$. Counting $+1$ for beginning $(A_+)$ and $-1$ for ending $(A_-)$, we include in n all such end-point contributions. Notice that only the endpoints contribute to the source of $C$ — the actual world lines of particles do not seem to matter!

†This has been discussed in several papers by Hoyle and Narlikar. See, for example, Hoyle and Narlikar (1963).

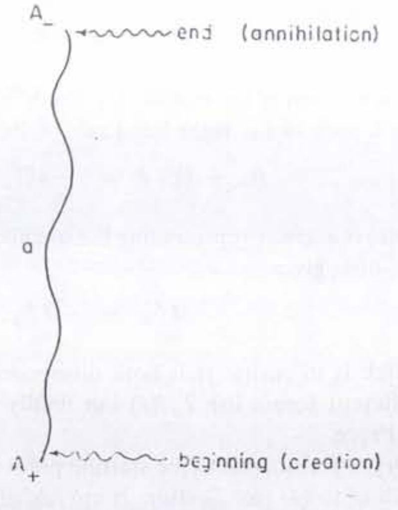



FIGURE 4.1 A typical world line may begin at $A_+$ and end at $A_-$. These points respectively act as positive and negative sources of the C-field.

The variation of world lines confirms this further. The last term of (10) makes no contribution to the action. The world lines therefore are geodesics. However, at each end the following must hold:

$$m_a \frac{da^i}{da}\bigg|_{A_\pm} = C^i\bigg|_{A_\pm} \tag{13}$$

This confirms also that even at a creation (or annihilation) event total energy is conserved. The energy-momentum of the created particle is balanced by that of the C-field.

The variation of geometry ($g_{ik} \to g_{ik} + \delta g_{ik}$) gives the field equations

$$R_{ik} - \tfrac{1}{2} g_{ik} R = -8\pi G\,[\underset{(m)}{T_{ik}} - f(C_i C_k - \tfrac{1}{2} g_{ik} C^l C_l)]. \tag{14}$$

Thus we have

$$\underset{(C)}{T_{ik}} = -f(C_i C_k - \tfrac{1}{2} g_{ik} C^l C_l). \tag{15}$$

Since $f > 0$ the C-field has *negative* energy.

That a negative energy reservoir is needed to sustain a steady state through matter creation is clear from the fact that we do not want the reservoir to be depleted by matter creation. A positive energy reservoir will be depleted by (i) the creation of matter and by (ii) the expansion of the universe (which tends to dilute the reservoir). A negative energy reservoir is augmented by (i) and diluted by (ii). By a proper adjustment a steady state is reached.



For the Robertson–Walker line element the field equations become

$$2\frac{\ddot{S}}{S} + \frac{\dot{S}^2 + k}{S^2} = 4\pi G f \dot{C}^2, \tag{16}$$

$$3\frac{\dot{S}^2 + k}{S^2} = 8\pi G(\varrho - \tfrac{1}{2} f \dot{C}^2). \tag{17}$$

It is interesting to note that (1) is a solution of Eqs. (16) and (17) for

$$\dot{C} = 1, \quad \varrho = m^2 f = \frac{3H^2}{4\pi G} = \varrho_0 \quad \text{(say)}. \tag{18}$$

(Here $m$ = mass of a typical particle). For reasons stated earlier, however, we shall choose the solution

$$S = \exp(Ht), \quad k = 0. \tag{19}$$

It is worth pointing out another aspect of the problem. It is sometimes argued—cf. Hawking and Ellis (1973) p. 90—that a negative energy field would pose insurmountable problems of quantization. It is certainly true that in flat space-time, quantum transitions in a negative energy field would lead to a cascading process. However, when we couple such fields to gravity in the Einstein framework there is an interesting dynamical feed back. In curved space-time any cascading has a counteracting effect. As the energy increases in magnitude it exerts a larger and larger repulsive force and this leads to an expansion of space-time which in turn dilutes the cascading effect. Thus the gravitational interaction of the C-field exerts a controlling influence on the negative energy instability. The quantum behaviour of negative energy fields is still to be investigated fully, however.

## 4.5 The C-field cosmology

The main attractive feature of the Bondi–Gold approach was that it produced the solution (19) uniquely from the PCP. In the Hoyle approach this uniqueness is lost, since the field equations (14) admit many solutions. However, this is compensated by the fact that many investigations of the type mentioned at the end of Section 4.3 can be undertaken in this approach. One result is immediately seen from (18) where there is a unique relation between $\varrho$ and $H$. We summarize the results of some of these investigations, with appropriate references.

4.5.1 *Two classes of solutions* One consequence of the C-field equations is seen from (9). Taking $T_m{}^{ik} = \varrho u^i u^k$ for dust, we have

$$(\varrho u^k)_{;k} u^i + u^k u^i{}_{;k} = f C^i C^k{}_{;k}. \tag{20}$$

Suppose in a given region, there is *no* creation of matter. Then (11) leads to

$$C^k_{;k} = 0, \quad (\varrho u^k)_{;k} = 0 \quad u^k u^l_{;k} = 0. \tag{21}$$

We will call such solutions, solutions of Class I.

On the other hand, from (13) we see that if there is a net creation in any given region, we have $u^i \propto C^i$ and

$$u^k(\varrho u^i)_{;i} = f C^k C^i_{;i}, \quad u^k u^l_{;k} = 0. \tag{22}$$

From the first of (22) it follows that the velocity vector is hypersurface orthogonal, the hypersurfaces being

$$\Sigma^{(C)} \equiv \{C = \text{constant}\}. \tag{23}$$

Thus $\varrho$ = constant along $\Sigma^{(C)}$, and there is *no* spin in the sense described in Chapter 2. We will call such solutions, solutions of Class II.

4.5.2 *Stability and Mach's principle* It can be shown (Hoyle and Narliker, 1963) that if we consider the general first order perturbations of the SS universe, these die out exponentially in fractions of the Hubble time scale. The perturbations may be considered in the following form:

$$\begin{aligned} &\text{density:} && \varrho = \varrho_0 + \varrho_1 \\ &\text{flow vector:} && u^i = (1, 0, 0, 0) + u_1{}^i \\ &\text{line element:} && dS^2 = (1 + h_{44})dt^2 - \exp(2Ht)(\delta_{\mu\nu} + h_{\mu\nu})dx^\mu dx^\nu \end{aligned} \tag{24}$$

where $\varrho_1$, $u_1^i$, $h_{44}$ and $h$ are functions of space-time but are small enough to be ignored in second and higher orders. Over any specified proper volume these quantities die out in various exponential modes of the type $\exp(-nHt)$, $n \geqslant 1$.

This analysis has relevance to Mach's principle. We have already noted that if creation of matter is permitted, i.e., if we have Class II solutions, the spin tensor is eliminated. Thus Godel type solutions, which are anti-Machian, will not be possible. The stability analysis shows that any first order irregularity, i.e., a departure from homogeneity and isotropy is wiped out too. This led Hoyle and Narlikar (1963) to make the following conjecture.

Suppose we have the universe in an arbitrary state at $t = 0$, when creation process is turned on. The *C*-field begins to produce new matter normal to the surfaces $C$ = constant. The negative energy of the *C*-field also produces repulsion and hence expansion. Gradually the expansion dilutes the initial irregularities while the *C*-field produces a "new deal" of regularly distributed matter. There is a clear analogy here with a battery driven current replacing the initial transients in an electric circuit. The outcome is that as $t \to \infty$, the universe tends to the steady state form which represents the eventual outcome of all Class II solutions.



This conjecture, stated in the general form still needs to be proved rigorously. If proved it will explain why we see the universe in the highly regular form—in a form where the background of remote galaxies determines the local inertial frame. This result, first noted in the last century by the physicist-philosopher Ernst Mach, forms the basis of what is known as *Mach's principle*. Mach interpreted it to imply a causal connection between the inertia of a typical object and the large scale structure of the universe.

In Section 2 we saw that the spinning models of the type first discussed by Godel (1949) violate this observational basis of Mach's principle. Einstein himself believed that general relativity would turn out to incorporate Mach's ideas which had influenced him favourably in earlier days. Indications like the existence of spinning models showed, however, that this belief was not going to be justified. In the later part of his life Einstein began to think Mach's principle untenable, or at least inconsistent with the general development of physics. In any case it became obvious that general relativity as such could not be reconciled with or incorporate Machian ideas. The C-field described here seems to give a plausible link between the ideas of Mach and Einstein. In Sections 5 and 6 we will consider other, more radical approaches to Mach's principle.

4.5.3 *The avoidance of singularity* The Penrose-Hawking theorems and other singularity theorems (Hawking and Ellis, 1973) require some kind of positivity of energy. The *C*-field energy tensor violates these energy conditions with the result that we expect singularites of space-time to be absent in many solutions of (14).

The steady state solution is of course one example, and it is an example of Class II solutions. There exists cases of Class I type also. Hoyle and Narlikar (1964a) have discussed the classical Oppenheimer-Snyder problem of the collapsing dust ball in the presence of a *C*-field (Oppenheimer and Snyder, 1939). It is well-known that the interior of the collapsing homogeneous dust ball can be described by the Robertson-Walker line element. The scale factor $S$, in the absence of the *C*-field, satisfies the equation of the form

$$\dot{S}^2 = \frac{A}{S} + B, \; A, B \text{ constants}, \; A > 0. \tag{25}$$

A Class I solution in the presence of the *C*-field modifies it to

$$\dot{S}^2 = \frac{A}{S} + B - \frac{\mu}{S^4}, \; \mu > 0. \tag{26}$$

Notice that whereas in (25) $\dot{S} \to \infty$, $S \to 0$, the corresponding situation does not exist for (26). There a bounce occurs for a minimum $S$ when $\dot{S} = 0$. For $B < 1$, we have a finitely oscillating solution. This is possible only because the $C$-field has negative energy.

Narlikar (1974) has shown how the Class I and II solutions can be combined to produce non-singular Friedmann-like solutions. By solving (11) for a delta-function of time it is possible to mathematically describe the big-bang creation. The subsequent behaviour of the universe is through an equation of the type (26). This universe has a non-singular "beginning" and tends asymptotically to the standard Friedmann models.

## 4.6 The different variants of the steady state universe

Although the strict steady state implies conformity with the PCP at the smallest scale of space and time the reality of the situation forces the cosmologist to apply it with varying degrees of relaxation. We have already mentioned this towards the end of Section 4.2. Here we will review briefly the different variations of the steady state theme both from the fundamental conceptual point of view and from the astrophysical angle.

4.6.1 *McCrea's interpretation* If we substitute the steady state line element into the left hand side of Einstein's equations

$$R_k^{\,i} - \tfrac{1}{2}\delta_k^{\,i} R = -8\pi G T_k^{\,i} \tag{27}$$

we get for $T_k^{\,i}$ the expression

$$T_k^{\,i} = \begin{pmatrix} -p & & & \\ & -p & & \\ & & -p & \\ & & & \varrho \end{pmatrix} \tag{28}$$

where

$$\varrho = \frac{3H^2}{8\pi G} \quad p = -\frac{3H^2}{8\pi G}. \tag{29}$$

McCrea (1951) interpreted the steady state model as a universe with positive densith and *negative* stresses (which lead to expansion).

McCrea argued that the negative stress in (29) being uniform all over the universe, does not produce any observable force (of $-\nabla p$ type) in the universe. However, it does exert gravitational repulsion through Einstein's equations and leads to expansion of the universe; the rate at which this



pressure does work is (per unit volume)

$$|p| \cdot 3\frac{\dot{S}}{S} = 3H\varrho. \tag{30}$$

The right hand side is simply the rate of creation of matter. Thus continuous creation is the manifestation of the work done by negative stresses. McCrea suggested that this stress could be related to vacuum stress of quantum field theory.

4.6.2 *The electric universe* Lyttleton and Bondi (1959) attempted to explain the cosmological expansion by postulating a small difference between the magnitudes of the electric charges of the proton and the electron. Denoting the electron charge by $-|e|$ and the proton charge by $|e| + |\delta e|$ where

$$|\delta e/e| \sim 10^{-19} \tag{31}$$

we see that a typical hydrogen stom will be positively charged with charge $|\delta e|$. Bondi and Lyttleton postulated a steady state universe with a continuous creation of hydrogen atoms. These atoms will repel each other and cause a universal expansion. The Maxwell equations and the electromagnetic energy tensor have to be modified to take account of this charge excess and the non-conservation of charge. It is possible to obtain the steady state model as a solution of the resulting equations.

Laboratory experiments (Hughes, 1964) have, however, ruled out a charge excess of even this small order. The alternative possibility suggested by Lyttleton and Bondi that a comparable number excess (of say protons over electrons) would produce the same result is of course untestable and hence not so attractive as a scientific hypothesis.

4.6.3 *The bubble universe* Hoyle and Narlikar (1966a) explored a model which involved a radical departure from the steady state concept. In this the old $C$-field formalism was adopted but with the coupling constant $f$ raised by $\sim 10^{20}$. Of course this gives a highly dense universe ($\varrho \sim 10^{-8}\,\mathrm{g\,cm^{-3}}$) with an extremely short time scale ($H^{-1} \sim 1$ year!). However, the authors proposed that we live in a "bubble" in this dense universe, a bubble being a part of the space where the creation has been temporarily switched off. Such a bubble expands as a Friedmann universe would, except that its expansion starts from the density of $\sim 10^{-8}\,\mathrm{g\,cm^{-3}}$ instead of from an infinitely dense state.

In this model the investigations of the behaviour of the $C$-field near massive objects led to the conclusion that creation rate could be pushed up in the strong gravitational fields of such objects (Hoyle and Narlikar, 1966b).

Purely gravitational considerations produced a spectrum of the type $dE/E^3$ for the number of particles created in the energy range $(E, E + dE)$. A match with the cosmic ray spectrum could thus be achieved (Hoyle and Narlikar, 1966a, b).

4.6.4 *The hot universe* Gold and Hoyle (1958) suggested that the continuous creation of matter might be in the form of neutrons. This idea conserves charge in the process of creation although the baryon number is not conserved. A neutron decays into a proton, an electron and an antineutrino

$$n \rightarrow p + e^- + \bar{\nu}. \tag{32}$$

The conservation laws of energy and momentum force the electron to take up most of the kinetic energy so that the electrons have a high kinetic temperature of $\sim 10^9$ °K. Gold and Hoyle proposed that this high temperature could be made to work heat engines in the cosmos which would result in the formation of large condensations of the size of $\gtrsim 50$ mpc. They argued that a temperature gradient can cause condensations which pure gravity is unable to do in the expanding universe.

From these condensations groups of galaxies in numbers $\sim 10^3 - 10^4$ form at a time. The observations of superclustering (Abell, 1974) are consistent with such a hypothesis. The idea is also useful in resolving one curious age difficulty associated with the steady state universe (King, 1961; Hoyle and Narliker, 1962a). If the universe is in a steady state we should expect an age $\tau$ distribution of galaxies according to the law

$$N(\tau)d\tau \propto \exp(-3H\tau)d\tau. \tag{33}$$

The average age of a galaxy in such a distribution should be $\sim 1/3H^{-1}$. This is too low ($\sim 4 \times 10^9$ y) compared to the age of our galaxy or of other galaxies in our neighbourhood. However, if galaxies form in superclusters we must add to this figure the period from the formation of the supercluster to the formation of the first galaxy. Also, the galaxies in our own neighbourhood could lie in a supercluster of age $\sim H^{-1} - 2H^{-1}$.

This picture introduces large scale inhomogeneities in the universe and thereby reduces one attractive feature of the steady state model, viz. its ability to make clear testable predictions. Hoyle and Narlikar (1961, 1962b) showed that even a super-Euclidean slope of the log $N$-log $S$ curve for radio sources can be explained within the framework of this model (see Section 10.3).

Gould and Burbidge (1963) pointed out one serious difficulty of this model. A kinetic temperature of $\sim 10^9$ °K will lead to an X-ray background at a level much higher than what is observed. Thus the hot universe idea seemed doomed by the X-ray observations. However, it turns out that the



Burbidge-Gould calculation depends on a high power of Hubble's constant ($\sim H^3$). A lowering of the value of $H$ from 100 km s$^{-1}$/Mpc used by Gould and Burbidge to half this value as suggested by Sandage and his collaborators would remove the seriousness of the discrepancy between the observed X-ray background and that predicted by the hot universe of Gold and Hoyle. This was pointed out by Field (1972).

## 4.7 Concluding remarks

One of the useful by-products of the steady state cosmology was the subject of stellar nucleosynthesis. As pointed out in Section 2.10, Gamow, and his collaborators had proposed a workable but not entirely successful model of synthesizing elements from neutrons and protons in the early big bang.

The steady state model does not have such a hot era, and hence it cannot generate elements that way. This posed a challenge to the sympathizers of this cosmology to find an alternative scenario for element synthesis. A successful attempt in this direction was finally made by Burbidge, Burbidge, Fowler and Hoyle (1957) (B²FH) who showed that stars in their different stages of evolution are able to manufacture all the elements. They were able to obtain a reasonable fit between theory and the observed abundances.

It is believed by many that the stellar nucleosynthesis cannot explain the observed abundances of He and D, and that for this purpose the big bang is necessary. This categorical conclusion is drawn on the assumption that the observed He/H and D/H abundance ratios (which are really seen in certain parts of the Galaxy only) are universal. We will return to this issue in Section 10. Whatever the status of this investigation in future, it cannot be denied that the idea of stellar nucleosynthesis was prompted by the steady state universe.

Although the steady state cosmology by virtue of its being vulnerable to observational tests has faced many critical attacks, the most effective one to date is the discovery of the cosmic microwave background first discovered in 1965 (Penzias and Wilson, 1965). As discussed in Section 2 this background is believed to have carried a signature of the early high temperature era of the big bang. Lacking such an era how can the steady state model explain this background radiation?

It may well be that future observations turn out to show that the background is relatively local, i.e., it does not extend beyond the Galaxy or the local group. It might also turn out that the background spectrum is not that of a blackbody. If these possibilities do not materialize, then for its survival the steady state theory will have to find an alternative (astrophysical) explanation for this background. The situation in which the theory finds itself today is therefore remarkably similar to that 25 years ago when the problem of formation of elements posed a similar threat to its survival.

Whatever its ultimate fate it is interesting to record that many of the ideas of the steady state cosmology are finding applications in altogether different contexts. For example, the concept of baryon nonconservation was explicitly discussed in the hot universe model (*op. cit*). At that time the particle physicists considered it "unphysical" because the concept of conserved baryon member was deeply rooted in their theories. Now the picture has changed. As discussed in Section 10, the baryon number is no longer regarded as a conserved quantity.

There is also a considerable similarity in the presently popular inflationary universe (cf. Section 10.3) and the bubble universe of Section 4.6.3. This point has been highlighted by Narlikar (1984) who has also emphasized that the decay of perturbations discussed in Section 4.5.2 (of Eq. (24)) is the same as the "cosmic baldness hypothesis" put forward in the context of the inflationary model (cf. Barrow, 1984).

Finally, several authors (cf., e.g., Atkatz *et al.*, 1982; Vilenkin, 1982; Brout *et al.*, 1980) have arrived at *C*-field type energy tensors with negative energies by discussing quantum scale fluctuations of spacetime geometry. We will refer to this point in Section 10.4.

## 5. THE BRANS–DICKE THEORY

In 1961 Carl Brans and Robert H. Dicke proposed a theory of gravitation which was based on Machian ideas. Since its inception this theory has attracted considerable attention partly because it makes clearly testable predictions for the weak field case (applicable to the Solar system) and partly because cosmologies based on it lead to a slow decrease of the gravitational constant.

We will first present two versions of the theory which are needed in the subsequent discussion. We will then consider several classes of homogeneous and isotropic world models and finally a general cosmological solution lacking these symmetries. Its observational status will be reviewed in Section 10 along with other cosmological theories.

### 5.1 Motivation for the Brans–Dicke theory

The Brans–Dicke theory makes an attempt to incorporate Mach's principle into gravitational theory (Brans and Dicke, 1961). Mach's principle can be interpreted to mean that the inertia of every particle must arise from interactions with all other particles in the universe. If this is true, the inertial mass $m$ of a particle should depend upon the distribution of matter about it, and therefore should be a function of space-time position.



It is of course impossible to compare masses at different points of space-time, for this would involve carrying a standard mass from one point to another for the purpose of comparison, and the standard itself would change from point to point. One method of getting over this difficulty is to use the characteristic mass provided by gravitation,

$$\left(\frac{hc}{G}\right)^{1/2} \simeq 2.16 \times 10^{-5}\,\text{g}, \tag{1}$$

where the symbols carry their usual meaning, to obtain the dimensionless number

$$m\left(\frac{G}{hc}\right)^{1/2}, \tag{2}$$

which can be compared at different points. Any change in this number can be taken to be indicative of a change in the particle mass $m$ with position. However, the change in (2) could equally well be attributed to a change in the other constants appearing in the expression. If it is decided that $h$ and $c$ will always be taken to be constant† (for, any change in them would mean sweeping changes in quantum theory and in special relativity) then either $m$ or $G$ or both could vary to give different values to the above dimensionless number at different points of space-time. In other words, Machian interactions could be taken to manifest themselves either through a change in particle masses when compared to the unit mass $(hc/G)^{1/2}$ defined at each point, or through a change in the unit of mass itself, i.e., through a change in $G$.

In the original version of their theory, Brans and Dicke (1961) considered the possibility of having a varying "constant" of gravitation and fixed particle mass. This involves replacing $G$ by the reciprocal of a scalar field $\phi$ and obtaining the field equations for the metric and the scalar field. Later, Dicke (1962a) considered a version of the theory in which $G$ is truly a constant and the particle masses vary and are represented by a scalar mass field. The two versions of the theory are related by a space-time dependent change of units, i.e., a conformal transformation. We will briefly consider the two versions of the theory in the next two sections.

### 5.2 The Brans–Dicke theory with varying $G$

This is a scalar-tensor theory of gravitation in which the varying constant of gravitation $G$ is represented by a scalar field. It has been shown by Sciama

†A set of physical 'constants' may always be defined as constant if they cannot be combined to form one or more dimensionless numbers (Brans and Dicke, 1961).

(1953) that Mach's principle leads to a simple relation of the form

$$\frac{GM}{Rc^2} \sim 1, \tag{3}$$

where $M$ is the finite mass of the visible universe and $R$ is its radius. This equation suggests that in a linear approximation, the reciprocal of the gravitation constant $G$ is determined locally as a linear superposition of contributions from the matter in the universe. If there exists a scalar field $\phi$ such that

$$G \sim \phi^{-1}, \tag{4}$$

a simple wave equation for $\phi$ with a scalar matter density as source would lead to a relation roughly in accord with (3). This led Brans and Dicke to replace the gravitation constant $G$ in the action of general relativity ($c = \hbar = 1$ in all subsequent equations),

$$S = \int\left(\frac{1}{16\pi G} R + L\right)\sqrt{(-g)}d^4x, \tag{5}$$

by $\phi^{-1}$ and to add a term corresponding to the Lagrangian density of the scalar field. The action of the Brans–Dicke theory is in fact given by

$$S = \int\left(\phi R + \omega\frac{\phi_{,k}\phi^{,k}}{\phi} + 16\pi L\right)\sqrt{(-g)}d^4x, \tag{6}$$

where $\omega$ is a dimensionless constant and $L$ is the usual Lagrangian density describing any matter and fields that may be present.

It is noteworthy that $L$ does not contain the field $\phi$ or its derivatives and is identical to the Lagrangian density in (5). This means that the equations of motion of particles and field equations (other than the equations for the metric $g_{ik}$) will be the same as in general relativity. In particular, free massive particles will move along timelike geodesics and massless particles will move along null geodesics, i.e., the equivalence principle is preserved. It also follows that, as in general relativity (Landau and Lifshitz, 1975), the energy-momentum tensor $T^{ik}$ of the matter and fields satisfies the local conservation equations

$$T^{ik}{}_{;k} = 0. \tag{7}$$

The equations for the gravitational field are obtained by varying the $g_{ik}$. The result is

$$R_{ik} - \tfrac{1}{2}g_{ik}R = -\frac{8\pi}{\phi}T_{ik} - \frac{\omega}{\phi^2}(\phi_{,i}\phi_{,k} - \tfrac{1}{2}g_{ik}\phi_{,k}\phi^{,k}) - \frac{1}{\phi}(\phi_{;ik} - g_{ik}\Box\phi). \tag{8}$$



Variation of $\phi$ leads to

$$R - \frac{2\omega}{\phi}\Box\phi + \frac{\omega}{\phi^2}\phi_{,k}\phi^{,k}_{,} = 0. \tag{9}$$

Contracting (8) with $g^{ij}$ and using it in (9) then gives

$$\Box\phi = \frac{8\pi}{(2\omega + 3)}T, \tag{10}$$

where

$$T = g^{ik}T_{ik} \tag{11}$$

is the trace of the energy-momentum tensor. Equations (8) and (10) may be taken to be the basic equations of the Brans–Dicke theory. It is clear from these equations and the remarks made above that the role of the scalar field $\phi$ is restricted to its effects on the gravitational field equations (8). Once $g_{ik}$ is obtained, it completely determines the geometric properties of space-time and the effects of gravitation on physical systems as in general relativity.

### 5.3 The Brans–Dicke theory with varying particle masses

In this version of the theory, units are so chosen that $G$ is a constant. Machian interactions will then show up through varying particle masses and these are taken to depend on a scalar "mass field". This form of the theory is obtained from the version presented in the previous section by making a conformal transformation

$$g_{ik} \to g^*_{ik} = \lambda(x)g_{ik}, \tag{12}$$

where the conformal function is to be appropriately chosen (for the meaning of such a transformation see Section 6).

Under the conformal transformation (12) the scalar field $\phi$ which has dimension† (Length)$^{-2}$ transforms to

$$\phi^* = \lambda^{-1}\phi. \tag{13}$$

Choosing $\lambda \sim \phi^{-1}$ will make $\phi^*$ a constant in the new system of units, i.e., $G$ will be a constant. If the Lagrangian density in Eq. (6) is of the form (say) corresponding to interacting charged particles, $a, b, c, \ldots$

$$L = \sum_a \frac{1}{\sqrt{(-g)}}\int(m_a + eA_iu^i)\delta^4(x - x_a(\tau))d\tau + \frac{1}{16\pi}F_{ik}F^{ik}, \tag{14}$$

†In a system of units with $c = \hbar = 1$.

then in the new conformal frame

$$L^* = \lambda^{-2}L. \tag{15}$$

Using Eqs. (12), (13), (15) in (6) and putting

$$\phi^* = \text{constant} \tag{16}$$

leads to the action $S$ written in terms of the starred variables:

$$S = \int \left( \phi^* R^* + \frac{1}{2}(2\omega + 3)\phi^* g^{*ik} \frac{\lambda_{,i}\lambda_{,k}}{\lambda^2} + 16\pi L \right) \sqrt{(-g)^*}\, d^4x; \tag{17}$$

this can be written as

$$S = \int (R^* + 16\pi G_0 (L^* + L_\lambda^*)) \sqrt{(-g)^*}\, d^4x, \tag{18}$$

with

$$L_\lambda^* = \frac{(2\omega + 3) g^{*ik} \lambda_{,i}\lambda_{,k}}{32\pi G_0 \lambda^2}, \quad G_0 = \text{constant} = \phi^{*-1}. \tag{19}$$

In the form (18), the action is the same as that of general relativity except for the term $L_\lambda$ which is the Lagrangian density of the $\lambda$ field. The equations for the gravitational field which follow from (18) are

$$R^*_{ik} - \tfrac{1}{2} g^*_{ik} R^* = -8\pi G_0 T^*_{ik}, \tag{20}$$

which are the same as Einstein's equations. The energy-momentum tensor $T_{ik}^*$ includes, however, a contribution from the $\lambda$-field,

$$T_{(\lambda)ik}^* = \frac{(2\omega + 3)}{16\pi G_0 \lambda^2} (\lambda_{,i}\lambda_{,k} - \tfrac{1}{2} g^*_{ik} g^{*lm} \lambda_{,l}\lambda_{,m}), \tag{21}$$

and obeys the usual conservation equation

$$T^{*ik}_{\ \ ;k} = 0, \tag{22}$$

where the covariant derivative is evaluated using the starred metric. The field equation for $\lambda$ is

$$\Box^* (\log \lambda) = \frac{8\pi G_0}{(2\omega + 3)} T^*, \tag{23}$$

where the D'Alembertian $\Box^*$ and the trace $T^*$ of the energy momentum tensor are evaluated using the starred metric.

Particle masses have the dimension (Length)$^{-1}$. It follows that in the new frame the mass of the $i$th particle is given by

$$m_i^* = m_i \lambda^{-1/2}, \tag{24}$$



where $m_i$ is the constant particle mass in the conformal frame with varying $G$. The equations of motion for an uncharged, spinless, massive particle obtained from (18) are

$$m\left(\frac{d^2x^i}{ds^2} + \Gamma_{kl}^{*i} u^k u^l\right) = h^{ik} m_{,k}, \tag{25}$$

where

$$h^{ik} = g^{ik} - u^i u^k \tag{26}$$

is the projection operator into the rest space of the particle, with $u^i$ the particle four-velocity and $s$ the proper time. A massive particle is prevented from ever moving on a geodesic because of a force due to the varying particle mass which is always present. For massless particles, however, the affine parameter can always be chosen such that the particle trajectories are described by the standard equations of null geodesics.

The two pairs of equations (8, 10) and (20, 23) provide alternative descriptions of the scalar-tensor theory of gravitation which are related by a conformal transformation. It must be emphasized however that the theory is not conformally invariant: the conformal transformation does not generate new solutions of the theory, it merely provides an alternative way of looking at the same solutions. The conformal frame most suitable to the problem under consideration is to be used. However, some care has to be exercised in the conformal transformation if there is a zero or infinity in the mass field $\lambda$ or the scalar field $\phi$, for then a singular conformal function would have to be used. A singular conformal transformation can lead to matter or curvature singularities or geodesic incompleteness in the new frame when these are not present in the original one (see Section 7 for examples of such behaviour in a different context).

### 5.4 Solar system experiments to determine the value of $\omega$

Brans and Dicke (1961) have obtained an exact solution of their theory which corresponds to the static, spherically symmetric space-time around a point mass. Using this, it is possible to calculate as in general relativity the gravitational redshift, bending of light and the relativistic precession of the perihelion of Mercury. Of these only the last can be of use in setting a limit on the size of $\omega$, because of the accuracy available in the observation of the precession.

The value predicted for the perihelion precession in the Brans–Dicke theory is

$$\left(\frac{3\omega + 4}{3\omega + 6}\right) \times \text{(the value predicted by general relativity).} \tag{27}$$

When contributions due to planetary perturbations and other known effects are subtracted from the observed precession of the perihelion of Mercury, there is a residual value of 42.6" $\pm$ 0.9"/century which remains to be accounted for. It is normally believed that the residual precession is entirely relativistic in origin; and it then agrees well with the value 43.3"/century predicted by general relativity. Dicke and Goldenberg however believed that part of the residual precession could be due to the oblateness of the Sun (Dicke and Goldenberg, 1967; Dicke, 1974). They concluded from photoelectric observations of the Sun that its polar diameter $r_p$ must be shorter than its equatorial diameter by $\Delta r/r = 4.51 \pm 0.34 \times 10^{-5}$, and took this as evidence for the existence of a Solar quadrapole moment of $J_2 = 2.47 \pm 0.23 \times 10^{-5}$. This Solar oblateness shold contribute about 7 percent to the observed precession and the relativistic effect should be proportionately less than the value predicted by general relativity. This is achieved by taking $\omega \simeq 6$ in the Brans–Dicke theory. Such a value of $\omega$ would lead to a value for the light deflection which is $\sim 6$ percent less than the value predicted by general relativity. However recent measurements of the rotation of the solar interior by Hills *et al.* (1982) have led to the value $J_2 = (5.5 \pm 1.3) \times 10^{-6}$. A more recent measurement by Duvall *et al.* has led to the even smaller value of $J_2 = (1.7 \pm 0.4) \times 10^{-7}$. This means that the gravitational quadrupole moment of the Sun contributes a negligible amount to any Solar system test, and $\omega$ is restricted to large values.

A good lower limit on the value of $\omega$ has come from the Viking relativity experiment of Reasenberg *et al.* (1979), using radio ranging to Viking landers on Mars and the Viking spacecraft in orbit around it (see Section 11.1). The result of the experiment implies the metric parameter $\gamma = 1.000 \pm 0.002$, which implies that

$$\omega \geqslant 500. \tag{28}$$

Such a large value of $\omega$ makes the scalar-tensor theory much less compelling than it would be for values of $\omega$ close to unity. For large $\omega$, equation (10) can be written as

$$\phi = 0(1/\omega), \tag{29}$$

for which a solution is

$$\phi = \frac{1}{G} + 0(1/\omega),$$

with $G$ constant. Using this in Eq. (8) gives

$$R_{ik} - \tfrac{1}{2}g_{ik}R = -8\pi G T_{ik} + 0(1/\omega). \tag{30}$$

It follows that as $\omega$ becomes large the Brans–Dicke theory goes over into general relativity.



## 5.5 Cosmological models in the Brans–Dicke theory

We have seen in the previous section that recent experiments lead to a substantially larger value for $\omega$ than originally estimated. This means that the scalar field, even if present, can have only a small effect on the scale of the Solar system. But the scalar field can have a considerable effect on the cosmological scale irrespective of the value of $\omega$. There are cosmological models in which the effects due to the scalar field dominate over the effects due to matter and curvature in the early stages of the evolution. The reason for this can be easily understood in the conformal frame of varying partical masses: the $\lambda$-field acts as a cosmological fluid of great rigidity with the equation of state $\varrho = p$, and the energy density due to this field dominates over the energy density of matter in the early stages. The effects of the scalar field can speed up or slow down the expansion considerably over the rates predicted for corresponding models in general relativity, leading to large changes in the amounts of various elements synthesized in the cosmic fireball. The scalar field also changes the nature of the singularity at the origin of the universe. These facts make the study of Brans–Dicke cosmological models worthwhile in spite of the small effects expected locally.

We will consider in this section some of the cosmological solutions of the Brans–Dicke theory available in the literature. As in general relativity, most of the work is in the area of homogeneous and isotropic models, because of their tractability and their possible relevance to the real universe. The possibilities in this case are more numerous than in general relativity because of the increased complexity of the field equations, but analytical solutions are possible only in the simplest cases and numerical integration has to be used in most circumstances. After considering a variety of homogeneous and isotropic solutions, we will consider an inhomogeneous and anisotropic cosmological model. Some observational implications of the Brans–Dicke models will be considered in the next section.

5.5.1 *Homogeneous and isotropic cosmologies* In these models it is imagined that the matter content of the universe is in the form of a homogeneous and isotropic perfect fluid with the energy-momentum tensor

$$T^{ik} = (\varrho + p)u^i u^k - pg^{ik}, \tag{31}$$

where $\varrho$ is the total energy density of the fluid, $p$ its pressure and $u^i$ the fluid four-velocity. The assumed symmetry of the model means that the space-time geometry is described by the Robertson–Walker line-element

$$ds^2 = dt^2 - S^2(t)\left[\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right], \tag{32}$$

in the usual $r, \theta, \phi$ co-ordinates centred at some arbitrary origin. Particles of the fluid have constant $r, \theta, \phi$ co-ordinates so that $u^i = (1, 0, 0, 0)$, and $t$ is the time shown by synchronized clocks carried on the fluid particles. The 3-surfaces of constant curvature are homogeneous and isotropic. $k$ takes the values $+1$, 0, $-1$ when the surfaces are of positive, zero and negative curvature respectively. The function $S(t)$ is to be determined by the field equations.

For the line-element (32) the space–space components of the field equations (8) lead to the single second order equation

$$\frac{2\ddot{S}}{S} + \frac{\dot{S}^2}{S^2} + \frac{k}{S^2} = -\frac{8\pi p}{\phi} - \frac{2\dot{\phi}\dot{S}}{\phi S} - \frac{\omega\dot{\phi}^2}{2\phi^2} - \frac{\ddot{\phi}}{\phi}, \tag{33}$$

and the time–time component leads to

$$\frac{\dot{S}^2}{S^2} + \frac{k}{S^2} = \frac{8\pi\varrho}{3\phi} - \frac{\dot{\phi}\dot{S}}{\phi S} + \frac{\omega\dot{\phi}^2}{6\phi^2}. \tag{34}$$

The field equation (10) for $\phi$ and the conservation equations (7) give

$$\frac{1}{S^3}\frac{d}{dt}(\dot{\phi}\, S^3) = \frac{8\pi}{(2\omega + 3)}(\varrho - 3p) \tag{35}$$

and

$$\dot{\varrho} = -\frac{3\dot{S}}{S}(\varrho + p) \tag{36}$$

respectively. Due to the homogeneity of the space-time, all quantities in the above equations are functions only of the time $t$ and ( . ) indicates differentiation with respect to $t$. Taking the derivative of Eq. (34) and using Eqs. (35) and (36), Eq. (33) can be obtained. The fundamental equations of homogeneous and isotropic Brans–Dicke cosmologies can therefore be taken to be Eqs. (34)–(36) together with an equation of state $p = p(\varrho)$.

Equation (35) can be integrated to give

$$\dot{\phi} S^3 = \frac{8\pi}{(2\omega + 3)}\int_0^t (\varrho - 3p)S^3\, dt + C, \tag{37}$$

where $C = \dot{\phi}(0)S^3(0)$ is the constant of integration. For $C = 0$, a three parameter family of solutions satisfying

$$\dot{\phi} S^3 \to 0 \text{ as } t \to 0 \tag{38}$$

is obtained. A four parameter family of solutions is obtained when $C \neq 0$.



A) *Cosmological models with zero space curvature* ($k = 0$) *and* $C = 0$ The simplest model of this class, a pressure free dust universe, was obtained by Brans and Dicke (1961). Assuming power law solutions of the form

$$S = S_0\left(\frac{t}{t_0}\right)^A, \qquad \phi = \phi_0\left(\frac{t}{t_0}\right)^B, \tag{39}$$

it follows from (36), using $p = 0$, that

$$\varrho = \varrho_0\left(\frac{t_0}{t}\right)^{3A}. \tag{40}$$

Using these equations in (34) and (35) gives

$$A = \frac{2\omega + 2}{3\omega + 4}, \qquad B = \frac{2}{3\omega + 4}, \tag{41}$$

and

$$\frac{\varrho_0 t_0^2}{\phi_0} = \frac{(2\omega + 3)B}{8\pi}. \tag{42}$$

A solution of the field equations is also given by $A = 2$, $B = -4$. But in this case it follows from (42) that either the energy density $\varrho$ or the scalar field $\phi$ (i.e., $G$) would be negative for $\omega > 0$. As $\omega \to \infty$, the solution (41), (42) gives

$$S \sim t^{2/3}, \qquad \phi = \text{constant}, \tag{43}$$

which is the Einstein–de Sitter solution of general relativity. This is to be expected since in the limit of $\omega \to \infty$, the Brans–Dicke theory reduces to general relativity.

Nariai (1968) has obtained power law solutions of the type given in (39) for the equation of state

$$p = n\varrho, \qquad 0 \leqslant n \leqslant 1/3. \tag{44}$$

For these solution Nariai obtains, with the condition that $\varrho$ and $\phi$ should be positive for $\omega > 0$,

$$A = \frac{2\omega(1 - n) + 2}{3\omega(1 - n^2) + 4}, \tag{45}$$

$$B = \frac{2(1 - 3n)}{3\omega(1 - n^2) + 4}, \tag{46}$$

$$\frac{\varrho_0 t_0{}^2}{\phi_0} = \frac{3[\omega^2(1 - n)^2 + 3\omega\{(1 - n)^2 - (1 - 3n)^2/18\} + 2 - 3n]}{2\pi[3\omega(1 - n^2) + 4]^2}. \tag{47}$$

For $n = 0$ this reduces to the dust universe of (41), (42) and for $n = 1/3$ to the radiation universe of general relativity,

$$S \sim t^{1/2}, \qquad \phi = \text{constant.} \tag{48}$$

In the case of the models with non-zero space curvature, $k \neq 0$, the curvature term drops out in comparison with the other terms in equation (34) as $t \to 0$, and the solutions given above are good approximations to the solutions with $k \neq 0$ for vanishingly small $t$.

B) *$\phi$-dominated models* (Greenstein, 1968) These are the homogeneous and isotropic cosmological models with $C \neq 0$ in Eq. (37). If matter is pressure free, it follows from Eq. (36) that $\varrho \propto S^{-3}$; for non-zero pressure, as $t \to 0$† matter obeys the extreme relativistic equation of state $\varrho \simeq 3p$. In either case, for small $t$

$$\frac{8\pi}{(2\omega + 3)} \int_0^t (\varrho - 3p) S^3(t) dt \ll C, \tag{49}$$

and it follows from Eq. (37) that

$$3A + B - 1 = 0, \qquad t_0 = \frac{S_0{}^3 \phi_0 B}{C}. \tag{50}$$

for sufficiently small $t$, for the power law solutions (39).

The strongest dependence of $\varrho$ on $S$ is for radiation, $\varrho \propto S^{-4}$. Using Eqs. (39) and (50) in (34) with this form of $\varrho$, and noting that $A < 1$ for physically interesting solutions ($\varrho$, $G$ positive for reasonable positive values of $\omega$), it follows that as $t \to 0$ the curvature and energy terms can be neglected in (34) and we get

$$A^2 = -AB + \frac{\omega B^2}{6}. \tag{51}$$

It follows from (50) and (51) that

$$A = \frac{\omega + 1 \mp \sqrt{(2\omega/3) + 1}}{3\omega + 4}, \tag{52}$$

and

$$B = \frac{1 \pm 3\sqrt{(2\omega/3) + 1}}{3\omega + 4}, \tag{53}$$

†We always assume that clocks are set such that $t \to 0$ as $S \to 0$.



where the upper sign corresponds to $C$ positive and the lower sign to $C$ negative. For positive values of $C$, $\phi \to 0$ as $S \to 0$ and for negative values of $C$, $\phi \to \infty$ as $S \to 0$. These results are model independent, i.e., for sufficiently small values of $t$ the effects of space curvature and matter are overwhelmed by those of the rapidly changing $\phi$ field, irrespective of the value of $k$ and the equation of state for the matter, and the solution is given by (52) and (53).

As time advances the integral in Eq. (37) dominates over $C$ and the solution goes over into the $C = 0$ case. But if $C$ is very large, its effects could be significant even at the present time. These effects are a reduction in the age of the universe and increase (decrease) in $\dot{\phi}/\phi = \dot{G}/G$ at the present epoch, corresponding to $C$ very large and positive (negative). If $C$ is negative $(\dot{\phi}/\phi)$ can be reduced significantly below the values for models with $C = 0$, and can be made negative, giving rise to an increase in the gravitational "constant" over recent epochs.

C) *The solutions of Gurevich, Finkelstein and Ruban* Gurevich *et al.* have obtained a class of flat space solutions for the equation of state

$$p = n\varrho, \qquad 0 \leqslant n \leqslant 1. \tag{54}$$

We will present here their solutions without going into the details of the calculations. It will be seen that Nariai's solutions (Eqs. (45)–(47)) are obtained as a special case. Gurevich *et al.* have considered the consequences of having a negative $\omega$ (Gurevich, Finkelstein and Ruban, 1972). We will consider such cases also for completeness.

The authors introduce a time co-ordinate $\eta$ which is related to the Robertson–Walker time co-ordinate $t$ of Eq. (32) through

$$d\eta = R^{-3n} dt. \tag{55}$$

The nature of the solutions depends upon the sign of

$$\Delta = \frac{(1 - 3n)^2 (1 + \omega(1 - n))^2 (\eta_1/\eta_2 - 1)^2}{(1 + 2\omega/3)}, \tag{56}$$

where $\eta_1$, $\eta_2$ are constants of integration.

Defining

$$\tilde{\eta} = \frac{\eta}{\eta_2}, \tag{57}$$

the solution for $\Delta > 0$ $(\omega > -3/2)$ is

$$S = S_0 (\tilde{\eta} - \tilde{\eta}_+)^P (\tilde{\eta} - \tilde{\eta}_-)^P, \tag{58}$$

$$\phi = \phi_0 (\tilde{\eta} - \tilde{\eta}_+)^Q (\tilde{\eta} - \tilde{\eta}_-)^Q, \tag{59}$$

where

$$P = \frac{\omega}{3[(1-n)\omega + 1 \mp \sqrt{(1+2\omega/3)}]} \tag{60}$$

$$Q = \frac{1 \mp \sqrt{(1+2\omega/3)}}{[(1-n)\omega + 1 \mp \sqrt{(1+2\omega/3)}]}, \tag{61}$$

$$\tilde{\eta}_{\pm} = \frac{-\tilde{B} \pm \sqrt{(\Delta)}}{2A}, \quad A \neq 0, \tag{62}$$

$$\tilde{B} = 3\sigma(1-n) + (1-3n)\frac{\eta_1}{\eta_2}, \tag{63}$$

$$2A = (1-3n) + 3(1-n)(1+(1-n)\omega). \tag{64}$$

The density of matter is given by

$$\varrho = \frac{M}{S^{3(1+n)}}, \qquad M = \frac{(2\omega+3)\phi_0}{8\pi\eta_2^2 R_0^{3(1-n)} A}. \tag{65}$$

This solution always has a singularity at $\eta = \eta_+$ at which the scale factor $S$ is zero and the density of matter $\varrho$ is infinitely large.

If $\Delta < 0 (\omega < -3/2)$, the solution is with $A \neq 0$

$$S = S_0[(\tilde{\eta} + \tilde{\eta}_-)^2 + \tilde{\eta}_+]^{(1-n)\omega + 1/(2A)} \times \exp\left[\pm \frac{\sqrt{(2/3|\omega| - 1)}}{A} \tan^{-1} \frac{\tilde{\eta} + \tilde{\eta}_-}{\tilde{\eta}_+}\right], \tag{66}$$

$$\phi = \phi_0[(\tilde{\eta} + \tilde{\eta}_-)^2 + \tilde{\eta}^+]^{(1-3n)/2A} \times \exp\left[\mp \frac{3(1-n)\sqrt{(2/3|\omega| - 1)}}{A} \tan^{-1} \frac{\tilde{\eta} + \tilde{\eta}_-}{\tilde{\eta}_+}\right], \tag{67}$$

where

$$\tilde{\eta}_- = \tilde{B}/2A, \qquad \tilde{\eta}_+ = \sqrt{(|\Delta|/2A)}, \tag{68}$$

and all the other constants are defined as in the previous case. The expression for the matter density $\varrho$ is the same as in Eq. (65).

In the $\Delta < 0$ case there is no singularity for any value of the parameter $\tilde{\eta}$, and it can be continued from $-\infty$ to $\infty$. Over this inverval $S$ reduces from $\infty$ to a value $S_{min}$ and again increases to $\infty$. There is no matter or curvature singularity in this model. The reason for this becomes clear in the conformal frame of varying particle masses. It follows from Eq. (21) that the scalar field $\lambda$, in the homogeneous case, has the energy density



$$\varrho_\lambda = \frac{(2\omega+3)\dot{\lambda}^2}{32\pi G_0 \lambda^2}. \tag{69}$$

$\varrho_\lambda$ is negative for $\omega > -3/2$ and acts repulsively and causes a bounce.

In the case $\Delta = 0$ ($\tilde{\eta} = 1$), the solution when expressed in terms of the $t$ co-ordinate reduces to Nariai's solutions.

5.5.2 *An inhomogeneous cosmological solution of the Brans–Dicke theory* All known cosmological solutions of general relativity have a space–time singularity. It was believed at one time that these singularities were a result of the special symmetries, like homogeneity and isotropy, which were inherent to the known models. This conjecture was investigated, and was proved to be wrong, in the singularity theorems of Hawking, Penrose and Geroch (Hawking and Ellis, 1973), on one hand, and in the work of Belinskii, Khalatnikov and Lifshitz on the other (Lifshitz and Khalatnikov, 1963; Belinskii, Khalatnikov and Lifschitz, 1970).

Belinskii *et al.* argued that if the space-time singularity is inherent to general relativity, and not a consequence of any special assumptions underlying the known solutions, then the singularity must occur in a general solution of Einstein's equations. They considered a general solution to be one which contains eight arbitrary functions of the space co-ordinates — one function to describe the density of matter on some initial surface, three functions to describe its four-velocity field, and four functions to describe any free gravitational waves that may be present. Belinskii *et al.* obtained such an inhomogeneous, anisotropic, general solution as a generalization of Kasner's homogeneous, anisotropic solution (Kasner, 1921). This solution showed that (looking backwards in time) a curvature and matter singularity at $t = 0$ was approached in a complicated, oscillatory fashion. As $t \to 0$, the solution passes through an infinite number of "Kasner like" epochs, in each of which there is contraction along two axes and expansion along the third. When the solution passes from one Kasner like epoch to another, the axes along which there is expansion and contraction are interchanged.

Belinskii and Khalatnikov (1973) have obtained an analogous solution of the Brans–Dicke theory and shown that the presence of the scalar field changes completely the nature of the singularity. We will consider their solution in this section. In order to obtain this solution, it is convenient to work in the conformal frame of varying particle masses. It is easy to make a conformal transformation to the more familiar conformal frame of constant particle masses at the end of the calculation.

In order to adhere to the notation of Belinskii *et al.* we introduce a field $\kappa$ defined by

$$\kappa = \left(\frac{2\omega + 3}{2}\right)^{1/2} \log \lambda. \tag{70}$$

The energy-momentum tensor of Eq. (21) then reduces to

$$T^*_{(\lambda)ik} = \frac{1}{8\pi G_0}(\kappa_{,i}\kappa_{,k} - \tfrac{1}{2}g_{ik}\kappa_{,l}\kappa^{,l}). \tag{71}$$

It is assumed that *there is no matter present*. The field equations (20) and (23) can then be written as

$$R_{ik}^* = -\kappa_{,i}\kappa_{,k}, \tag{72}$$

and

$$\Box^*\kappa = 0. \tag{73}$$

A homogeneous Kasner-like solution of these field equations is

$$ds^{*2} = dt^2 - t^{2p_1}dx^2 - t^{2p_2}dy^2 - t^{2p_3}dz^2, \tag{74}$$

$$\kappa = q \log t, \tag{75}$$

where $-\infty < x, y, z < \infty$, $0 < t < \infty$ and $p_1, p_2, p_3, q$ are constants satisfying

$$p_1 + p_2 + p_3 = 1, \quad p_1^2 + p_2^2 + p_3^2 = 1 - q^2. \tag{76}$$

With the exponents arranged in the sequence $p_1 \leqslant p_2 \leqslant p_3$ (this can be done by suitably interchanging the axes), the ranges of the constants are

$$-1/3 \leqslant p_1 \leqslant 1/3, \quad 0 \leqslant p_2 \leqslant 2/3, \quad 1/3 \leqslant p_3 \leqslant 1,$$
$$-\sqrt{(2/3)} \leqslant q \leqslant \sqrt{(2/3)}. \tag{77}$$

The Kasner solution of general relativity is obtained by putting $q = 0$. In this case the ranges of the exponents in (74) are

$$-1/3 \leqslant p_1 \leqslant 0, \quad 0 \leqslant p_2 \leqslant 2/3, \quad 2/3 \leqslant p_3 \leqslant 1. \tag{78}$$

The exponent along the $x$-axis is always negative and there is always contraction along this axis with increasing $t$. In contrast to this, in the case $q \neq 0$ the exponent $p_1$ can take both positive and negative values, and there can be either expansion or contraction along the $x$-axis as $t$ increases. The possibility of positive $p_1$ leads to a singularity of a simple nature in the general solution.

The inhomogeneous solution is obtained only in the limit $t \to 0$. In this case, in synchronous co-ordinates ($g_{00} = 1$, $g_{0\mu} = 0$, $\mu = 1, 2, 3$) the line element and the field $\kappa$ are given by



$$ds^{*2} = dt^2 - (t^{2p_1}l_\alpha l_\beta + t^{2p_2}m_\alpha m_\beta + t^{2p_3}n_\alpha n_\beta)dx^\alpha dx^\beta, \tag{79}$$

and

$$\kappa = q \log t + \kappa_0, \tag{80}$$

where $p_1, p_2, p_3, q$ are subject to the conditions (76) but are now functions of the three spatial co-ordinates. The three vectors **l**, **m**, **n** and $\kappa_0$ are also arbitrary functions of the spatial co-ordinates.

The principle terms in Eq. (73) and in the time–time and space–space components of Eqs. (72) are of order $t^{-2}$; these are identically satisfied by virtue of the relations (76). The space–time components of (72) are also identically satisfied to order $\log t/t$, but reduce in order $t^{-1}$ to three relations between the arbitrary three dimensional functions contained in the solution (74) and (75). Together with Eqs. (76), these provide five relations between the fourteen arbitrary functions **l**, **m**, **n**, $p_1, p_2, p_3, q$ and $\kappa_0$. The synchronous nature of the co-ordinates being used is preserved under arbitrary three dimensional co-ordinate transformations, and these can be used to fix three of the nine functions **l**, **m**, **n**. It then follows that there remain in the solution (79), (80) six arbitrary functions, which is just the number required for a general solution of the Brans–Dicke theory when matter is absent (when matter is present this number is ten).

When the exponents $p_1$, $p_2$, $p_3$ are all positive, as $t \to 0$ the solution approaches monotonically a curvature singularity at $t = 0$ in the same manner as does the homogeneous solution (74), (75). In the case of the general relativistic solution, $p_1$ is always negative. This causes a perturbation term to develop in the equations $R_{ik} = 0$; this term destroys one Kasner like epoch and brings about a transition to another, with the negative exponent transferred to one of the other two axes. In the present case as well, it is possible for the exponent $p_1$ to be negative. However it can be shown (Belinskii, Khalatnikov and Lifshitz, 1970) that if an initial state is specified with $p_1$ negative, a finite number of transitions from one Kasner like epoch to another occur, at the end of which all the exponents take on positive values. The transitions then stop and there is a monotonic evolution towards the singularity. The infinitely many oscillations present in the general relativistic case are therefore absent.

Matter can be written into the empty solution of general relativity without affecting the evolution of the metric for vanishingly small values of the time $t$, for then the curvature terms dominate over the matter terms in the field equations. The same is the case for the general solution of the Brans–Dicke theory. In contrast to this, the effects of the scalar field cannot be ignored in comparison with the curvature terms. The reason for this can best be understood by writing the energy-momentum tensor (17) as

$$8\pi G_0 T^{*ik}_{(\lambda)} = (\kappa_{;l}\kappa_{;}^{\;l}) \frac{\kappa_{;}^{\;i}}{\sqrt{(\kappa_{;}^{\;l}\kappa_{;l})}} \frac{\kappa_{;}^{\;k}}{\sqrt{(\kappa_{;}^{\;l}\kappa_{;l})}} - g^{ik}(\kappa_{;}^{\;l}\kappa_{;l}/2). \tag{81}$$

Comparing with (31), we find that the scalar field acts like a perfect fliud with four-velocity $\kappa_{;}^{\;i}/(\kappa_{;}^{\;l}\kappa_{;l})$ and the equation of state

$$\varrho = p = \tfrac{1}{2}\kappa_{;l}\kappa_{;}^{\;l}. \tag{82}$$

The conclusion that matter plays a minor role near the singularity is valid (Gurevich, Finkelstein and Robson, 1972) only for an equation of state which satisfies $p \leqslant 2\varrho/3$. It is not valid for the fluid with equation of state (82), and analysis shows the $T^{*}_{(\lambda)44}$ is of the same order as $R_{44}{}^{*} - 1/2g_{44}{}^{*}R^{*}$ (i.e., $\sim t^{-2}$).

The solution (79), (80) can be transformed to the conformal frame of constant particle mass. In the notation of Section 3, if $\lambda$ is the conformal function which takes us from the frame of constant masses to the frame of varying particle masses, then from Eqs. (70) and (80),

$$\lambda = \exp\left\{\left(\frac{2}{2\omega + 3}\right)^{1/2}\kappa\right\} = \exp(\kappa_0)t^{q}\sqrt{2/(2\omega + 3)}. \tag{83}$$

The metric and the scalar field $\phi$ in the new conformal frame can be obtained using Eqs. (12) and (13).

### 5.6 The observational status of the Brans–Dicke cosmologies

The predictions of the various cosmological models we have been considering have to be compared with observations to see if any Brans–Dicke model can reproduce the properties of the real universe. Experience with standard cosmological models has shown that the data now available is insufficient to distinguish between various cosmological models using the apparent magnitude–redshift relation, angular diameter–redshift relation etc. The best that an observational cosmologist can do at present is to look for some direct effects due to the presence of the scalar field. The simplest such effect is of course the non-constancy of the gravitational "constant" $G$. Since $G \sim \phi^{-1}$, we have $\dot{G}/G = -\dot{\phi}/\phi$, and if the rate of change of $G$ can be accurately measured, it will be possible to limit the cosmological models to some extent. We have seen that the presence of $\phi$ can effect greatly the evolution of the universe in its early epochs. This can change the fractional abundances of elements synthesized in the primordial fireball. Establishing that the observed abundances of $He^4$ and $H^2$ are cosmological will therefore help to limit the possible Brans–Dicke models. We will return to these possibilities in Section 11.



Whatever the outcome of such observational tests for the Brans–Dicke theory, it cannot be denied that the theory has, for over a decade and a half, provided an interesting alternative to general relativity and to standard cosmology.

## 6. DIRAC COSMOLOGY

We will consider in this section Dirac's Large Numbers Hypothesis and its consequences to cosmology. This idea was first proposed by Dirac in 1937 and taken up again by him in 1973. During this period of 36 years cosmology had advanced considerably as a science and today these ideas are of greater interest than they were in the 1930s. It is now possible to have a clear picture of how the predictions of this cosmology can be checked observationally.

### 6.1 The large numbers hypothesis

Experimental physics provides us with a number of constants with dimensions. The values of these constants depend on the system of units used and therefore cannot be of any fundamental significance. But these constants can be combined together in various ways to give dimensionless numbers which are independent of the system of units used in measurement. These numbers, like the reciprocal of the fine structure constant $(e^2/\hbar c)^{-1} \cong 137$, or the ratio of the proton mass to the electron mass $m_p/m_e \cong 1836$, reflect intrinsic properties of nature. It is generally believed that some future theory will enable these numbers to be calculated from purely numerical factors like $4\pi$, etc., which occur in the equations of physics.

The dimensional constants can also be combined together to give dimensionless numbers which are very large. These numbers involve the gravitation constant $G$ or constants which pertain to properties of the universe as a whole. The most important of these are: (i) The ratio of the electric force between an electron and a proton to the gravitational force between them, $e^2/Gm_pm_e \cong 10^{40}$. (ii) The age of the universe $t_0 \cong 2 \times 10^{10}$ yrs expressed in terms of some natural unit of time, say the time $t_e$ taken by light to travel across the classical electron radius $e^2/mc^2$. The age expressed in this way is $\cong 10^{40}$. This number could vary by a few orders of magnitude depending on which particular natural unit of time is chosen, but in every case it is an enormously large number, about the same as $e^2/Gm_pm_e$. (iii) The number of nucleons $N$ in the universe. This number is clearly finite for a closed universe, but for open universe models it can be infinitely large. In such cases "the number of nucleons in the universe" is taken to be the number in a volume of some characteristic radius like $c/H$, where $H$ is Hubble's constant (at this radius the recession velocity of galaxies is $c$). If $\varrho$ is the proper density of matter and $m_p$ the proton mass,

$$N \simeq \frac{4\pi}{3}\left(\frac{c}{H}\right)^3 \frac{\varrho}{m_p}. \tag{1}$$

$N$ is estimated to be $\sim 10^{78}$ so that $\sqrt{N}$ is remarkably close to the other large numbers, and we can write

$$\sqrt{N} \simeq \frac{e^2}{Gm_p m_e} \simeq \frac{t_0}{t_e} \simeq 10^{40}, \tag{2}$$

where the equalities hold to within a few orders of magnitude.

These extremely large numbers seem to be on a completely different footing from the dimensionless numbers like $(e^2/\hbar c)^{-1}$, which by comparison can be considered to be of the order of magnitude unity. It seems unlikely that any theory can lead to these large numbers starting with factors like $2\pi$; see Eddington (1953) for an attempt.

An unorthodox explanation has been suggested by Dirac (1937, 1938, 1973a, b, 1974, 1975a, b) for the large magnitude of these large numbers and their approximate equality. He points out that of the three numbers in (2), the last is by definition not a constant, for it is the age of the universe (we are assuming that there was a "big bang" and the time elapsed is counted from that as the origin). This means either that the approximate equality of the numbers at the present epoch is a coincidence, or that the first two numbers must also be dependent on time so that the equality is preserved at all times. Dirac accepts the second alternative because it provides a natural explanation for the magnitude of the numbers in (2): these numbers are very large simply because the universe is very old.

Dirac has generalized this idea in what he calls the *Large Numbers Hypothesis* (LNH): any large number which is at the present epoch of the order of magnitude $(t_0/t_e)^k$ must be proportional to $t^k$ with some coefficient close to unity. The coefficients are not necessarily constant for all times $t$, but for large values of the cosmological time $t$ (we will always consider $t$ to be expressed in terms of $t_e$ except when otherwise stated) the time dependence can be neglected and the coefficients taken to be constant.

The Large Numbers Hypothesis has far reaching consequences for physics and cosmology. Some of these were discussed by Dirac (1937, 1938) in the late thirties (see also Bondi, 1961). In recent years Dirac (1973a, b, 1974, 1975a, b) has revived the LNH and has attempted to reconcile it with general relativity by introducing the concept of two space-time metrics, one of which is relevant to mechanical phenomena and the other to atomic processes.



Canuto and others have recently worked out some cosmological and astrophysical consequences of the LNH (Canuto and Londenquai, 1977; Canuto, undated; Steigman, 1978). We will consider in the following sections the two metric theory and some of the applications of the LNH relevant to cosmology.

## 6.2 Variation of the gravitation 'constant' and the law of expansion of the universe

It follows from Eq. (2) and the LNH that $e^2/Gm_p m_e \sim t$. Dirac assumes that the atomic quantities are always constant in terms of the atomic units which are always used in experimental measurements. It follows that in such units

$$G \sim \frac{1}{t} \tag{3}$$

i.e.

$$\frac{\dot{G}}{G} = -\frac{1}{t}. \tag{4}$$

If the value of Hubble's constant at the present epoch is taken to be $H_0 \cong 50\,\text{km sec}^{-1}\,\text{Mpc}^{-1}$, the present age of the universe $t_0$, for $q_0 = 0$ (we will see later that this is the $q_0$ value preferred by Dirac), is $\sim 2 \times 10^{10}$ yrs, which gives

$$\frac{\dot{G}}{G} = -5x10^{-11}\,\text{yr}^{-1}. \tag{5}$$

The variation of the gravitation "constant" is the most unambiguous prediction of Dirac's LNH. If observationally verified, it will not necessarily mean that the LNH is true, but it will be necessary to discard Einstein's theory of gravitation or at least to give a new interpretation of it to accommodate a varying $G$.

The LNH can be used to show that the universe cannot be of the type which expands to a maximumm size and contracts again (as in the case of the $k = 1$ Friedmann model). If the universe had a maximum size, this size expressed in terms of a natural unit of length (the classical radius of the electron $e^2/mc^2$, say) would be a large dimensionless number which would be a constant by definition. The LNH rules out such constant large numbers and therefore the universe cannot have a maximum size.

Dirac (1975b) has also used the LNH to determine the law of expansion of the universe, i.e., the time dependence of the scale factor $S(t)$. If the space sections of the universe are taken to be homogeneous and isotropic, the metric is of the Robertson-Walker type and $S(t)$ is the usual scale factor of the universe. If complications like anisotropy are introduced, $S(t)$ can be considered to be a representative length which determines the volume behaviour of an element of the cosmological fluid of galaxies. In what follows we will always consider cosmological models of the Robertson–Walker type (see Section 2).

The distance at time $t$ between our own galaxy and another is

$$d = xS(t) \tag{6}$$

where $x$ is a constant depending on which particular galaxy is chosen. According to the LNH, the asymptotic expression for $S(t)$ cannot contain any very large constants, and must therefore be of the form $t^n$ or log $t$. The recession velocity of the galaxies must therefore be of the form

$$\dot{d} = nxt^{n-1} \quad \text{or,} \quad xt^{-1}. \tag{7}$$

Now, if $S \sim t^n$ with $n > 1$, there must be some time $t_f$ in the future at which $\dot{d}$ must exceed the speed of light $c$, even for the nearest galaxy. On the other hand, if $1 > n$ or if $S(t) \sim \log t$, there must have been a time $t_p$ in the past at which the recession velocity must have exceeded $c$. If we consider a galaxy for which $v = 10^{-3}c$ now, we have for $S \sim t^n$, $t_p = t_0 \cdot 10^{3/n-1}$, where $t_0$ is the present epoch. For $n = 3/4$ say, $t_p = 10^{28}$ which is a large number. It can similarly be shown that if $n > 1$, $t_f$ also is a large number $\sim 10^{40}$. If $n < 1$ or $n > 1$ we get large numbers $t_p$ or $t_f$ which are constant by definition and therefore not admissible by the LNH. Similarly $S \sim \log t$ also leads to a large number $t_p$ which is a constant. The LNH therefore rules out $S \sim t^n$ with $n > 1$ or $n < 1$ and $S(t) \sim \log t$. The asymptotic law of expansion must therefore be

$$S(t) \sim t. \tag{8}$$

It must be stressed here that the deduction $S \sim t$, or any other law deduced from the LNH, is valid only for large $t$. The large numbers can be considered to be equal to each other for large $t$ only after neglecting factors of order unity which can be time dependent. These factors will not be negligible for small $t$ (when the large numbers are no longer large) and will have to be taken into consideration in determining the time dependences of various quantities.

In deducing the law $S(t) \sim t$, Dirac and later Canuto *et al.* have argued that the moments of time $t_p$ and $t_f$ at which the recession velocity exceeds $c$ must be physically significant moments of time. But it is not clear that any physical significance should be attached to these moments of time for (i) the moments $t_f$ or $t_p$ are not unique to a given model and depend upon which two galaxies are chosen, (ii) the concept of velocity is not very meaningful in relation to two bodies separated by vast distances; in fact it is nowhere necessary to introduce the concept in cosmography, and (iii) even if some physical significance is attached to the concept of the velocity of recession $v$, nothing peculiar happens when $v > c$ in any of the standard models. If the moments $t_p$ and $t_f$ are indeed devoid of any physical significance, their being large dimensionless numbers which are constant will not conflict with the LNH, and it will not be possible to deduce that $n = 1$ in the manner we have done above. We will proceed with this reservation in mind.



### 6.3 The creation of particles

It follows from Eq. (2) that

$$N \sim t^2 \tag{9}$$

i.e., that the number of particles in the universe (in the sense defined above) increases with time. Dirac has taken this to mean that particles are continuously created in the universe. He suggests two possible modes of creation:

(i) *Additive creation*: in this mode the new particles are uniformly created throughout the universe, and therefore mainly in intergalactic space.

(ii) *Multiplicative creation*: in this mode new particles are created where matter already exists, in proportion to the amount already present; presumably the new matter will consist of the same kind of atoms as already existing at the place of creation.

Dirac has not developed a mathematical theory to describe the process of creation. But it is possible to consider qualitatively the observable effects which must be present if either creation mode is operative.

Consider, for example, the application of these ideas to a chunk of matter, e.g., a rock. In the case of multiplicative creation, there would be a fractional increase in the mass $M$ of the rock at the rate $\dot{M}/M = 2/t_0 = 10^{-10}$ per year at the present epoch. This rate of increase of mass is too small to be observed in the laboratory, but it is possible that the matter creation leads to discernable effects in the crystal structure of the rock over geological time scales. In a crystal, the new atoms must either occupy interstitial positions at vacancies or dislocations (see Gittus, 1975), or they must be added to new unit cells external to and part of the existing crystal structure. In the former case, it has been estimated (Towe, 1975) that for a crystal $3 \times 10^9$ years old more than 30 percent of the atoms would have to occupy interstitial positions. This is too many for any crystal to accept without lattice disruption. There should also be a marked difference between X-ray and pycnometrically determined densities, but experiment shows that there is a reasonable agreement between the two. On the other hand, if new atoms are added to new unit cells, it is necessary to create intercrystalline space to accommodate the new cells. If this is the case, the rock and all its crystal should be getting bigger and there should be a secular trend towards increasing grain size in the older rocks. But available evidence supports the opposite conclusion (Nanz, 1953).

In additive creation, new matter is created uniformly throughout space, and its density is expected to be too small to be directly observable. But it has been possible to study theoretically the consequences of the LNH and additive creation taken together for the evolution of a star. In additive creation the amount of matter created in a star over its lifetime is negligible and the stellar mass can be assumed to be constant. The gravitational "constant" $G$ of course varies as $t^{-1}$. Pochoda and Schwarzschild (1964) and Gamow (1967b) have studied numerically the evolution of the Sun under these assumptions. They find that there is no difficulty for the representation of the observed Sun through stellar evolution if the universe is at least 15 billion years old. Since it is believed now that the universe is about 20 billion years old, LNH + additive creation is not inconsistent with observation and stellar evolutionary theories.

We will consider in subsequent sections further consequences of choosing either creation process.

It has been pointed out by Steigman (1978) that it does not necessarily follow from (9) that particles are being continuously created. It was explained in Section 6.1 that by "number of particles in the universe" is meant the number of particles in a characteristic volume $V_H(t) \sim (c/H)^3$. If $n(t)$ is the number of particles per unit proper volume, we have (with $c = 1$)

$$N(t) = n(t)V_H(t) \sim \frac{n(t)}{H^3} \sim n(t)\cdot\frac{S^3(t)}{\dot{S}^3(t)}. \qquad (10)$$

Now, if $\mathcal{N}(t)$ is the number of particles per unit comoving volume we have $\mathcal{N}(t) = n(t)S^3(t)$, and using (10) we get

$$\mathcal{N}(t) \sim N(t)\dot{S}^3(t) \sim t^2\dot{S}^3(t). \qquad (11)$$

It follows from this that for $S(t) \sim t$, $\dot{S}(t) =$ constant and the number of particles in a given co-ordinate volume increases and we have particle creation. If on the other hand $S(t) \sim t^{1/3}$, $\mathcal{N}(t)$ is constant. In this case there is no particle creation and the increase in the number of particles in the universe $N$ is entirely due to the time dependence of $S(t)$. For any other time dependence of $S(t)$ there is particle creation with $\mathcal{N}(t) \sim t^2$. In an early version of his theory Dirac proposed that the total amount of matter in the universe was conserved and was led precisely to the time dependence $S(t) \sim t^{1/3}$ for the scale factor.



## 6.4 The two metrics

The variation of $G$ and the creation of matter predicted by the Large Numbers Hypothesis seem to conflict with Einstein's theory of gravitation which is based on a constant $G$ and the conservation of energy. It is therefore necessary to either modify Einstein's theory, or to reinterpret it suitably so that it is consistent with the LNH. Dirac has chosen the latter alternative, for he believes that general relativity, which is very successful in accounting for observations, should not be abandoned altogether. Dirac has attempted to reconcile general relativity with the LNH by introducing two metrics, the Einstein metric $ds_E$ and the atomic metric $ds_A$.

All measurements in the laboratory are made using atomic standards. In particular, distances and time intervals are measured using wavelengths of spectral lines, lattice spacings in crystals, etc. These intervals lead to a metric for space-time which Dirac calls the *atomic metric* $ds_A$. Dirac proposes that the metric which enters Einstein's equations is not $ds_A$, but another metric $ds_E$ which he calls the *Einstein metric*. In contrast to $ds_A$, the Einstein metric cannot be directly observed, but it reveals its existence by determining the course of mechanical phenomena. For instance, planets are expected to move along the geodesics determined by $ds_E$.

In a given physical situation, $ds_E$ is obtained by solving Einstein's equations. All measurements pertaining to $ds_E$ are said to be in Einstein units. In these units $G$ is constant by definition. The masses of macroscopic bodies whose structure is determined gravitationally (planets, stars etc.) are also defined to be constant so that these bodies will move along geodesics, in accordance with Einstein's theory. In making measurements of course, it is necessary to introduce atomic apparatus. Fundamental atomic parameters like particle charges, masses, Planck's constant, etc., are then defined to be constant, and atomic units are said to be used. Quantities which are constant in the Einstein units need not be constant in atomic units and vice versa. In particular, $G$ acquires a time dependence in atomic units when Eq. (2) is used. The introduction of the two metrics allows the calculation of space–time geometry using Einstein's equations, but all quantities have to be transformed to atomic units before comparison is made with observations. We will obtain below the time dependence of some basic quantities in each system of units.

We will use in what follows the subscript $E$ to indicate quantities expressed in Einstein units. The absence of a subscript will mean that atomic units are being used. The time dependencies obtained in the previous sections are all in atomic units.

In Einstein units, we have by definition

$$G_E = \text{constant}, \quad M_E = \text{constant} \qquad (12)$$

where $M_E$ is the mass of any astronomical body. In atomic units, the charge and mass of elementary particles are taken to be constant

$$e = \text{constant}, \; m = \text{constant}. \tag{13}$$

The velocity of light $c$ is taken to be constant ($= 1$) in both units. The work of Bahcall and Schmidt (1967) shows that the value of $e^2/\hbar c$ ($= 1/137$) does not vary appreciably over a large range of redshifts, i.e., over large volumes of the universe. It follows from (13) that

$$\hbar = \text{constant}, \tag{14}$$

where $\hbar$ is (Planck's constant)/$2\pi$.

The total mass of an object in Einstein units is

$$M_E = m_E N, \tag{15}$$

where $m_E$ is the mass of a nucleon and $N$ is the total number of nucleons in the object. $N$ is a dimensionless number and therefore independent of the system of units. In the case of multiplicative creation, the number of nucleons $N$ in the object increases as $t^2$; in the case of additive creation we have $N$ = constant. Since we must have $M_E$ = constant for any macroscopic body, the nucleon mass $m_E$ must have the time dependence

$$m_E \sim t^{-2} \text{ (multiplicative creation)} \tag{16}$$

and

$$m_E = \text{constant (additive creation)}. \tag{17}$$

From Eqs. (2) and (12)

$$\left[\frac{e^2}{m^2}\right]_E \sim t, \tag{18}$$

and it follows from Eqs. (16), (17) and (18) that

$$e_E \sim t^{-3/2} \text{ (multiplicative creation)} \tag{19}$$

and

$$e_E \sim t^{1/2} \text{ (additive creation)}. \tag{20}$$

Since $e^2/\hbar c$ is a dimensionless constant (note that $c = 1$ but we sometimes introduce it for convenience) we also have

$$\hbar_E \sim t^{-3} \text{ (multiplicative creation)} \tag{21}$$

and

$$\hbar_E \sim t \text{ (additive creation)}. \tag{22}$$



The connection between the two metrics can be determined by considering the motion of the Earth around the Sun in the Newtonian approximation. The basic equation is

$$GM = v^2 r \tag{23}$$

where $M$ is the mass of the Sun, $r$ the radius of the Earth's orbit and $v$ its velocity. This equation holds equally well in either system of units since it is a "balance of forces" equation. The velocity $v$ is a dimensionless number ($v < 1$) and $v = v_E$ = constant. The mass $M_E$ and the "constant" of gravitation $G_E$ are constant in the Einstein units and therefore

$$r_E = \text{constant}. \tag{24}$$

In atomic units $G \sim t^{-1}$ and $M \sim t^2$ for multiplicative creation and $M$ = constant for additive creation. We therefore have

$$r \sim t \text{ (multiplicative creation)} \tag{25}$$

and

$$r \sim t^{-1} \text{ (additive creation)}. \tag{26}$$

If $r$ and $r_E$ are taken to be equal at $t = 1$, we get $r = r_E t$ and $r = r_E t^{-1}$ for the cases of multiplicative and additive creation respectively.

In the case of multiplicative creation the Earth recedes from the Sun, when the distance is measured in atomic units, and the whole Solar system expands. In the case of additive creation the Earth approaches the Sun and the Solar system contracts. The change in the radius of the orbit can produce change in the Earth's surface temperature over geological time scales. For a discussion of this point see Canuto and Londenquai (1977), Roxburgh (1977) and the references given there.

The results (25) and (26) can be straightaway generalized to give the connection between the metrics $ds_A$ and $ds_E$:

$$ds_A = t\, ds_E \text{ (multiplicative creation)} \tag{27}$$

and

$$ds_E = t\, ds_A \text{ (additive creation)}. \tag{28}$$

## 6.5 Cosmological models in Einstein units

In Einstein units the metric $ds_E$ of every cosmological model has to be a solution of Einstein's equations.

We will first consider the case of multiplicative creation. The scale factor $S(t)$ which determines the distances between galaxies has the time dependence

$S(t) \sim t$, in atomic units (this time dependence is independent of the mode of creation). We see then from Eq. (27) that all distances will be constant in Einstein units and we therefore need a static solution of Einstein's equations. The only static cosmological solution (assuming homogeneity and isotropy) is the static Einstein universe (Weinberg, 1972). In this case if the energy density and pressure of matter are not to be negative, it is necessary to introduce the cosmological constant $\lambda_E$. The line element is

$$ds_E^2 = dt_E^2 - \left[\frac{dr_E^2}{1 - r_E^2/R_E^2} + r_E^2 d\Omega^2\right], \qquad (29)$$

with

$$\lambda_E = R_E^{-2}, \varrho_E = \left[\frac{\lambda}{4\pi G}\right]_E, \qquad (30)$$

for a dust filled universe with zero pressure and energy density $\varrho_E$.

Since we are considering the case of multiplicative creation, the line element corresponding to (29) in atomic units is

$$ds_A^2 = t^2 dt_E^2 - t^2\left[\frac{dr_E^2}{1 - r_E^2/R_E^2} + r_E^2 d\Omega^2\right]. \qquad (31)$$

Using

$$\frac{dt}{t} = dt_E \qquad (32)$$

and introducing a new radial co-ordinate $r = r_E/R_E$ (note that the co-ordinates are dimensionless numbers), we get

$$ds_A^2 = dt^2 - S^2(t)\left[\frac{dr^2}{1 - r^2} + r^2 d\Omega^2\right], \qquad (33)$$

with

$$S(t) = S_0 t, S_0 = R_E = \lambda_E^{-1/2}. \qquad (34)$$

A comparison of (33) with the standard Robertson–Walker line element shows that the $t$ = constant 3-surface must have positive curvature ($k = 1$). This is the only value of $k$ consistent with the demands of the LNH and multiplicative creation.

It is seen from (32) that referred to the time $t_E = \log t$, which Dirac calls the dynamic time, the Einstein universe has always existed. It is everywhere free from matter and curvature singularity. In contrast to this, referred to the atomic units there is a matter and curvature singularity at $t = 0$, when all the



matter is together. This difference in behaviour of the two metrics *vis-a-vis* the singularity occurs because of the singular conformal transformation which connects the two metrics (see Section 7).

At first sight it might seem inconsistent that a static solution is obtained when the number of particles $\sim t^2$. But there is no inconsistency if it is remembered that in the Einstein units particle masses decrease as $t^{-2}$. The masses of astronomical bodies in which the creation takes place are therefore constant, the density of the cosmological fluid does not change with time and a static solution is permitted.

In the atomic metric the observed redshifts in the light coming from distant galaxies is attributed to the expansion of the universe. The redshift being a dimensionless quantity will have the same value in the Einstein metric as well, and we have to understand why redshifts occur in the static universe. To see this, it must be remembered that measurements are always made using atomic clocks. If $\Delta t_A$ is a unit of time in atomic units, the corresponding unit in Einstein units is $\Delta t_E = t^{-1}\Delta t_A$, and atomic clocks are continually speeding up with respect to Einstein clocks. For a given atomic time interval $\Delta t_A$, the Einstein time intervals at epochs $t_0$ and $t_e$, measured with atomic clocks, are in the proportion

$$\frac{(\Delta t_E)_0}{(\Delta t_E)_e} = \frac{t_0}{t_e} \qquad (35)$$

Since the unit Einstein intervals are different at different epochs, the wavelength of a spectral line changes with epoch, and for $t_0 > t_e$ we have

$$1 + z = \lambda_0/\lambda_e = t_0/t_e = S(t_0)/S(t_e) \qquad (36)$$

which is the same as the relation we have for the redshift in the atomic metric.

In the case of additive creation, matter is created uniformly throughout space. We cannot now have conservation of mass by letting particle masses vary, for that would lead to a change in the masses of astronomical bodies and that is not permitted in Einstein units. Dirac suggests that to have the conservation of energy in this case, it must be supposed that together with positive mass, a uniform distribution of negative mass is created, so that the total density of the created matter is zero. In order not to have violent disagreement with observation, Dirac suggests that the negative mass is not observable. It must not interact with other atoms, except gravitationally, and must not have any physical effects, apart from producing a curvature of space. The newly created positive mass can be expected to be in the form of hydrogen atoms. These condense to form the stars and galaxies that we observe. The negative mass remains uniform and unobservable (it is interesting to compare the distribution of negative matter with the sea of negative energy electrons which Dirac postulated when interpreting his wave equation for the electron, and with the creation field of Hoyle and Narlikar (see Section 4)).

The total density of matter is zero apart from local irregularities arising from the condensations of the H atoms. If these irregularities are smoothed out a model is obtained in which the space-time, referred to the Einstein metric, is flat. It is just the Minkowski space-time and the Einstein metric becomes the Minkowski metric.

### 6.6 Concluding remarks

We have been considering the Large Numbers Hypothesis and its consequences to Cosmology. It is difficult to test directly some consequences of the LNH like particle creation. On the other hand the LNH makes some predictions, like $G \sim t^{-1}$ and $q_0 = 0$, which are in principle capable of being tested in straightforward ways. In applying the standard cosmological tests like the $m - z$, $\theta - z$ relations to verify if $q_0 = 0$, it is necessary to allow for the time dependences expected from the LNH, in addition to the possible evolutionary effects present in the standard models.

A further test of the LNH can be made by examining the nature of the cosmic microwave background spectrum. It can be shown that there are a number of effects in Dirac's cosmology which tend to distort the spectrum of the microwave background if it is initially a blackbody spectrum. This means that if it is verified that the microwave background is of the black-body type, the LNH would be inconsistent with observations. We will consider in greater detail some of these questions in Section 11.

It must be noted that the LNH is independent of the two-metric theory of Dirac. The two metrics have been introduced as a possible theoretical framework to reconcile a time dependent $G$ with Einstein's general relativity. If it is necessary to discard the two-metric theory either because it clashes with observations or because a more satisfactory theoretical framework is found, the Large Numbers Hypothesis will in no way be affected.

## 7. THE HOYLE–NARLIKAR COSMOLOGIES

We now discuss a theory of gravitation and cosmology which is perhaps more strongly rooted in Mach's principle than any other theory discussed so far. To introduce this theory we return to Mach's principle.

### 7.1 Another interpretation of Mach's principle

In Sections 4 and 5 we came across Mach's principle in two different contexts. We now look at this principle from another angle which leads to a new



gravitation theory. These different interpretations are to some extent due to the inherent vagueness of Mach's own statement of the principle: that the property of inertia is in some way a result of the background provided by the universe in the large. How the connection

$$\text{background} \Rightarrow \text{inertia}$$

is to be established quantitatively was never stated by Mach in any of his discussions (Mach, 1960). Physicists following Mach (Einstein included) have interpreted Mach's principle with varying degrees of confidence and skepticism. Even those who consider the idea as a significant one differ in their mathematical interpretation of it. The interpretation given in this section is perhaps the most direct expression of Machian concepts in quantitative form. The approach to this interpretation is via the following argument.

Consider the second law of Newton in the obvious notation:

$$\mathbf{P} = m\mathbf{f}. \tag{1}$$

The force $\mathbf{P}$ acting on a body of mass $m$ produces an acceleration $\mathbf{f}$. The statement of the law of course presupposes a frame of reference. In Newton's formulation the absolute space was postulated to serve the purpose of the background reference frame. An *inertial* frame, i.e., a frame *unaccelerated* relative to the absolute space will also serve the purpose. An accelerated frame of reference will, however, require the introduction of the so-called inertial forces. For instance, even assuming that the external force $\mathbf{P}$ is unaffected, the law given by (1) as seen through a frame with acceleration $\mathbf{a}$ relative to the absolute space, would look like

$$\mathbf{P} - m\mathbf{a} = m\mathbf{f}' \tag{2}$$

where $\mathbf{f}'$ is the acceleration relative to the non-inertial frame. The inertial force $-m\mathbf{a}$ indicates that a non-inertial frame has been used.

Newton had been puzzled by the unique status of the absolute space (and the inertial frames) and was not entirely satisfied that he had to *postulate* the existence of such a fundamental entity.

Later Mach noted that the absolute space was in fact the background framework of the distant stars (see Section 1). Following his view that the concept of inertia depends on the second law of motion which in turn requires the distant background to provide a reference frame, we can argue that inertia *results* from the distant background. To see this, consider the application of (1) to an isolated particle in an otherwise empty universe. With $\mathbf{P} = 0$ we get

$$m\mathbf{f} = 0. \tag{3}$$

The usual conclusion from this is $f = 0$. This is manifestly nonsense: for without a background it is meaningless to conclude that the particle moves with uniform velocity! The alternative conclusion of (3) is $m = 0$. In this case the particle has zero inertia and a totally indeterminate motion. Hence in Machian spirit we should look for an expression for inertia in terms of the background, which leads to $m = 0$ for the above empty universe scenario.

The hypothesis that the inertia of a particle arises from sources far away in the universe looks like an "action at a distance" concept. Since this had already proved unsuccessful in electromagnetism, Einstein (1949) was doubtful of its validity. However, since then action at a distance has been made a viable concept in electrodynamics (Hoyle and Narlikar, 1974) and the extension of the concept to inertia does not seem so unreasonable now.

## 7.2 Inertia as a direct particle field

Hoyle and Narlikar (1964b, 1966c) defined the inertia of a typical particle in terms of the rest of the particles in the universe and made this definition the starting point of a theory of gravitation. Let the universe consist of a large number of particles labelled a, b, c, . . . with masses $m_a, m_b, m_c, \ldots$ As in general relativity the space-time geometry is assumed to be Riemannian with the metric tensor $g_{ik}$. Let $a^i$ denote the co-ordinates of particle $a$ and $ds_a$ the element of its proper time so that

$$ds_a^{\ 2} = g_{ik}\, da^i da^k. \tag{4}$$

Hoyle and Narlikar then define the mass of particle $a$ at a world point $A$ by the formula

$$m_a(A) = \lambda_a \sum_{b \neq a} m^{(b)}(A) \tag{5}$$

where $m^{(b)}(A)$ is the contribution of inertia from the particle $b$. Thus the mass of $a$ arises from the "rest of the universe" as suggested by Mach. $\lambda_a$ is a coupling constant.

To fix ideas further it is necessary to define the inertia functions. This is done by the approach of *direct particle fields*, in analogy with the Fokker action for electrodynamics. There the direct particle potential at a typical point $A$ due to an electrically charged particle $b$ is given by the formula

$$A_i^{(b)} = \int e_b \bar{G}_{ik}(A, B)\, db^k \tag{6}$$

where $\bar{G}_{ik}(A, B)$ is a symmetric bi-vector Green's function (with index $i$, operating at $A$ and $k$, at $B$) of the wave equation

$$\Box_A \bar{G}_{ik}(A, B) + R^l_{\ i}(A)\bar{G}_{lk}(A, B) = [-\bar{g}]^{-1/2} \bar{g}_{ik}\, \delta_4(A, B). \tag{7}$$



Here $\bar{g}_{ik}(A, B)$ is the parallel propagator (Synge, 1960) between $A$ and $B$ and $\bar{g}(A, B)$ is the determinant of the matrix $\bar{g}_{ik}(A, B)$. The details of this have been discussed elsewhere (Hoyle and Narlikar, 1974). It can be shown that a full theory of electrodynamics, classical and quantum, can be built out of such an action at a distance concept in which the direct particle fields and potential have no degrees of freedom *independent* of their source particles.

Returning to inertia, the definition of $m^{(b)}(A)$ can be given in the formula

$$m^{(b)}(A) = \int \lambda_b \tilde{G}(A, B)\, db, \tag{8}$$

where $\tilde{G}(A, B)$ is the symmetric biscalar Green's function of the wave equation

$$[\Box + \tfrac{1}{6}R]_A\, \tilde{G}(A, B) = [-\bar{g}]^{-1/2}\, \delta_4(A, B). \tag{9}$$

Note that corresponding to a vector interaction in electrodynamics we now have a scalar interaction. The term $R/6$ in the wave operator is necessary to achieve *conformal invariance* as in electrodynamics.

Conformal invariance has a particular significance in a Machian theory of variable masses as well as in a direct particle theory.

## 7.3 Conformal invariance

It is usually supposed that the units measuring any physical quantity are expressible as products of powers of three elementary units – of length (L), mass (M) and time (T). However, two fundamental constants $c$ (from special relativity) and $h$ (from quantum theory) signifying limitations on measurements placed by nature, really reduce the number of fundamental units to one. Taking this unit to be the mass unit, we can write, for example:

$$\begin{gathered}\text{length} \sim M^{-1}, \quad \text{time} \sim M^{-1}, \quad \text{electric field} \sim M^2, \\ \text{gravitational constant} \sim M^{-2}, \quad \text{magnetic moment} \sim M^{-1}, \\ \text{etc.} \quad \text{etc.}\end{gathered} \tag{10}$$

Thus in principle any quantity can be "measured" if we fix a unit of mass. Taking this, for example, to be the mass of the electron, all physical quantities are expressible as numbers. However, do we have the guarantee that this unit is the same all over the universe?

In a theory using the Newtonian concept of inertia this is *assumed* to be so. But in a Machian theory, inertia may change owing to a changing background and hence a comparison of physical quantities becomes a meaningless exercise unless our physical interpretation does not depend on the unit used. Since a change of mass unit results in a change of length unit we may restate the above requirement in these words: "the physical theory of interpretation should be *conformally* invariant".

In geometrical terms if the line element measured in one set of units looks like

$$ds^2 = g_{ik}\,dx^i dx^k, \tag{11}$$

then in another set of units all lengths are scaled and

$$ds \to ds^+ = \Omega ds \tag{12}$$

where $\Omega$ is a scale factor which may be a space-time dependent quantity. It is of course implicit in the conformal transformation (13) that it is non-singular, i.e.,

$$\Omega \neq 0, \quad |\Omega| \neq \infty. \tag{13}$$

We will call such a conformal function as "admissible". The physical theory, i.e., the equations describing it, should not have to take account of the arbitrary function $\Omega$. This is the essence of conformal invariance. Since the mathematical equations describing physics are usually second order differential equations $\Omega$ should be at least $C^{(2)}$.

From (11) and (12) we see that $ds = 0 \Leftrightarrow ds^+ = 0$, i.e., null lines are conformally invariant. Null geodesics are also conformally invariant. Since the main support of a direct particle interaction (i.e., of the Green's function describing it) is along null cones, the conformal invariance has a special significance for direct particle theories. Just as null cones are invariant *locally* under a Lorentz transformation, so they are invariant globally under a conformal transformation (Hoyle and Narlikar, 1972c).

### 7.4 The gravitational equations

The usual inertial term in the action (cf. Section 2) now takes the form:

$$\mathscr{A} = -\sum_a \int m_a ds_a = -\sum_{a<b}\sum \iint \lambda_a \lambda_b \tilde{G}(A, B)\, ds_a ds_b. \tag{14}$$

The remarkable aspect of this theory is that this term *also* contains gravitation! The expression (14) in fact contains the action for a system of gravitating particles. The Machian paradox of the single particle in empty space does not arise. In (6) $m_a = 0$ for a single particle $a$ in an otherwise empty space. In fact if the number of particles is less than two, (14) tells us that the action itself cannot be defined: there is no physics in such a case!

The gravitational equations follow by the usual prescription $\delta\mathscr{A}/\delta g_{ik} = 0$. It is convenient to express the resulting equations in terms of the following space-time functions:

$$m(X) = \sum_a m^{(a)}(X) = \tfrac{1}{2}[m^{(\mathrm{ret})}(X) + m^{(\mathrm{adv})}(X)], \tag{15}$$



$$\phi(X) = m^{(\mathrm{ret})}(X) m^{(\mathrm{adv})}(X), \quad m_{,k} = \frac{\partial m}{\partial x^k} \text{ etc.}, \tag{16}$$

$$N(X) = \sum_a \int \delta_4(X, A)[-\bar{g}(X, A)]^{-1/2} ds_a. \tag{17}$$

Here the suffixes "retarded" and "advanced" denote respectively the retarded and advanced inertial contributions (from all particles) at a typical point $X$. We will take $\lambda_a = 1$ for all $a$ in the following presentation. Later we will investigate variable $\lambda_a$. The gravitational equations are

$$\tfrac{1}{6}\phi[R_{ik} - \tfrac{1}{2}g_{ik}R] = -T_{ik} + \tfrac{1}{6}[g_{ik}\Box\phi - \phi_{;ik}] + \tfrac{1}{2}[m_{,i}^{(\mathrm{ret})} m_{,k}^{(\mathrm{adv})} + m_{,k}^{(\mathrm{ret})} m_{,i}^{(\mathrm{adv})} - g_{ik} m_{,p}^{(\mathrm{ret})} m_{,q}^{(\mathrm{adv})} g^{pq}]. \tag{18}$$

These are supplemented by the "source" equation for $m(X)$:

$$\Box m + \tfrac{1}{6}Rm = N. \tag{19}$$

It will be noticed that the *trace* as well as divergence of (18) vanishes identically. This leaves us one equation short if we wish to determine all the unknowns, i.e., $g_{ik}$ and $m$.

The situation is, however, analogous to that of general relativity. As in relativity, the co-ordinate invariance means that we need not expect a unique answer for $g_{ik}$. A change of co-ordinates produces another equally valid form of $g_{ik}$. In the same way here the conformal invariance implies that if $[g_{ik}, m]$, is a solution, then so is $(\Omega^2 g_{ik}, \Omega^{-1}m)$ if $\Omega$ is an admissable conformal function.

Suppose that the universe as a whole produces a response such that

$$\sum_a m^{(a)(\mathrm{ret})}(X) = \sum_a m^{(a)(\mathrm{adv})}(X) = m(X). \tag{20}$$

This response condition is similar to that of Wheeler–Feynman absorber theory of radiation which is automatically called into play in any action at a distance theory (Wheeler and Feynman, 1945). For details see Hoyle and Narlikar (1974). The gravitational equations now become

$$\tfrac{1}{2}m^2[R_{ik} - \tfrac{1}{2}g_{ik}R] = -T_{ik} + \tfrac{1}{6}[g_{ik}\Box m^2 - m^2_{;ik}] + [m_{,i}m_{,k} - \tfrac{1}{2}g_{ik}m_{,}^{l}m_{,l}]. \tag{21}$$

We will assume that (20) holds, and suppose that it is possible to find an admissible conformal function $\Omega$ such that

$$m^{(\mathrm{ret})}\Omega^{-1} = m_0(\text{constant}). \tag{22}$$

We will refer to this frame as the Einstein conformal frame (ECF). The equations (21) then reduce to

$$R_{ik} - \tfrac{1}{2}g_{ik}R = -\kappa T_{ik} \tag{23}$$

where

$$\kappa = \frac{6}{m_0^2}. \tag{24}$$

That the gravitational constant has turned out to be positive is a definitive outcome of the theory. The choice of coupling constant $\lambda_a\lambda_b, \ldots$ as unity does not play any role in this; any arbitrary sign of $G$ produces the same result. This is not the case in relativity wherein the sign of $G$ is arbitrarily fixed. The above theory also does not allow the arbitrariness of the $\lambda$-term; $\lambda = 0$ is the only possibility consistent with linearity and conformal invariance. In this context we refer the readers to the criticism of Deser and Pirani (1965) and its refutation by Hoyle and Narlikar (1966c).

It is interesting to note that unlike the Brans–Dicke theory, here the predictions are the same as of general relativity for such experimental tests as the bending of light, the perihelion precession, etc. This is *not*, however, an attractive feature of the theory which is now open to the criticism that nothing new has been gained. In the rest of this chapter we shall consider those features in which the theory genuinely differs from relativity.

The above approach is entirely in terms of the direct particle fields. It is, however, possible to recast the theory as a proper field theory according to the prescription given by Narlikar (1968). There is an "inertial" field $m(X)$ which, along with $g_{ik}$, is described through an action principle:

$$\mathscr{A} = \int [\tfrac{1}{12} Rm^2 + m^i m_i + \mathscr{L}](-g)^{1/2} d^4x \tag{25}$$

where $\mathscr{L}$ has the same interpretation as in Eq. (2.63). The discussion of most of the remaining topics in this chapter is applicable to the field version as well, except where explicitly stated otherwise.

## 7.5 The cosmological singularity

We note that the transition to general relativity was made through Eq. (22). The transition is valid provided $\Omega$ is admissible. It may well be that $\Omega$ is zero or infinite at a set of points of four dimensional measure zero. In that case care is needed in the interpretation of the solution of Einstein's equations where the part of the space-time manifold *includes* such points.

One example of this is the neighbourhood of the world line of the typical particle $a$. Since $a$ contributes to $m(X)$, it is expected that $m(X)$ will diverge in the neighbourhood of the world line of $a$. Hence in such cases a more careful analysis of the equations is needed. This has been done by Hoyle and Narlikar (1974) and by Islam (1968).



Cosmologically, however, the interesting situation arises in the neighbourhood of hypersurfaces $m = 0$. How would a relativist interpret such hypersurfaces?

This is illustrated by the example of the simples standard cosmological model: that of a dust filled homogeneous and isotropic universe. Using (20) and (21) we see that the following is a solution:

$$g_{ik} = \eta_{ik},\ m \propto \tau^2, \tag{26}$$

where $\tau$ is the time co-ordinate for the Minkowski spacetime $\mathscr{M}$. Note that $-\infty < \tau < \infty$ represents the whole range of $\tau$ during which $m(\tau)$ decreases from $\infty$ to 0 and then increases to $\infty$. Note that even though the space-time is flat, this universe has the phenomenon of redshift. This is because with the characteristic frequencies of atomic lines increasing in proportion to $m$, a wave emitted at $\tau_1 > 0$ and received at $\tau > \tau_2$ will have a smaller frequency than a similar wave emitted at $\tau$.

If we choose to describe this model in ECF we have to take

$$\Omega \propto \tau^2. \tag{27}$$

It is easy to see that with a co-ordinate transformation we are able to describe the new line element as

$$ds^2 = dt^2 - Bt^{4/3}(dx^2 + dy^2 + dz^2) \tag{28}$$

where $B(>0)$ is a constant and $t \propto \tau^3$. (28) is none other than the simplest Friedmann cosmological solution, the so-called Einstein–de Sitter model (see Section 2).

The transformation to ECF is however not admissible over the entire space-time $\mathscr{M}$. At $\tau = 0$ $(t = 0)$, $\Omega = 0$ and hence it is not correct to carry through the ECF across the zero mass surface. The penalty paid for the "illegal" use of the ECF is the appearance of space-time singularity at $t = 0$. Also, the conventional ideas of the origin of the universe relate to this singularity at $t = 0$. Looked at from the wider concept of conformal invariance the "universe" in fact extends to "the other side" of the zero mass surface, to $t < 0$, and hence does not really originate at $t = 0$. We will return to the observational implications of this concept in Section 9.

The big bang event in the Friedmann models is a highly symmetric singularity, and it was conjectured by some that by reducing the symmetries of a cosmological model the singularity may be averted. Extensive work on space-time singularities in general relativity in recent years has led to the view that singularity is an inherent feature of general relativity, and that the Friedmann case is not an isolated exception (Hawking and Ellis, 1973). This leads therefore to the following question in the above reinterpretation in terms of a zero mass surface: "Is this reinterpretation of space-time singularity of general relativity as a consequence of insisting on the use of ECF at a zero mass surface, peculiar to the Friedmann singularity, or is it a general feature of *all* relativistic singularities?"

A. K. Kembhavi (1978) has demonstrated with the help of several specific examples that the second of the two possibilities is likely to be correct. The basic approach in Kembhavi's work is as follows.

Given a spacetime manifold $\mathcal{M}_{(GR)}$ in general relativity with the space-time singularity, it is possible to find as a solution of the equations (21) the manifold $\mathcal{M}_{(NS)}$ with the following properties:

(i) $\mathcal{M}_{(NS)}$ has no curvature singularity.
(ii) $\mathcal{M}_{(NS)}$ is geodesically complete.
(iii) $\mathcal{M}_{(NS)}$ is cogeodesically complete.†
(iv) There exists a conformal transformation from $\mathcal{M}_{(NS)}$ to $\mathcal{M}_{(GR)}$, i.e., there exists a function $\Omega$ connecting their metrics:

$$ds_{(GR)} = \Omega ds_{(NS)},$$

where $\Omega$ is admissible except on the $m = 0$ surfaces in $\mathcal{M}_{(NS)}$.
(v) All the singularities of $\mathcal{M}_{(GR)}$ arise from the inadmissability of $\Omega$, i.e., they can be identified with the zero mass surfaces of $\mathcal{M}_{(NS)}$.

A rigorous proof of this result has not been obtained, but the diverse cases of singularity considered by Kembhavi suggest that this result must hold in general. The cases considered are outlined below.

7.5.1 *The anisotropic Bianchi Type-I Cosmology* This is the dust universe which expands homogeneously with shear but no rotation. The manifold has the line element

$$ds^2_{(GR)} = dt^2 - A^2(t)dx^2 - B^2(t)dy^2 - C^2(t)dz^2. \tag{29}$$

A general solution for $A$, $B$, $C$ can be obtained but the particular case of "pancake" singularity at $t = 0$ illustrates the general case (Kembhavi, 1977).

$$A(t) \propto t(t + a)^{-1/3}, \quad B(t) = C(t) \propto (t + a)^{2/3}, \tag{30}$$

†The "cogeodesic" corresponds to the track of a particle obtained by varying its world line in the action (14). It is seen that a typical particle $a$ follows the "cogeodesic" given by

$$\frac{d}{ds_a}\left(m_a \frac{da^i}{ds_a}\right) + m_a \Gamma^i_{kl} \frac{da^k}{ds_a}\frac{da^l}{ds_a} = g^{ik} \cdot \frac{\partial m_a}{\partial x^k}.$$

In the ECF a cogeodesic coincides with the geodesic.



where $a$ is a constant $> 0$. The corresponding $\mathcal{M}_{(NS)}$ is given by a similar line element

$$ds^2_{(NS)} = d\tau^2 - X^2(\tau)dx^2 - Y^2(t)dy^2 - Z^2(\tau)dz^2, \tag{31}$$

with

$$\Omega = \frac{t^2}{t^2 + 1}, \quad \tau = t - \frac{1}{t}. \tag{32}$$

Although the $t = 0$ in $\mathcal{M}_{(GR)}$ corresponds to $\tau = -\infty$ in $_{(NS)}$ the transformation is not a trivial co-ordinate transformation. The $m = 0$ surface in $\mathcal{M}_{(NS)}$ is seen to be non-singular and coincides with the $t = 0$ singularity of $\mathcal{M}_{(GR)}$. The manifold $\mathcal{M}_{(NS)}$ satisfies the criteria of nonsingularity outlined on p. 398.

7.5.2 *The cosmological singularity of Belinskii, Khalatnikov and Lifschitz* Belinskii, Khalatnikov and Lifshitz (1970) (abbreviated to BKL) have obtained a general solution of Einstein's equations in the sense that it contains eight arbitrary functions. This solution is a generalization of the Kasner (1921) universe which is a homogeneous anisotropic empty space solution of Einstein's equations:

$$ds^2 = dt^2 - t^{2p_1}(dx^1)^2 - t^{2p_2}(dx^2)^2 - t^{2p_3}(dx^3)^2 \tag{33}$$

with

$$p_1 + p_2 + p_3 = p_1^2 + p_2^2 + p_3^2 = 1. \tag{34}$$

The three parameters $p_\mu$ are expressible in terms of a single parameter $u$:

$$p_1(u) = \frac{u}{1 + u + u^2}, \quad p_2(u) = \frac{1 + u}{1 + u + u^2}, \quad p_3(u) = \frac{u(1 + u)}{1 + u + u^2}. \tag{35}$$

In the BKL solution a synchronous co-ordinate system is chosen and the line element is written in the form

$$ds^2_{(BKL)} = dt^2 + g_{\mu\nu}dx^\mu dx^\nu, \tag{36}$$

with

$$g_{\mu\nu} = -[a^2 1_\mu 1_\nu + b^2 m_\mu m_\nu + c^2 n_\mu n_\nu]. \tag{37}$$

Here $\mathbf{1} = (1_1, 1_2, 1_3)$, **m**, **n** are functions of the space co-ordinates and a, b, c are functions of time. BKL found that the solution (whether the universe is empty or not) tends to a singular epoch $t = 0$, through a succession of

Kasner like regimes. During any regime, a, b, c behaves as

$$a \sim t^{p_1}, \quad b \sim t^{p_2}, \quad c \sim t^{p_3}. \tag{38}$$

However, the parameters $p_1, p_2, p_3$ change through the successive epochs with the result that the universe expands along one direction and contracts along the other two as $t \to 0+$ (backwards in time!); but these directions are themselves changing. It is therefore not possible to say along which directions the final singularity is attained.

Kembhavi (1978) finds that with $\Omega \sim t^n$, $n \geqslant 3$ it is possible to construct a nonsingular $\mathcal{M}_{(NS)}$ with the properties mentioned on p. 398. $\mathcal{M}_{(NS)}$ contains a zero mass surface at $t = 0$.

7.5.3 *The Taub–NUT universe* This universe (Taub, 1951) has many properties which have made its study of interest more to mathematicians than to physicists. It is an empty universe with the line element

$$ds^2_{(GR)} = U^{-1}(t)dt^2 - 4l^2 U(t)(d\psi + \cos\theta\, d\phi)^2 - W(t)(d\theta^2 + \sin^2\theta\, d\phi^2) \tag{39}$$

where

$$\left.\begin{aligned} U(t) &= \frac{(t - t_1)(t_2 - t)}{t^2 + l^2}, \quad l = \text{constant}, \\ W(t) &= t^2 + l^2. \end{aligned}\right\} \tag{40}$$

The Taub space-time is confined to $t \in (t_1, t_2)$ while the NUT (short for its three authors Newman, Tamburino and Unti, 1963) extension refers to $t > t_2$ and $t_1 > t$. The space-time is singular at $t_1$, $t_2$ but without the divergence of any curvature invariants! The singularity comes from the incompleteness of certain geodesics as $t \to t_1$ or $t \to t_2$. The universe is unstable against the introduction of matter (Misner and Taub, 1969).

Taking the Taub part as $\mathcal{M}_{(GR)}$, it is possible to get $\mathcal{M}_{(NS)}$ with the necessary properties for the conformal function

$$\Omega = (t - t_1)^{-n}(t_2 - t)^{-n}, \; n > 1, \; t \in (t_1, t_2). \tag{41}$$

Moreover $\mathcal{M}_{(NS)}$ is stable against the introduction of matter so that it is not inconsistent with Mach's principle. It thus becomes possible to attach a semblance of physical plausibility to $\mathcal{M}_{(NS)}$ as the limiting case of a matter filled universe whose density tends to zero.

These cases suggest that most singularities in GR arise from the "wrong" choice of conformal frame, in a non-singular space-time with zero mass surfaces. It may well be that there is a "hard core" case of singularities of general relativity which do not arise this way. If such a class exists we may be able to gain further insight into the nature of singularities. Another advantage of this type of investigation is that the discussion of physics near or at a singularity is precluded in the relativistic framework. In the above conformal approach the problem is conceptually (if not operationally) simpler: it boils down to discussing the physics of interacting particles of zero masses. In some cases, as we shall see in Section 10 it is even possible to carry the discussion across the zero mass surface.



## 7.6 Anomalous redshifts

The Hoyle–Narlikar theory can be modified in a simple manner to enable one to understand the anomalous redshift observations of quasistellar objects (QSOs in brief).

Before going to this modification let us consider one difficulty that arises with the theory presented so far. It is easily verified that the response condition (7.20) is *not* satisfied for the simple Friedmann cosmology; that is, the advanced and the retarded contributions to inertia do not cancel and divergencies arise. This had been pointed out by Hawking (1965). It was subsequently shown by Hoyle and Narlikar 1972d) that the divergences can be eliminated by going to a path integral formulation of the quantum theory of the above direct particle action. Here we show that by a simple modification the divergence can be eliminated even in the classical framework.

Consider the Friedmann model which can be described by a flat minkowski line element and a variable mass as in (26). Instead of a fixed coupling constants consider constants $\lambda$ which switch sign. Thus along the world line of a typical particle $a$, identify a "switch-over" point $A_0$ and redefine (14) in the form

$$\mathcal{A} = -\sum_{a<b}\sum \iint \varepsilon(A, B_0)\varepsilon(B, A_0) G(A, B)\, da\, db \tag{42}$$

where $\varepsilon(A, B_0) = +1$ if $A$ lies to the future of $B_0$ and $\varepsilon(A, B_0) = -1$ if $A$ lies to the past of $B_0$. This definition presupposes a synchronous co-ordinate system which is usually possible in cosmology. In the Einstein–de Sitter model we get the required sign value for mass if all particles have their switch-over points on the surface $\tau = 0$. The divergence is avoided by the simple artifact that when the particles $a$ and $b$ (see Figure 7.1) are very far apart the advanced and retarded contributions of $b$ to the inertia of $a$ have the opposite signs and hence they cancel each other.

Narlikar (1977) has used this idea to investigate the problem of anomalous redshifts of quasi stellar objects (QSOs). Suppose a particle does not have a switch-over point at $\tau = 0$ as for the whole universe. In Figure 7.2 are shown

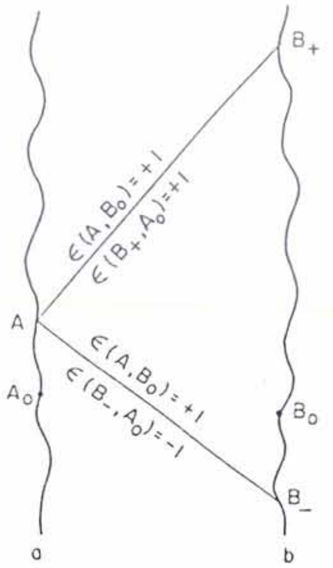



FIGURE 7.1 The world lines of particles $a$ and $b$ are shown, with the switchover points at $A_0$ and $B_0$ respectively. At $A$ the inertial contribution of $B_+$ is positive and of $B_-$, negative.

three particles $w$, $p$, and $q$. $w$ and $q$ have their switch-over points at $\tau = 0$ while $p$ has it at $\tau = \tau_1 > 0$. If $w$ makes an observation at the present epoch $\tau_0$, he sees $p$ and $q$ at epoch $\tau_2$. Calculations show (Narlikar, 1977) that the masses of $p$ and $q$ as seen by $w$ will respectively be

$$m_p = k(\tau_2 - \tau_1)^2, \quad m_q = k\tau_2^{\,2}, \quad k = \text{constant} \tag{43}$$

while $w$ will estimate his own mass to be

$$m_w = k\tau^2. \tag{44}$$

Since redshifts arise in this theory from variable masses, the redshifts of $p$ and $q$ will respectively be

$$z_p = \frac{\tau^2}{(\tau_2 - \tau_1)^2} - 1, \quad z_q = \frac{\tau^2}{\tau_2^{\,2}} - 1, \tag{45}$$

as seen by $w$. Thus it is possible in this model for two extragalactic objects physically near each other to show different redshifts. In the above example we may identify the low (and normal) redshift object $q$ with a galaxy and the high (and anomalous) redshift object $p$ with a QSO. We will return to this idea in Section 11.



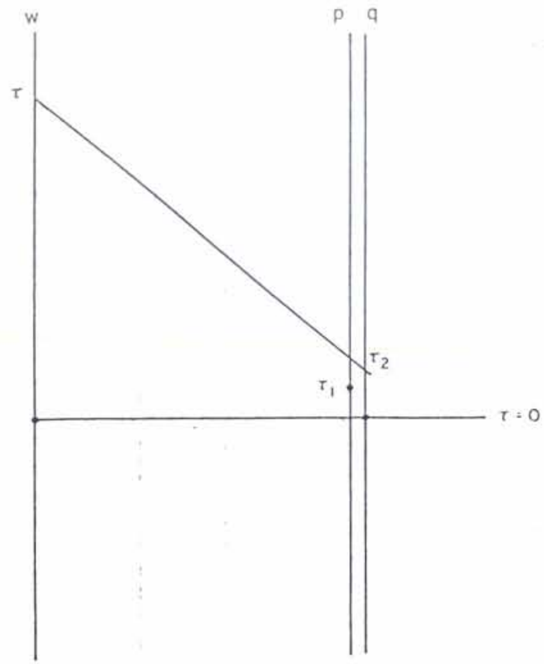



FIGURE 7.2 The anomalous redshift of $p$ arises in the above scenario because its zero mass epoch occurs at $\tau_1 > 0$. (See text for details.)

## 7.7 Cosmology with variable gravity

In 1972 Hoyle and Narlikar had considered two types of models within the framework of their gravitation theory (Hoyle and Narlikar, 1972a, b). The first type was essentially the same as those given by general relativity with the more sophisticated interpretation of space-time singularity as discussed in Section 7.5. The second type admitted creation of matter and led to a new type of model. However, this approach was partly kinematic in the sense that there were no equations to describe creation as in the C-field cosmology (see Section 4). These models led to a variable gravitational constant. Later in 1974 Hoyle and Narlikar used another approach to show how variable gravity models can be obtained from their equations (Hoyle and Narlikar, 1974). The difference between these models and those discussed in Section 7.5 is best illustrated by going over to the simple homogeneous and isotropic cosmology with Minkowski space-time. Suppose we do not require at the outset that (20) holds and we make a power series expansion in the $\tau$-parameter.

$$\left.\begin{aligned} m^{(\mathrm{ret})} &= a + b\tau + C\tau^2 + \ldots, \\ m^{(\mathrm{adv})} &= -a + (B - b)\tau + C\tau^2 + \ldots, \end{aligned}\right\} \tag{46}$$

where $a$, $b$, $C$, $B$ etc. are constants and we ensure that $\tau = 0$ is the zero mass surface. From the gravitational equations we also have in the flat space

$$\Box\phi - 6g^{pq} m_p^{(\mathrm{ret})} m_q^{(\mathrm{adv})} = 0. \tag{47}$$

Using (46) we find that at least one of the two $m^{(\mathrm{ret})}$ and $m^{(\mathrm{adv})}$ must vanish at $\tau = 0$. Without loss of generality we may take $b = 0$ (i.e., $m^{(\mathrm{ret})} = 0$ at $\tau = 0$). If we also have the response condition (20) we then get

$$a = 0, \ B = 0, \tag{48}$$

and we are back to the case considered in Section 7.5. If we *do not* follow the response condition then (48) need not hold and we get new models.

Consider what happens for small $\tau$. Since from the gravitational equations (18)

$$G = \frac{3}{4\pi\phi} \propto [-a^2 + aB\tau + BC\tau^3 + C^2\tau^4 + \ldots] \tag{49}$$

while $m = 1/2B\tau + C\tau^2 + \ldots$ we see that going to a constant—$m$ conformal frame will *not* produce a constant $G$. If we choose a conformal function $\Omega \propto m$, we find that in the new frame the mass function $m^+ = m_0 =$ constant, but the gravitational constant behaves as

$$G^+ \propto \frac{(1/2B\tau + C\tau^2 + \ldots)^2}{(-a^2 + aB\tau + BC\tau^3 + C^2\tau^4 + \ldots)}. \tag{50}$$

We will now take first the case $a = 0$. For $\tau \gg B/C$, we see that $G^+ \sim$ constant, and the universe tends to the Einstein–de Sitter form discussed in Section 7.5. For $B/C \gg \tau$, however, we have

$$G^+ \propto \tau^{-1} \propto t^{-1/2}, \tag{51}$$

and the line element as

$$\begin{aligned} ds^{+2} &= (\text{constant})\tau^2[d\tau^2 - dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\Phi^2)] \\ &= dt^2 - At[dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\Phi^2)]. \end{aligned} \tag{52}$$

with

$$A = \text{constant}.$$

In terms of the cosmological parameters $H_0$ and $q_0$ we have at the epoch $t = t_0$

$$H_0 = \frac{2}{t_0}, \quad q_0 = 1, \quad \frac{\dot{G}^+}{G^+} = -H_0. \tag{53}$$

Thus for small $\tau$ we have a variable gravitational constant.



If $a \neq 0$, (49) predicts a *negative* gravitational constant close to $\tau = 0$. Large repulsive forces close to $\tau = 0$ have the effect of removing "initial" inhomogeneities by evaporating large condensations of matter.

However, as shown by Hoyle and Narlikar (1974), these models do not satisfy the observed experimental tests of general relativity (see Section 9) and so long as these tests are as accurate as they are claimed to be, such models appear to be ruled out. Such tests therefore provide a confirmation of the response condition (20).

## 8. MATTER–ANTIMATTER COSMOLOGIES

We consider in this chapter cosmological models which are based on the assumption that the universe contains as much anti-matter as matter. We will consider in some detail the symmetric models of Alfven and Klein, and Omnes, and then the observational evidence for the existence of antimatter in the universe.

### 8.1 Motivation

It follows from relativistic quantum mechanics that for every particle there must be an antiparticle, and that when a particle is created or destroyed so is an antiparticle. We should expect from this that the universe contains as many antiparticles as particles. Antiparticles can combine to form antiatoms and antimolecules which have the same physical properties as atoms and molecules, and there is no reason why there should not be stars and galaxies made entirely of antimatter. It would be very difficult observationally to distinguish these from their "normal" counterparts because electromagnetic radiation behaves symmetrically with respect to matter and antimatter.

All objects in our neighbourhood are known to be made of matter and not antimatter, and it is an assumption of standard cosmology that this asymmetry extends to the whole universe. In order to have at the present epoch a world made only of matter, it is necessary to postulate at the time of the big band a small excess of particles over antiparticles ($\sim 1$ part in $10^9$). All the antiparticles would be annihilated by an equal number of particles during the dense phase, and all subsequent epochs would be matter dominated. The 3°K background can be identified with the annihilation radiation. Such a process would mean that particle–antiparticle symmetry is violated at the moment of the origin of the universe.

It is the view of people like Alfven, Klein, Omnes and their co-workers that the universe is symmetric between particles and antiparticles. They have tried to build cosmological models which evolve from a symmetric initial state. In all such models it is necessary to provide a mechanism to separate matter from antimatter on a scale consistent with observation, for if the matter and antimatter remained uniformly mixed together, there will be total annihilation in the universe. It is also necessary to study the possible effects of the presence of antimatter, such as the annihilation radiation which must occur when matter and antimatter come together, and to compare these with observations. We will review in the subsequent parts of this section some matter–antimatter cosmological models and their predictions. We will see in the last section what light observations shed on the possible large scale presence of antimatter in the universe.

## 8.2 The cosmological model of Alfven and Klein

This model is the culmination of the ideas of O. Klein (Klein, 1958; Alfven and Klein, 1963; Alfven, 1965). In it the initial state of the universe is taken to be a thin, very low density plasma which is a homogeneous and perfectly symmetric mixture of matter and antimatter. Because of the low density, the rate of annihilation is very low. The plasma occupies a volume which is very much larger than the present volume of the metagalaxy, i.e., the observable universe.

For definiteness Alfven and Klein consider a homogeneous, low density, symmetric sphere of protons and antiprotons. It is assumed that the initial temperature is non-zero and that a magnetic field is present. The first is necessary to cause matter–antimatter collisions, and the second to provide a mechanism for the separation of matter from antimatter. The sphere collapses under its own gravity, and as the density increases the annihilation rate goes up. The ensuing radiation exerts a pressure opposing gravity and turns the collapse into an expansion. It is assumed that Newtonian gravity is applicable throughout the development of the sphere of matter and anti-matter representing the metagalaxy.

As discussed in Section 3 the sphere contracts under its own gravity which is proportional to the distance $r$ from the centre. The velocity of a particle at a distance $r$ from the centre is given by

$$v = \eta r, \quad \eta = \tfrac{2}{3}(t_0 - t)^{-1} \tag{1}$$

where $t_0$ is the time at which all the mass would be concentrated at the centre if the collapse were to continue unchecked until then. The uniform density of the sphere at time $t$ is

$$\varrho = \frac{1}{6\pi G}(t_0 - t)^{-2}. \tag{2}$$



As the collapse proceeds, the protons and antiprotons in the sphere annihilate each other, and the energy production due to annihilation in a sphere of radius $r$ per unit time is

$$P = \frac{4\pi}{3} r^3 n_p \cdot 2m_p c^2 \cdot \frac{1}{T_p} = \frac{8\pi}{3k_p} \cdot m_p c^3 n_p^2 \sigma_0 r^3, \tag{3}$$

where $m_p$ is the proton mass and $n_p$ the proton number density, $T_p$ is the average lifetime of the proton (antiproton) and $\sigma_0 = \pi d^2$, where $\sigma_0$ is the classical electron radius. $T_p$ is related to $\sigma_0$ through

$$T_p = (n_p c \sigma_0)^{-1} k_p, \tag{4}$$

where $k_p$ is a function of the proton energy which is of the order unity up to relativistic energies. The outward flow of momentum per unit area per unit time due to the energy produced is

$$F = \frac{P}{4\pi r^2 c} = \frac{c^2 \varrho^2 \sigma_0 r}{6k_p m_p}, \tag{5}$$

where $\varrho = 2n_p m_p$ is the matter density.

The annihilation results in the production of mesons, neutrinos and high energy $\gamma$-rays (see Section 8.3 for the details). After the $\mu$ and $\pi$ mesons decay there are about four $e^{\pm}$ of energy around 100 MeV each per annihilation. These spiral in the magnetic field and emit synchrotron radiation until they annihilate each other and emit $\gamma$-rays. If all the energy flow were carried by low energy electromagnetic radiation, its pressure effect on the gas would be given by the Thomson cross-section $(8\sigma_0/3)$ of the electrons. But for $\gamma$-rays approaching 1 MeV the cross-section decreases by an order of magnitude, until above 10 MeV it increases again due to pair production. A considerable fraction of the energy also goes into neutrinos. Taking these factors into account Alfven and Klein suggest that the force per electron at radius $r$ should be

$$f_e = \varepsilon \cdot \frac{8\pi\sigma_0 F}{3} = \frac{4\varepsilon\varrho^2 \sigma_0^2 c^2 r}{9k_p m_p}, \tag{6}$$

with $\varepsilon$ between 0.1 and 1.0. This radiation pressure exerts an outward force $f_r dV$ on a volume element $dV$, with

$$f_r = \frac{8\varepsilon\varrho^2 n_e \sigma_0^2 c^2 r}{9k_p m_p}, \tag{7}$$

where $2n_e$ is the number of electrons and positrons per unit volume. This force opposes the gravitational force

$$f_g = -\frac{4\pi\varrho^2 G r}{3} \tag{8}$$

per unit volume, and the resultant may be expressed as

$$\varrho\left[\frac{\partial v}{\partial t} - v\frac{\partial v}{\partial r}\right] = f_r + f_g. \quad (9)$$

Upon using Eqs. (1), (7) and (8) and taking as units of length, mass and time the quantities $(c^2d^2/Gm_p) = 6.3 \times 10^{26}$ cm, $(d^2/Gm_p) = 6.7 \times 10^8$ yrs and $(c^4d^2/G^2m_p) = 8 \times 10^{54}$ gms, the equation of motion reduces to

$$\dot{\eta} + \eta^2 = \frac{4\pi}{3}\left[\frac{2\pi\epsilon n_e}{3k_p} - 1\right]. \quad (10)$$

The equation for the time rate of change of $\varrho$ is

$$\dot{\varrho} = -3\eta\varrho - \frac{\pi}{2k_p}\varrho^2, \quad (11)$$

where the last term represents the decrease in the mass density due to proton–antiproton annihilation. The equation for the time rate of change of the electron (positron) number density is

$$\dot{n}_e = -3\eta n_e - \frac{\pi}{k_e}n_e^2 + \frac{\pi\varrho^2}{2k_p}, \quad (12)$$

where the second term on the right represents the decrease in $n_e$ due to $e^+ - e^-$ annihilation and the last term represents the $e^\pm$ created by $p - \bar{p}$ annihilation.

8.2.1 *The solutions of Bonnevier* Bonnevier (1965) has obtained solutions of this system of equations which show that it is possible in some cases for the radiation pressure to overwhelm gravity and turn the implosion into an expansion. If $n_e$ and $\varrho$ are assumed to be related by $n_e = a\varrho$, (12) reduces to (11) for

$$a = a_0 = \frac{k_e}{4k_p}\left[1 + \sqrt{\left(1 + \frac{8k_p}{k_e}\right)}\right], \quad (13)$$

where $k_e$ plays the same role in determining $e^\pm$ lifetime as $k_p$ does for $p\bar{p}$ (see Eq. (4)). $k_e$ is of the order unity for non-relativistic energies, but increases logarithmically for relativistic energies. If $a$ differs from $a_0$ it will tend to $a_0$, because $\dot{a} > 0$ for $a < a_0$ and $\dot{a} < 0$ for $a > a_0$, and it is sufficient to use $n_e = a_0\varrho$. There are then only two independent equations. It is convenient to introduce the variables $R$ and $M$ which are related to $\eta$ and $\varrho$ by

$$\eta = \frac{\dot{R}}{R}, \quad \varrho = \frac{3M}{4\pi R^3}, \quad (14)$$



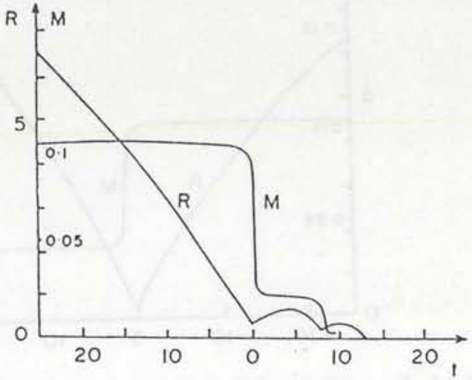



FIGURE 8.1 Time variation of the radius $R$ and mass $M$ of the metagalaxy for $K = 2$, $k_p = 1$, $\tau = 0$, where $K$ is the ratio between the force due to radiation pressure and the gravitational force, $k_p$ is proportional to the annihilation cross-section for protons and $\tau$ is the time delay between the annihilation and the force exerted on the electrons. The metagalaxy shrinks indefinitely with a superimposed oscillation. Almost all the matter–antimatter content is annihilated. Units used are time $6.7 \times 10^8$ years, length 210 Mpc. Mass $4 \times 10^{21}\,M_\odot$.

and satisfy the equations

$$\frac{\ddot{R}}{R} = \frac{M}{R^2} - \frac{2k_pKM}{\pi R^2}, \quad \dot{M} = -\frac{3M^2}{8k_pR^3}, \quad (15)$$

where

$$K = \frac{2\pi\epsilon a_0}{3k_p}. \quad (16)$$

$R$ can be taken to be the radius of the metagalaxy at time $t$ and $M$ the mass. Bonnevier has numerically integrated the equations (15) under the assumption that the sphere collapses at a very early epoch with zero inward velocity and infinite radius; the constant $t_0$ in Eq. (1) is set equal to zero. The initial mass $M_0$ is taken to be 0.1 in the units $8 \times 10^{54}$ gma mentioned above. For $k_p = 1$, $K = 2$ the solution shown in Figure 8.1 is obtained.

It is seen that the metagalaxy collapses, consumes a part of its mass, expands and again enters a contracting phase. For larger values of $K$ ($K$ is of order of magnitude unity) the mass consumption is smaller and the expansion phases are of longer duration. There is no value of $K$ for which the expansion continues for ever.

It has been assumed throughout the above discussion that there is no time lag between annihilation and the establishment of the electromagnetic field exerting a force on the electrons. Bonnevier has suggested that such a time

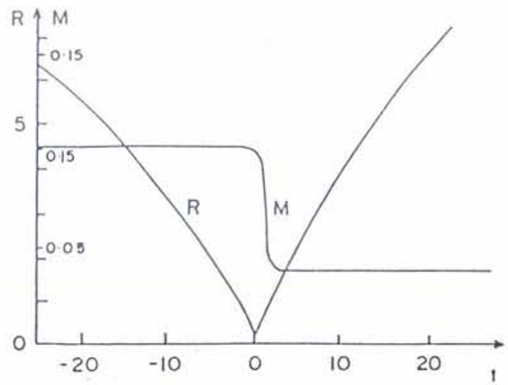



FIGURE 8.2 Time variation of the radius $R$ and the mass $M$ of the metagalaxy for $K = 2$, $k_p = 1$, $\tau = 0.4$ (see Figure 1 for the meaning of these constants and the units used). There is outward motion with finite velocity and finite mass for $t \to \infty$.

lag can be allowed for in a crude approximation by replacing $M(t)$ in the equations (15) by $M(t - \tau)$, with $\tau$ taken to be 0.4. If this is done, the solution for $k_p = 1$, $K = 2$ is of the form shown in Figure 8.2; it represents a metagalaxy that expands for ever after the bounce at $t = 0$.

Steigman (1973) has pointed out a difficulty with the Alfven–Klein models. He defines a free fall time for the collapse by

$$t_{ff}^{-1} = \frac{1}{\varrho}\frac{d\varrho}{dt} = (24G\varrho)^{1/2} \cong 2 \times 10^{-3}\varrho^{1/2}, \tag{17}$$

where Eq. (2) is used. The rate of annihilation of $p\bar{p}$ pairs may be written as

$$t^{-1}_{\text{ann}} = n_p\sigma v \cong 6 \times 10^{-7}\,\sigma v\varrho, \tag{18}$$

and the ratio of the two times is

$$\frac{t_{\text{ann}}}{t_{ff}} \cong \frac{3 \times 10^{-27}}{\sigma v\varrho^{1/2}}. \tag{19}$$

During the early stages of contraction, $\varrho$ is very low and $t_{\text{ann}} \gg t_{ff}$ so that there is hardly any annihilation. When the density is high the annihilation is important, and the bounce must occur at a density $\varrho_b$ such that $t_{\text{ann}} \cong t_{ff}$, i.e.

$$\varrho_b \cong \frac{3 \times 10^{-54}}{(\sigma v)^2}. \tag{20}$$

For a low temperature plasma, $T \lesssim 100°\text{K}$, $\sigma v \sim 10^{-11}\,\text{cm}^3\,\text{sec}^{-1}$, and varies as $T^{1/2}$. Also at low temperatures hydrogen and antihydrogen are formed and the atom–antiatom annihilation rate coefficients are very large, with $\sigma v \gtrsim 10^{-10}\,\text{cm}^3\,\text{sec}^{-1}$. This gives



$$\varrho_b \lesssim 3 \times 10^{-34}\,\text{gm}\,\text{cm}^{-3}, \tag{21}$$

i.e., the density at the bounce $\varrho_b$ is at least two orders of magnitude less than the present density of visible matter. However if we put $v \cong c$ in Eq. (20) as Alfven and Klein do (see Eq. (4)), $\sigma v \cong \sigma c \cong 10^{-15}\,\text{cm}^3\,\text{sec}^{-1}$ and $\varrho_b \cong 10^{-23}\,\text{gm}\,\text{cm}^{-3}$, which is many orders of magnitude more than the present density.

8.2.2 *The relativistic solutions of Laurent and Soderholm* The equations of Klein and Alfven for the evolution of the metagalaxy and the solutions of Bonnevier were obtained on the assumption that Newtonian gravity and nonrelativistic mechanics are applicable throughout the development of the metagalacy. Laurent and Soderholm (1969) have used the framework of general relativity to study the same physical processes. They consider the collapse of a sphere with initial mass $M_0$, the metric inside which is given in standard Schwarzschild co-ordinates by

$$ds^2 = e^{\nu}dt^2 - e^{\lambda}dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\Phi^2) \tag{22}$$

where $\nu$ and $\lambda$ are functions of $r$ and $t$.

The sphere is taken to collapse from an original thin cloud of protons, antiprotons, electrons and positrons. The effects of annihilation, scattering of radiation on matter, interaction of radiation and gravitation, and the interaction of gravitation and matter are all treated using the appropriate general relativistic equations. The system of equations is integrated numerically. The free parameters used are (i) the initial mass $M_0$ (ii) the inverse lifetime of matter divided by the density of matter $a$ and (iii) the radiation scattering cross-section per unit mass $\beta$. $a$ and $\beta$ often appear in the combination $k = c\beta/a$. The standard value of $a$ is $a_0 = c\pi d^2/2m_p$, and the standard value of $\beta$ is $\beta_0 = 8\pi d^2/3m_p$. For these values $k$ is $4/\sqrt{(3)}$.

The solutions obtained for $a = a_0$, $\beta = \beta_0$, $k = k_0$ and various values of the mass parameter are shown in Figure 3. The units of $M_0$, $R$ and $T$ are $8.7 \times 10^{53}$ gms, $1.4 \times 10^{8}$ lyrs and $1.4 \times 10^{18}$ yrs respectively, where $R$ is the Schwarzschild radius at the surface and $T$ the Schwarzschild time. A bounce at $R \cong 0.25$ occurs in all the solutions shown. In contrast to the non-relativistic solutions, it is not necessary to introduce an arbitrary timelag between the annihilation and the exertion of force in order to produce the bounce. For $M_0 \ll 1$ (e.g., $M_0 = 0.005$) the solutions are nearly the same as Bonnevier's solutions. The radius at which the bounce occurs, $R \cong 0.25$, corresponds in ordinary units to $R \cong 3.2 \times 10^{25}$ cms. For $M_0 \cong 0.22$, the

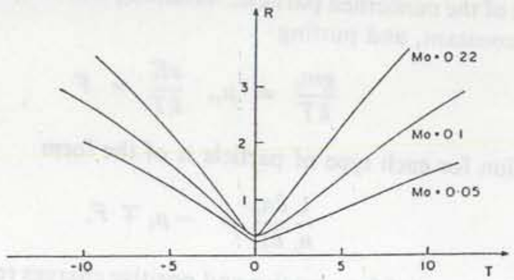



FIGURE 8.3 Schwarzschild radius $R$ at the periphery plotted against the Schwarzschild time $T$, for the standard value $k = k_0 = 4/\sqrt{3}$, with different choices of the initial mass $M_0$.

Schwarzschild radius is $R_S \cong 2.8 \times 10^{25}$ cms, i.e., the radius at the bounce is comparable to the Schwarzschild radius. This makes it clear that it is indeed necessary to use the general relativistic framework to study the evolution of the metagalaxy.

For $k = k_0$, the maximum value of $M_0$ for which a bounce can occur is $M_0 \cong 0.24$, which is about $2 \times 10^{53}$ gms, which is an order of magnitude less than the estimated mass of the observed part of the universe. Steigman[5] has arrived at the same order of magnitude for the maximum mass which can bounce. His calculations are independent of the detailed nature of the equations and their solutions. It is possible to increase the maximum permissible mass by increasing the values of $a$ and $\beta$ above their standard values, but it is not clear that these high values are physically meaningful. The smallness of the maximum initial mass permissible for a bounce to occur is a serious limitation of the Alfven–Klein models.

8.2.3 *Separation mechanisms in the Alfven–Klein model* We have seen that the initial state of the Alfven–Klein model is taken to be a homogeneous mixture of particles and antiparticles. It is necessary to provide a mechanism for the separation of the homogeneous mixture into regions containing only matter or antimatter, for otherwise the universe would soon annihilate. The interstellar medium in a galaxy has a particle density of $n \cong 1$ particle $\text{cm}^{-3}$. If the galaxy were made of protons and antiprotons homogeneously mixed, the lifetime of the galaxy would be of the order of magnitude of the proton (antiproton) lifetime against annihilation, which would be $T_0 \cong$ (average density $n \times$ annihilation rate coefficient $\sigma c)^{-1} \cong$ (1 part. $\text{cm}^{-3} \times 1.5 \times 10^{-15}\,\text{cm}^3\,\text{sec}^{-1})^{-1} \cong 2 \times 10^7$ yrs. It is clear that if the galaxies were all homogeneous mixtures the universe would have annihilated long ago. There is some evidence to show that the separation matter and antimatter in the universe is at least on the scale of clusters of galaxies, and it is necessary to



provide a mechanism which would achieve this state beginning with a homogeneous mixture.

It was speculated by Goldhaber (1956) that our "cosmos" could be a part of the universe where nucleons prevail over antinucleons as a result of a very large statistical fluctuation, which is compensated by an opposite situation elsewhere. Alpher and Herman (1958) have shown that this possibility can apparently be ruled out. They supposed that the presently observable universe derived from a state early in the universal expansion in which thermodynamic equilibrium prevailed between nuclei and antinuclei, and considered an ensemble of finite comoving volume elements $V$, each containing $N$ protons, antiprotons, neutrons and antineutrons, a total of $4N$ particles per volume element. If $a$ is the probability of annihilation, averaging over the ensemble of finite volume elements should yield $4aN$ as the mean number of particles annihilating per finite volume element $V$. The standard deviation of the mean number of nuclei undergoing annihilation is $\sigma = [4a(1 - a)N]^{1/2}$ which has a maximum at $a = 1/2$, with $\sigma_{max} = N^{1/2}$. In a finite volume element $V$, the excess of nucleons over antinucleons will be of the order of $\sigma_{max} = N^{1/2}$.

Suppose that at some early time the temperature of the universe was of the order of $10^{10}\,°\text{K}$, adequate enough to support thermodynamic equilibrium among nucleons, antinucleons and radiation. In such a state the number of nucleons and antinucleons should have been of the order of the number of photons. Assuming the radiation field to be blackbody, the photon density $N_\gamma = (kT/hc)^3 \cong 10^{31}\,\text{cm}^{-3}$ at $T = 10\,°\text{K}$. For a volume element $V$ with radius equal to the present radius of the universe, $\sim 10^{10}$ lyrs, the total number of particles $\cong 10^{31} \times 10^{84} \cong 10^{115}$. For this $N$, $\sigma_{max} = N^{1/2} \cong 10^{57}$, which is no more than the number of particles in the solar system.

The calculations of Harrison (1967) also show that statistical fluctuations are insufficient to produce separation between matter and antimatter on the galactic scale. The fluctuations considered by Harrison are spatial fluctuations of the number densities of baryons and antibaryons. Because of the inhomogeneity, the baryon number in a given co-ordinate volume $V$, $\Delta N = N - \bar{N}$, could be positive or negative, where $N$ and $\bar{N}$ are the number of baryons and antibaryons respectively in $V$. Even though $\Delta N$ could be a very small fraction of $N + \bar{N}$ in the early epochs, as the universe expands the baryons and antibaryons annihilate each other and $\Delta N$ grows to be a considerable fraction of the number of baryons and antibaryons in $V$, making it possible for separate regions of matter and antimatter to be formed.

Since baryon number is conserved we have $\Delta N_{then} = \Delta N_{now} = N_{now}$. If the initial fluctuations occur when matter and antimatter are still in equilibrium, we have $N_{then} = (N_\gamma)_{then} = (N_\gamma)_{now}$, since the number of photons $N_\gamma$ is a comoving volume $V$ is conserved as the universe expands. It follows that

$$\left.\frac{\Delta N}{N}\right|_{\text{then}} = \left.\frac{N}{N_\gamma}\right|_{\text{now}} = \left.\frac{n}{n_\gamma}\right|_{\text{now}} \tag{23}$$

where $n$ and $n_\gamma$ are the proper number densities of baryons and photons respectively. It is known that $(n/n_\gamma)_{\text{now}} \cong 10^{-9}$, so that a small fluctuation

$$\left.\frac{\Delta N}{N}\right|_{\text{then}} \cong 10^{-9} \tag{24}$$

in the early epochs of the universe is amplified by annihilation to $(\Delta N/N)_{\text{now}} = 1$. But even such a small fluctuation cannot be produced statistically. If $\Delta N$ is statistical then we have $\Delta N \cong N^{1/2}$. Now if we demand the largest possible statistical fluctuation in the universe we get

$$\Delta N_{\text{then}} \cong N^{1/2}{}_{\text{then}} \cong (N_\gamma)^{1/2}{}_{\text{then}} \cong (N_\gamma)^{1/2}{}_{\text{now}} \cong 10^{44}$$

i.e.

$$\left.\frac{\Delta N}{N}\right|_{\text{then}} \cong 10^{-44} \tag{25}$$

which makes the largest possible statistical fluctuation many orders of magnitude smaller than the fluctuation required to produce complete separation now.

Alfven and Klein have considered the possibility of separating matter from antimatter using gravitation and electromagnetic fields. We will briefly review their separation mechanism. A detailed account may be found in Alfven (1965).

Alfven and Klein have studied the properties of a plasma consisting of a mixture of matter and antimatter, called an *ambiplasma*, which has four constituents, $n_p^+$ protons, $n_p^-$ antiprotons, $n_e^+$ positrons and $n_e^-$ electrons per unit volume. The number densities are so small that annihilation can be neglected. The charge neutrality condition is taken to be satisfied so that

$$n_p^+ + n_e^+ = n_p^- + n_e^- . \tag{26}$$

The plasma is acted upon by a gravitation field with intensity $g$ in the $-z$ direction. In general this polarizes the ambiplasma so that an electric field $E$ is produced. Assuming all variables to depend only on the co-ordinate $z$, each component of the plasma satisfies the equation

$$\frac{\partial}{\partial z}(n_i kT) = (-gm_i \pm e_i E)n_i \tag{27}$$



where $k$ is Boltzmann's constant, $T$ the temperature and $m_i$ the mass and $e_i$ the charge of the concerned particle. Assuming that the plasma is isothermal, i.e., $T$ = constant, and putting

$$\frac{gm_i}{kT} = \mu_i, \quad \frac{eE}{kT} = F \tag{28}$$

the equation for each type of particle is of the form

$$\frac{1}{n_i}\frac{\partial n_i}{\partial z} = -\mu_i \mp F, \tag{29}$$

with the $\mp$ sign used for negative and positive charges respectively. The four equations (29) are combined with the charge neutrality condition (26) to obtain solutions.

The solutions show that in the case of an asymmetric ambiplasma, which for example contains more matter than antimatter, there are three different regions. The lowest region and the highest region contain a symmetric ambiplasma with no electric field, but the intermediate region has a Rosseland field and contains almost no antimatter. Almost all the antiprotons are accumulated in the lowest region together with an equal number of protons, and almost all the positrons are accumulated in the upper region together with an equal number of electrons. The excess of matter accumulates in the intermediate region. A gravitational field therefore has the tendency to separate the heavy component of an ambiplasma from its light component, and to separate the excess of matter or antimatter from the rest of the ambiplasma.

If the plasma contains equal amounts of matter and antimatter, i.e., if it is symmetric, the solution is

$$n_p^- = n_p^+ = \nu_p \exp(-\mu_p z), \; n_e^- = n_e^+ = \nu_e \exp(-\mu_e z), \tag{30}$$

where $\nu$ is the value of $n$ at $z = 0$. The electric field is zero, and there is no coupling between the heavy component and the light component, i.e., the total number of protons–antiprotons is independent of the total number of electrons–positrons. It is seen that the gravitational field cannot separate the ambiplasma when it is symmetric into matter and antimatter. However, this is possible under the action of electromagnetic effects.

Consider an electric current antiparallel to $g$. This will remove positive charges from the top of the atmosphere, and since the highest region contains mainly positrons and electrons (the scale height of the $e^+ - e^-$ gas is 1840 times the scale height of the $p - \bar{p}$ gas), it is essentially the positrons that are removed. The current removes negative charges from the base of the atmosphere, and since the lowest region contains mainly heavy particles, the result will be essentially a loss of antiprotons. Hence the current will remove antimatter, leaving an asymmetric ambiplasma with an excess of matter which will accumulate in the intermediate region. A vertical current parallel to $g$ will bring about an excess of antimatter. Vertical currents may be produced by hydromagnetic effects which are caused by motion of magnetized ambiplasma during the development of the metagalaxy, or in connection with the formation of galaxies.

The separation of matter or antimatter out of a symmetric ambiplasma is similar to the process of electrolysis. The quantity $M$ of matter or antimatter which is separated by a current $I$ flowing during a time $T$ is

$$M = \frac{M_H IT}{e} \tag{31}$$

where $M_H$ is a constant.

Suppose that a current $I = \pi r^2 i$ flows along the axis of a circular cylinder with radius $r$. The magnetic field at the surface of the cylinder is

$$B = \frac{2I}{cr} \tag{32}$$

(it should be noted that the presence of a magnetic field does not change the equilibrium state of an isothermal atmosphere). Combining Eqs. (31) and (32) it follows that

$$M = \frac{M_H TcrB}{2e} \cong 0.5 \times 10^{-4}\, TrB. \tag{33}$$

The magnetic fields in space $B \leqslant 10^{-4}$ G. For $T$ of the order of $3 \times 10^9$ yrs,

$$M < 0.5 \times 10^9\, r\,\text{gm}. \tag{34}$$

Hence a separation on a galactic scale, $r \cong 10^{23}$ cms, would not be sufficient to separate one Solar mass.

A small scale turbulence with dimension 0.1 lyr or less can produce a high degree of separation between matter and antimatter. The mass separated by each turbulene element is not very large, but the total mass separated can be a large fraction of the whole ambiplasma.

If regions of matter and antimatter are brought together, the matter–antimatter annihilation at the interface produces a very hot layer between the matter and antimatter known as the Leidenfrost layer. A pressure is produced in this layer by the electrons and positrons which are products of $p\bar{p}$ annihilations, and there is an apparent repulsion between regions of matter and antimatter. There is no such repulsion between regions of similar



content. It is possible that because of this there is a coalescence of the small regions to larger volumes of matter and antimatter. A small-scale separation process may in this way ultimately lead to masses of pure matter and antimatter on the scale of galaxies and clusters of galaxies. A detailed theoretical analysis of these possibilities remains to be done. Some problems associated with matter–antimatter boundary layers have been considered by Lehnert (1977).

Nelson and Rowlands (1975) have considered a separation mechanism which is based on the two stream instability which may occur whenever two charged species of a plasma stream relative to each other. In the Alfven–Klein model the relative streaming of leptons and nucleons can occur in two ways. Firstly, the self-pressure of leptons produced in annihilations will counteract the gravitational attraction of the part of the spherical meta-galaxy acting on them, so that they will collapse less rapidly than the non-relativistic nucleons, thus leading to a relative drift in the radial direction. Secondly, the radiation produced in nucleon–antinucleon annihilations will act more strongly on leptons than on nucleons and in the absence of magnetic fields this will also produce a radial drift between leptons and nucleons.

A linear perturbation analysis is made of the cold fluid equations for a symmetric matter–antimatter plasma in which leptons stream with a velocity $v$ relative to nucleons. The analysis shows that a wave develops in which adjacent half-wavelengths are alternately enriched in protons and antiprotons, i.e., a separation occurs. Although perturbations of infinitely long wavelengths are unstable, the growth rate decreases with increasing wavelength, and the maximum growth rate occurs at the instability threshold. It turns out that the wavelength for maximum growth rate is of the order of a few light seconds only. This means that a coalescence mechanism between like cells would again be necessary to produce large units of matter or antimatter.

## 8.3 Baryon symmetric big bang cosmology

We consider in this section the baryon symmetric big bang cosmology developed by Omnes and others. The calculations which describe the separation and coalescence mechanisms operative in the early stages of this model are very involved. We will give here only a bare qualitative outline of some of the key features of this model. The interested reader is referred to the more detailed reviews by Omnes (1972), Omnes and Puget (1974), Stecker (1978) and Steigman (1976) and the original papers listed therein.

Omnes considers a big bang model which is initially at a very high temperature and density. Bosons and fermion-pairs exist in thermodynamic equilibrium with photons of the black-body radiation field at $T \geqslant Mc^2/k$, where $M$ is the rest mass of the particle considered (for nucleons $T \cong 10^{13}$ °K). When equilibrium exists the number of baryon–antibaryon pairs in the universe is the same as the number of photons. In contrast to this, at present the number of baryons is a small fraction of the number of photons, $\eta \equiv N_B/N_\gamma \cong 10^{-9}$, but this fraction is still very much larger than that which would remain in thermodynamic equilibrium with black-body radiation at 2.7°K. In fact the ratio $\eta \cong 10^{-9}$ would be reached at $T \cong 30$ MeV ($3.6 \times 10^{11}$ °K). There are two possible ways to account for this descrepancy: (1) there was initially a small excess (~ 1 part in $10^9$) of baryons over antibaryons (this possibility was discussed in Section 2) or (2) the universe is symmetric in particles and antiparticles and a separation of matter from antimatter occurs at $T \cong 30$ MeV. A baryon symmetric model based on the second possibility has been developed by Omnes and others, who outlined how small regions of matter and antimatter are formed in the dense phase of the universe and how these can grow to galactic size during the subsequent evolution.

According to this scheme, when the universe was above a critical temperature ($T > 300$ MeV, i.e., $3.6 \times 10^{12}$ °K) and critical density, a phase transition occurred, creating a structure in which there were domains containing mostly matter and mostly antimatter. Subsequent annihilation pressure in the Leidenfrost layers between the matter and antimatter regions led to coalescence of regions of the same type together, by pushing apart regions of unlike baryon number.

One of the main concepts of Omnes' model is that a nucleon and antinucleon pair within a certain distance $r_0$ ($\cong$ 1 Fermi) of each other will form a bound state in which the two particles lose their separate identity and behave like a meson with appropriate quantum numbers. The number density of mesons is determined by statistical equilibrium at temperature $T$ so that there is a statistical exclusion principle acting to keep nucleons and antinucleons from occupying the same cell. No such principle applies for particles with like baryon number, and the net effect is a repulsion between nucleons and antinucleons. In this situation it has been shown that a phase transition occurs – the symmetry between nucleons and antinucleons is spontaneously broken and the thermal plasma breaks up into two phases, a nucleonic phase and an antinucleonic phase. The result is that the universe is divided into an emulsion of domains, each containing mostly nucleons or mostly antinucleons.

The size of the domains of nucleons and antinucleons in the emulsion is characterized by the distance $d$ between two opposing boundaries. The magnitude of $d$ is controlled by the diffusion of baryons into the regions of antibaryons and vice versa. $d$ is given at any time $t$ by $d = (Dt)^{1/2}$, where the diffusion coefficient $D \cong v_t\lambda/3$, with $v_t$ the typical thermal velocity ($= c$) and $\lambda$ the typical mean free path ($\cong$ 1 Fermi). The time $t$ at which the phase separation occurs is $t \cong 10^{-5}$ sec, and $d \cong 10^{-4}$ cm.



As the universe expands the temperature falls below the critical value, the two phases become unstable and the matter and antimatter tend to mix and annihilate. This process is slowed down by the strong pressure produced by the annihilations at the boundary between matter and antimatter. Pions produced during the annihilations are much lighter than the annihilating nucleons and antinucleons and their momenta are very large; this produces a large flux of momentum through the boundary. Such a flux produces a pressure which is exerted up to distances from the boundary of the order of the mean free path $\lambda_\gamma^{\text{h.e.}}$ of the high energy photons produced by the decay of $\pi^\circ s$. When the temperature $T \cong 30$ keV, $\lambda_\gamma^{\text{h.e.}}$ becomes equal to the size $d$ of the separate domains in the emulsion. This signifies the end of the annihilation period. The annihilation pressure gradients now extend over whole cells of matter and antimatter, and momentum can be transmitted on macroscopic scales of the order of the cell size to the fluid as a whole. This is the beginning of the "coalescence period", and the size $d$ grows in such a way that $d \geqslant \lambda_\gamma^{\text{h.e.}}$.

When the temperature falls below 5000°K, electrons and protons combine to form atoms, and the ionization of matter decreases rapidly. The coalescence stops at the time of recombination because the velocity of sound falls sharply towards the thermal velocity of atoms. The coalescence motion would then become supersonic, and this is not allowed. It has been estimated that the total mass in a homogeneous system of matter (or antimatter) of the size $\lambda_\gamma^{\text{h.e.}}$ at the time of combination lies between the mass of a galaxy and that of a compact cluster of galaxies ($10^2$ galactic masses). Stecker and Puget (1972) have considered the formation of galaxies once the stage is reached at which combination takes place.

In the cosmological model of Omnes,

$$\eta \equiv \text{(number of nucleons in comoving volume } V\text{/number of photons in the same volume)}, \tag{35}$$

has the theoretical value

$$\eta = \phi_1\phi_2\phi_3\phi_4 \tag{36}$$

where $\phi_1$ is the ratio of between the densities $N$ and $N_\gamma$ of nucleons and photons in thermal equilibrium at the critical temperature for seperation, $\phi_2$ is the ratio $B/N$ between the excess baryonic density and the nucleonic density somewhat above the critical temperature, $\phi_3$ is the probability of the survival of nucleons during the annihilation period and $\phi_4$ the same quantity during the coalescence period. Numerically one has $\phi_1 \cong 0.4$, $\phi_2$ is poorly known and estimated to range around 0.1, lying within the limits $0.03 \leqslant \phi_2 \leqslant 0.3$, $\phi_3$ is estimated to lie within the limits $10^{-9} \leqslant \phi_3 \leqslant 3 \times 10^{-7}$, and $\phi_4 \cong 0.5$. This gives

$$3 \times 10^{-12} \leqslant \eta \leqslant 10^{-8}. \tag{37}$$

This theoretical range of $\eta$ includes the observed value of $\eta \cong 10^{-9}$. As was mentioned at the beginning of the present section, in standard cosmology the value of $\eta$ can only be explained by assuming an initial small excess of baryons over antibaryons ($\sim 1$ part in $10^9$).

It is believed by many cosmologists that there is a cosmic distribution of light elements like $H^2$, $He^4$, $Li^7$, etc. in the universe and that these light elements were synthesized in the early, hot phases of an evolving universe (see Section 10 for greater detail on the significance of primordial nucleosynthesis). The primordial abundances expected in any model can be compared with the observed cosmic abundances as a test of the particular model. It has been shown by Leroy, Nicolle and Schatzman (1973) that within Omnes' coalescence model, no nucleosynthesis can take place at all. This is because at the time nucleosynthesis would occur ($T \cong 0.1$ MeV), the mean size of the separate domains is smaller than the neutron diffusion length. The neutrons are therefore lost from the domains and annihilated at the boundaries, and nucleosynthesis does not occur. If the observed abundances of the light elements are indeed cosmological, the lack of nucleosynthesis in Omnes' model would be a serious drawback. It has been shown by Combes, Fassi-Fehri and Leroy (1975) that if the size of each cell in the emulsion is much larger than the neutron diffusion length, nucleosynthesis would take place. They have shown that if the observed mass fractional $He^4$ abundance ($\sim 0.22$–$0.32$) is to be accounted for, the minimal dimension of a region of matter necessary is $L \cong 2 \times 10^6 T^{-1}$ MeV cms, which for $T \cong 0.1$ MeV is $L \cong 2 \times 10^7$ cms. At this temperature, the size of the cell expected from Omnes' model is $\sim 1$ cm. A much more efficient coalescence mechanism than is present in Omnes' model is therefore needed to produce the required primordial element abundances.

The radiation emitted in the annihilation of matter and antimatter which takes place during the evolution of the universe in baryon symmetric cosmologies can cause distortion of the microwave background spectrum from its pure black-body form. Sunyaev (1974) has traced the effects due to the release of energy during various epochs. At high temperatures (redshift $z > 10^8$) when plenty of electrons and positrons are present and the density is high enough, there is plenty of bremsstrahlung radiation and absorption and thermal equilibrium can be re-established and the black-body nature of the microwave spectrum restored, however great may be the energy release.



At redshifts $z \leqslant 10^6$, these relaxation processes are absent and equilibrium can no longer be established. For $10^3 \leqslant z \leqslant {}^4$, the energy release leads to a decrease in intensity in the low-frequency (Rayleigh–Jeans) part of the spetrum and an increase in intensity in the high-frequency (Wien) part.

Detailed calculations for the distortions produced have been made by Sunyaev and Zeldovich (1970a, b). They introduce the parameters $y$ and $q$ with

$$y = \int_0 \frac{kT_e}{mc^2}\, d\tau \tag{38}$$

where $\tau$ is the optical depth due to Thomson scattering and $T_e$ the electron temperature, and

$$q = \int \frac{Q(t)}{\delta T_r^4(t)}\, dt, \tag{39}$$

where $Q(t)$ denotes the rate of energy release, $T_r$ the temperature of radiation, $\sigma T_r^4$ the energy density of radiation, and $t$ the cosmological time. The experimental data (Sunyaev, 1974) indicates that deviations of the microwave spectrum from the Planckian form cannot be larger than 30 percent in the region $\lambda = 50$–$20$ cms or larger than 10 percent in the region $\lambda = 20$–$1$ cm. Measurements in the high-frequency region and data obtained from the observation of interstellar molecules set a limit to the parameter $y < 0.15$ and to the magnitude of the energy release $q < 0.8$ during the stage $100 < z < 1400$. At earlier stages of the expansion, the observations give $q < 0.05$ in the range $1400 < z < 4 \times 10^4$ for $\Omega h^2 \sim 1$, ($\Omega = \varrho/\varrho_c$, $H_0 = h \times 50$ km sec$^{-1}$ Mpc$^{-1}$) and $q < 0.01$ in the range $4 \times 10^4 < z < 5 \times 10^5$. The theoretical calculations of Ramani and Puget (1976), who used specific coalescence models, show that the distortions expected from annihilations would be higher than the values estimated from observations unless $\Omega < 0.01$, which is an order of magnitude less than the value of $\Omega$ required to produce the observed spectrum of isotropic gamma radiation (the gamma ray spectrum is discussed in the next subsection). On the other hand, the calculations of Stecker and Puget (1973) produced consistency between theory and observation. Stecker (1978) has listed a number of papers in which the various points of view are presented. A detailed criticism of the various assumptions which go into Omnes' model and its predictions has been made by Steigman (1976).

Brown and Stecker (1979) (see also Senjanovic and Stecker, 1980; Stecker, 1981) have suggested that grand unified theories (GUTs; see Section 10 for a brief review) with spontaneous CP violation in the very early big bang can lead more naturally to a baryon-symmetric cosmology with domain structure, than to a totally baryon-asymmetric cosmology. The symmetry is broken in a randomized manner in causally independent domains, favouring neither a baryon nor an antibaryon excess on a universal scale. The exponential expansion of the universe after the symmetry breaking leads to a huge increase in the size of these domains, which may evolve further leading to the formation of matter and antimatter galaxies in separate regions of the universe.

## 8.4 Evidence for the existence of antimatter in the universe

There is firm evidence to show that none of the constituents of the Solar systems are made of antimatter. Direct contact by space probes with the Moon and the nearer planets has established that these like the Earth are made of matter. Information about the outer planets can be obtained using the Solar wind as a probe. Were any of the planets made of antimatter, annihilation of the Solar wind particles on their surface would have made them strong sources of gamma radiation, which they are not. It may therefore safely be assumed that there is no antimatter in the Solar system.

Direct or indirect evidence for the existence of antimatter outside the Solar system is difficult to obtain. If a star or a galaxy were to be made of antimatter, its physical properties, like the energy levels of the antiatoms and antimolecules in it, would be the same as in the case of similar bodies made of matter. The emission and absorption line systems produced in an object made of antimatter would not therefore shed any light on its nature. It would be of course possible to distinguish antimatter from matter by its response to a magnetic field, but for this the direction of the magnetic field will have to be known independently, and this is not possible for a field in a distant object. We have to depend for evidence primarily on cosmic rays and on the products of annihilation which are expected to reach the Earth if large quantities of antimatter and matter meet and annihilate in interstellar or intergalactic space.

*Direct evidence* Direct evidence for the existence of antimatter in the universe would be obtained if antiparticles and antinuclei were to be observed in primary cosmic rays. However, intensive searches have failed to detect any significant antimatter component in the cosmic rays (see Steigman, 1973, 1976) except for the possible detection of one candidate antinucleus (Hagstrom, 1977). Cosmic rays passing through a few gms cm$^{-2}$ of interstellar matter will produce antiprotons as secondaries. As a result it is expected that antiprotons in cosmic rays at a level of about 1 part in $10^4$ will be found. Since the upper limits on the antiproton component are of this order of magnitude, antiprotons are not very useful as a probe for antimatter. On the other hand



the production of heavy antinuclei as secondaries is negligible, and if the existence of the antinucleus believed to have been detected is confirmed, there would be firm direct evidence for the existence of antimatter somewhere in the universe.

It is difficult to interpret the information that the cosmic rays contain because their origin is still largely obscure. If as is now believed, the overwhelming bulk of cosmic rays are of galactic origin, the upper limits on the antimatter component rule out the existence of large amounts of antimatter in the Galaxy, at least to within about $10^3$–$10^4$ lyrs from the Sun. But the limits are consistent with a possible small extragalactic flux of cosmic rays which may have a substantial antimatter component.

*Indirect evidence* Polarized light passing through a medium will have its plane of polarization rotated through and angle $\Delta\theta$ which is proportional to the integral along the line of sight of the product of the electron density and the component of the magnetic field parallel to the line of sight. This is known as Faraday rotation. Positrons will produce an opposite rotation, and the net effect is

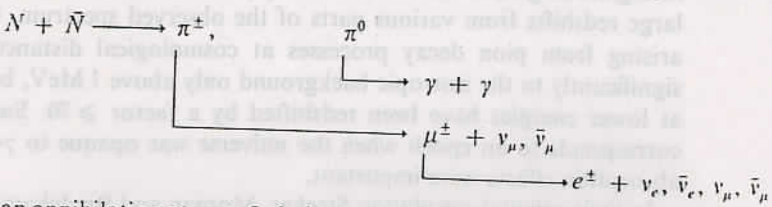
$$\Delta\theta = \int (n_{e^-} - n_{e^+}) B_{\parallel}\, ds \qquad (40)$$

where $n_{e^-}$ and $n_{e^+}$ are the densities of electrons and positrons respectively.

It is known that large scale Faraday rotation is present in the Galaxy (see Woltjer, 1967). There is also some evidence for extragalactic Faraday rotation (Reinhardt and Thiel, 1970). Steigman (1973, 1976) has interpreted this to mean that $\Delta n = n_{e^-} - n_{e^+} \neq 0$ along a typical line of sight, i.e., our Galaxy, external galaxies and the intergalactic medium do not contain equal amounts of matter and antimatter. This interpretation has been criticized by Elvius, Karlson and Laurent (1974). They point out that in the present state of the metagalaxy matter and antimatter must be in the form of separated cells. If the magnetic field in the cells is unidirectional, contributions from different cells tend to cancel. However, if $B_{\parallel}$ changes sign, as we pass from one cell to another, the Faraday rotation will be the same direction in both cells and no cancellation will occur. Thus the existence of antimatter in our Galaxy cannot be excluded. A similar argument applies to intergalactic space as well.

The simultaneous presence of matter and antimatter in some regions could be indirectly detected by the observation of the products of annihilation reaching the earth. In this the 0.5 MeV gamma rays produced in the annihilation of a possible $e^+ - e^-$ component of the matter–antimatter mixture cannot be used. 0.5 MeV $\gamma$-rays are also produced in the annihilation of positrons produced in cosmic ray secondaries, or in pulsars and supernovae, and the detection of such $\gamma$-rays cannot be taken as evidence for the presence of antimatter. Nucleon–antinucleon annihilations prove to be more useful in this respect.

The primary products of a nucleon–antinucleon annihilation are the pions $\pi^{\pm}$ and $\pi^0$. These finally end up as electrons, positrons, muon and electron neutrinos and antineutrinos, and gamma-rays. A typical decay scheme is




In an annihilation at rest, 5–6 pions are produced, with exactly equal number of $\pi^{\pm}$ and roughly equal number of $\pi^0$s. The spectra of all the secondary annihilation products are very similar, extending from several tens of MeV to several hundred MeV, and peaking between 100 and 200 MeV. About half of the energy released is carried away by neutrinos, one-third by $\gamma$-rays and one-sixth by the electrons and positrons.

The annihilation produced $e^+$ and $e^-$ are constrained to the region of production, either because of the presence of a magnetic field or because of scattering off microwave photons and any other radiation that may be present. The $e^+$ and $e^-$ cannot travel further than about 30 kpc from the region of production and are not useful in the detection of annihilations. The neutrinos, on the other hand, can travel to great distances, but their usefullness is severely limited by the difficulty encountered in their detection. This leaves $\gamma$-rays as the only probe presently available for any matter–antimatter annihilations that may be taking place. However, it has to be remembered that such annihilations are not the unique sources of $\gamma$-rays; these can also be produced by the decay of $\pi^0$ mesons produced in cosmic ray collisions and also by the inverse Compton scattering of starlight photons on cosmic rays. $\gamma$-rays observations can therefore best be used to set upper limits to annihilation rates.

A diffuse $\gamma$-ray spectrum has been reported from observations from the OSO–3 satellite. In this experiment Krushaar *et al.* (1972) observed an isotropic $\gamma$-ray background away from the galactic plane with an integral flux of $(3.0 \pm 0.9) \times 10^{-5}$ photons cm$^{-2}$ sec$^{-1}$ sr$^{-1}$ above 100 MeV. In addition to this there is a galactic component which is in good correlation with the distribution of hydrogen in the galaxy, and also a galactic centre component. Using observations from the SAS–II satellite, Fichtel *et al.* (1975) have reported an isotropic flux of $(19.3 \pm 0.26) \times 10^{-5}$ photons cm$^{-2}$ sec$^{-1}$ sr$^{-1}$ above 100 MeV. The integral flux reported here is somewhat less than the



OSO–3 result, but the two values agree within errors. The isotropic nature of this component of the $\gamma$-ray background makes it plausible that the radiation originated outside the Galaxy.

If it is assumed that the annihilations responsible for the radiation take place in an intergalactic gas, then the observations put an upper limit on the average annihilation rate $S$ per unit volume (Steigman, 1973).

$$S \lesssim 10^{-32}\,\mathrm{cm}^{-3}\,\mathrm{sec}^{-1}. \qquad (41)$$

In terms of the annihilation cross-section $\sigma$, the relative velocity $v$ of the annihilating particles, the mean squared intergalactic gas density $\langle n^2 \rangle$ and the antimatter fraction $f$, the annihilation rate per unit volume is

$$S = f\langle n^2 \rangle \sigma v. \qquad (42)$$

If it is assumed that the intergalactic space is filled with a cool neutral gas, then $\sigma v \gtrsim 10^{-10}\,\mathrm{cm}^3\,\mathrm{sec}^{-1}$ and $f\langle n^2 \rangle \leqslant 10^{-22}\,\mathrm{cm}^{-6}$. If the gas is assumed to be symmetric, $f = 1$ and it follows that $\langle n \rangle \lesssim \langle n^2 \rangle^{1/2} \lesssim 10^{-11}\,\mathrm{cm}^{-3}$, i.e., the gas must be very dilute. A similar limit to such a neutral intergalactic gas follows from the lack of Ly–$\alpha$ absorption in QSO spectra, if it is assumed that the QSO redshifts are of cosmological origin. The $\gamma$-ray limits therefore allow for such a gas, but it would represent only a small fraction of the matter content of the universe ($\lesssim 10^{-4}$ of the mass in the galaxies).

Observations of the X-ray background have suggested the presence of a hot intergalactic gas with temperature $T \cong 10^6\,°\mathrm{K}$ and mean squared density $\simeq 10^{-14}$. For such a gas the rate coefficient $\sigma v \cong 10^{-10}/T^{1/2}\,\mathrm{cm}^3\,\mathrm{sec}^{-1}$, and it follows that $f \lesssim 10^{-5}$. The $\gamma$-ray observations thus suggest that either the fraction of antimatter in the intergalactic medium is very small or there is no intergalactic medium. (For details and references, see Steigman, 1973).

It has been suggested by Gunn and Gott (1972) that within the Local Group there is a dilute hot gas with density $n \cong (1\text{–}3) \times 10^{-6}\,\mathrm{cm}^{-3}$ and temperature $T \cong 10^7\,°\mathrm{K}$. For such a gas Eqs. (41) and (42) give $f \lesssim 10^{-5}$. This means that neighbouring galaxies in our group are unlikely to contain antimatter. $\gamma$-ray observations of X-ray emitting clusters of galaxies, like the Perseus Cluster, show that the antimatter fraction for such clusters $\lesssim 10^{-5}$ (see Steigman, 1973). This indicates that if a significant quantity of antimatter is present in the universe, it must be separated from matter at least on the scale of clusters of galaxies.

Fichtel *et al.* (1977) have analyzed the SAS–II data to obtain the differential energy spectrum of the isotropic component of the diffuse gamma radiation above 35 MeV. They obtain for the differential spectrum

$$\frac{dJ}{dE} = 0.7 \times 10^{-7} \left[\frac{E}{100}\right]^{-3.4} \text{photons cm}^{-2}\,\text{sec}^{-1}\,\text{str}^{-1}\,\text{MeV}^{-1}, \qquad (43)$$

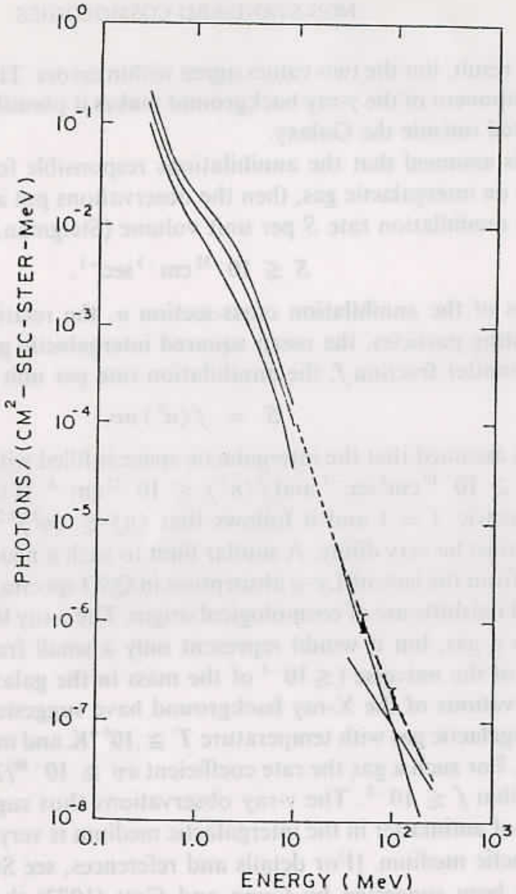



FIGURE 8.4 The energy spectra of diffuse gamma radiation reported by Trombka *et al.* (1977) at low energies and by Fichtel *et al.* (1975) at high energies, together with the deduced isotropic component and galactic component (steep solid line and flatter solid line respectively). Dashed lines are extensions of the SAS-2 results.

where $E$ is in MeV. This spectrum together with the data obtained by other workers down to 1 MeV and less is shown in Figure 8.4.

Stecker, Morgan and Bredekamp (1971) have calculated the $\gamma$-ray spectrum which is expected from the decay of $\pi^0 s$ produced in nucleon–antinucleon annihilations in Omnes' model. The spectrum peaks at 70 MeV if measured at the site of production. But since the annihilation and decay take place throughout the history of the universe, the spectrum has to be adquately redshifted to lower energies. To find the spectrum observed at the present epoch, it is necessary to integrate over all redshifts where $\gamma$-rays are produced. For $z \geqslant 100$, Compton interactions between $\gamma$-rays and the intergalactic gas may result in energy loss for the gamma rays. In general an integrodifferential equation involving $E_\gamma$ and $z$ has to be solved in order to obtain the expected spectrum. Absorption processes such as pair production mechanisms involving intergalactic gas and 2.7°K blackbody photons eliminate gamma rays with large redshifts from various parts of the observed spectrum. Gamma rays arising from pion decay processes at cosmological distances contribute significantly to the isotropic background only above 1 MeV, because $\gamma$-rays at lower energies have been redshifted by a factor $\geqslant 70$. Such a redshift corresponds to an epoch when the universe was opaque to $\gamma$-rays and the absorption effects were important.



In their original calculation Stecker, Morgan and Bredekamp (1971) used $n_0 = 10^{-5}\,\text{cm}^{-3}$ for the mean universal gas density and $H_0 = 100\,\text{km sec}^{-1}\,\text{Mpc}^{-1}$ for Hubble's constant. This corresponds to $\Omega = \varrho_0/\varrho_c = 1.0$, where $\varrho_0$ is the present matter density and $\varrho_c = 3H_0^2/8\pi G$ is the critical matter density required to close the universe. In a more recent calculation (see Stecker, 1978) the values $n_0 = 3 \times 10^{-7}\,\text{cm}^{-3}$, $H_0 = 50\,\text{km sec}^{-1}\,\text{Mpc}^{-1}$, which correspond to $\Omega = 0.1$, were used. The theoretical annihilation $\gamma$-ray spectrum is found to be compatible in both form and intensity to the observed spectrum above 1 MeV. The theoretical curve also predicts a hump apparent in the data around a few MeV. Other theories for the extragalactic component such as most galactic models, the primordial black hole hypothesis, conventional black body Compton radiation models etc. predict spectra which are significantly flatter than the observed spectrum (for references, see Fichtel *et al.*, 1977). Stecker *et al.*, interpret the success of their theoretical prediction of the $\gamma$-ray spectrum as evidence for the existence of antimatter in the universe and a pointer towards the correctness of Omnes' model. Steigman, (1976) has criticized this claim. He notes that the fit to the observed spectrum depends on the present density of the universe. The density determines the redshift at which the universe becomes opaque to ~100 MeV $\gamma$-rays and hence determines the energy at which the spectrum will turn over. He comments that by appropriately choosing the density and adjusting the annihilation rate as a function of epoch, any $\gamma$-ray spectrum can be produced.

## 9. AN ASSORTED COLLECTION OF NON-STANDARD COSMOLOGIES

In this chapter we mention briefly a number of interesting cosmological ideas suggested by different physicists and mathematicians from time to time. The discussion, out of space limitations, is necessarily brief. Our intention is to convey only a flavour of the different ideas in order to emphasize the scope of the field which opens out if we decide to enlarge our vision beyond the standard models. We begin with a discussion of the kinematic theory of E. A. Milne.

## 9.1 The kinematic relativity of E. A. Milne

In the cosmologies associated with general relativity and other gravitation theories, an attempt is made to describe the universe in terms of known physical theories. The aim of E. A. Milne's kinematic relativity proposed in 1935 was to deduce as much as possible about the structure of the universe merely from a cosmological principle and the basic properties of space and time and the propagation of light. In its dependence upon a cosmological principle, kinematic relativity is similar to Bondi and Gold's version of the steady state theory, but kinematic relativity covers not only cosmology but a great part of theoretical physics as well. The work of Milne and his coworkers is described in detail in two books by Milne (1935, 1948) and reviewed amongst others by Bondi (1960) and North (1965). Our sketch of the subject is heavily dependent on Bondi's very readable account.

In order to define and apply a cosmological principle, it is necessary to identify a set of fundamental observers for whom the principle will be valid. These observers are considered to be those located on "fundamental particles" which are taken to be the nuclei of galaxies. This last assumption is not free from ambiguities. Since Milne considered the fundamental observers to form a continuous set, not all fundamental observers can be identified with galactic nuclei. Milne has expressed the view that the system of fundamental particles plays the part of an imaginary homogeneous background against which the inhomogeneities and the random motions, which lead, for instance, to the formation of galaxies, have to be considered.

The chief concept used in kinematic relativity is the passage of time. It is assumed that each observer can causally order *local* evens, and label them with real numbers. The labelling process is arbitrary except that an earlier event is labelled with a lower number than a later event. Any system of labelling is called a clock, and the change from one system to another is called a regraduation of the clock.

Clocks are physically realized by using any repetitive device such as an atomic oscillator, a planetary orbit etc. Since it is impossible to compare intervals of time occurring at different epochs by laying them side by side, the concept of a uniform flow of time cannot be introduced *a priori*. It is therefore not obvious that different physical clocks, say an atomic and a dynamical clock, should have the same ratio of their periods at all times, and different time scales may exist.



Distance measurements are less fundamental in kinematic relativity than time measurements. Using light rays and the "radar" method, it is possible to measure spatial distances by using observations of time intervals between local events. An observer A sends out a light signal at an instant $t_1$ by his clock. This is reflected or replied to by B and arrives back at A at time $t_2$. Observer A then calls $(t_2 - t_1)/2$, multiplied by the conventional velocity of light, the distance of B from him at epoch $(t_1 + t_2)/2$. This measurement depends upon the clock A is using and a regraduation of A's clock necessarily implies a regraduation of all distance measurements. It is now standard to use the radar method in giving operational meaning to the notions of co-ordinates, distances, etc. (see Marzke and Wheeler, 1964; Anderson, 1967), but it was once criticized by Born (1943) as being far removed from all practical or indeed practicable measurements of distances.

A further measurement which is permitted in kinematic relativity is the measurement of angles, especially angles between light rays, by means of some instrument like the theodolite. This brings the number of space dimensions into the theory.

In the above framework, the cosmological principle adopted by Milne is that the totality of the observations that any fundamental observer can make by means of his clock and theodolite is identical with those any other fundamental observer can carry out.

Milne and Whitrow examined one-dimensional systems of observers to see whether the observers can regraduate their clocks in such a way that in some sense all of them keep the same time. It turns out that the problem has a well-defined solution if the relative motions of the observers satisfy certain conditions. Then there exists a universal or cosmic time for all the observers; the cosmic time is not unique, it admits regraduation.

Since the regraduation affects both time and distance measurements, the variation of the mutual distances of the observers depends on the particular type of clock used. Of the infinitely many time-scales available, the two most important for kinematic relativity are the $t$-time and the $\tau$-time. In $t$-time the relative motion of the observers is non-zero but unaccelerated. The zero of $t$-time may be chosen in such a way that the velocity distance relation $v = r/t$ is valid. The zero of $t$-time is a fundamental event at which the distance between all pairs of observers vanishes. It is the origin of the whole system. In $\tau$-time all the observers appear to be at rest. The transformation from $t$-time to $\tau$-time is

$$\tau = t_0 \log \frac{t}{t_0} + t_0, \qquad (1)$$

where $t_0$ is a constant. In $\tau$-time the origin of the whole system takes place in the infinite past.

Milne deduced that the transformation from one fundamental observer to another is the Lorentz transformation, when the $t$-time is used as the fundamental scale. This result is reasonable because the mutual accelerations vanish in the $t$-system. Since Lorentz transformations deal chiefly with electromagnetic phenomena, Milne has suggested that these follow $t$-time. Atomic clocks which are essentially electromagnetic in character indicate $t$-time, which may therefore be known as atomic time. The $\tau$-time is closely associated with dynamical processes.

When this argument is carried over to three dimensions, it turns out that additional assumptions are necessary to specify the curvature of three dimensional space. The choice is made that, in $t$-time, the transformation between any two fundamental observers should still be the Lorentz transformation. This implies that the four-dimensional space it flat, and this further implies that the motion of the observers is relatively unaccelerated and that their three-dimensional space is hyperbolic. The metric for the four-dimensional flat space-time may be variously written as

$$\begin{aligned} ds^2 &= d\bar{t}^2 - \bar{t}^2 \frac{d\bar{x}^2 + d\bar{y}^2 + d\bar{z}^2}{(1 - 1/4(\bar{x}^2 + \bar{y}^2 + \bar{z}^2))^2}, \\ &= dt^2 - dx^2 - dy^2 - dz^2, \\ &= e^{2\tau/t_0}\left[d\tau^2 - t_0^2 \frac{d\bar{x}^2 + d\bar{y}^2 + d\bar{z}^2}{[1 - (\bar{x}^2 + \bar{y}^2 + \bar{z}^2)^2/4]}\right], \end{aligned} \tag{2}$$

where the first form uses comoving co-ordinates and a universal time, the second locally defined measures of time and space in $t$-time, and the third locally defined measures of time and space in $\tau$-time. The transformation laws are

$$t = \bar{t}\,\frac{1 + (\bar{x}^2 + \bar{y}^2 + \bar{z}^2)^2/4}{1 - (\bar{x}^2 + \bar{y}^2 + \bar{z}^2)^2/4}, \quad x = \frac{\bar{x}\bar{t}}{1 - (\bar{x}^2 + \bar{y}^2 + \bar{z}^2)^2/4},$$

$$\tau = t_0 \log \frac{\bar{t}}{t_0}. \tag{3}$$

Since the co-ordinates $\bar{x}$, $\bar{y}$, $\bar{z}$ are constant for any observer, it follows that $x/t$ is constant. Hence the velocity distance relation takes the simple form

$$u = \frac{dx}{dt} = \frac{x}{t}, \tag{4}$$

i.e., in vector notation,

$$\bar{v} = \frac{\bar{r}}{t}. \tag{5}$$



It follows from the theory of the Lorentz transformation that the quantity

$$X = t^2 - x^2 - y^2 - z^2 \tag{6}$$

is an invariant. The invariant density $\varrho$, which is the number of particles in the invariant volume element $dxdydzdt/ds$, is a function only of $X$, and

$$n = \varrho(x)\,\frac{dt}{ds}, \tag{7}$$

where $n$ is the apparent density, i.e., the number of particles in the volume element $dxdydz$. From Eqs. (2) and (4) it follows that $(ds/dt)^2 = X/t^2$, and therefore that

$$n = t(X)X^{-1/2}. \tag{8}$$

Milne assumed that matter is conserved. It then follows from the equation of continuity for $n$ and Eqs. (5) and (8) that

$$\varrho = BX^{-3/2}, \quad n = BtX^{-2}, \tag{9}$$

where $B$ is a constant. The meaning of Eqs. (5) and (9) is that the velocity distance relation applies strictly, the density decreases in time but increases from the origin towards $X = 0$, and that the surface $X = 0$ represents the invariant border of the universe which is advancing at the speed of light.

The knowledge about the behaviour of the density completes the discussion of the substratum in kinematic relativity, as the question of pressure does not arise since no random motions have been introduced. The particularly simple structure of the substratum in kinematic relativity is used together with the cosmological principle to construct a dynamics and other branches of theoretical physics.

In kinematic relativity a free particle is defined as one which is not participating in the motion of the substratum. For such a particle it can be shown that acceleration

$$\frac{d^2\bar{r}}{dt^2} = \bar{f}(\bar{r}, \bar{v}, t) = (\bar{r} - \bar{v}t)\,\frac{Y}{X}\,G(X, \xi), \tag{10}$$

where

$$X = t^2 - \bar{r}^2, \quad Y = 1 - \bar{v}^2, \quad Z = t - \bar{v}\cdot\bar{r}, \quad \xi = \frac{Z^2}{XY}. \tag{11}$$

At this point it was assumed by Milne that since (10) involves a property of the substratum which was itself built up without reference to any dimensional fundamental constants, no such constants can appear in the theory. The non-dimensional constant $G$ cannot therefore depend on a dimensional argument like $X$, and hence $G = G(\xi)$. The vectorial part of (10) determines the direction of the acceleration $\bar{f}$, and this is along the line joining the free particle to that fundamental particle which has the same velocity as the free particle. Since $X$ and $Y$ are positive, the acceleration is towards the fundamental particle if $G < 0$ and away from it if $G > 0$. In the first case the system is stable in the sense that there is tendency for the velocities of free particles to diminish relatively to the local fundamental particles. In the case $G > 0$ the system is correspondingly unstable. It has been shown by Milne that $G \leqslant -1$.

Consider a swarm of free particles of different positions and velocities superposed on the substratum. It can be shown that the distribution function $f(x, y, z; \dot{x}, \dot{y}, \dot{z}; t)$ of these particles is of the form

$$f = Y^{-5/2} X^{-3/2} \psi(\xi), \tag{12}$$

where the dimensionless distribution function $\psi(\xi)$ is related to the acceleration function $G(\xi)$ through

$$G(\xi) = -1 - \frac{C}{(\xi - 1)^{3/2} \psi(\xi)}, \tag{13}$$

with $C$ a constant.

Milne interpreted (13) as a law of gravitation, since it determines the accelerations in terms of the distribution of the massive particles. This is not a general law of gravitation and it applies only to mass distributions satisfying the cosmological principle, i.e., those which appear the same to every fundamental observer.

Milne identified the first term in Eq. (13) with the gravitational effects of the substratum alone, and the second term with the gravitational effects of the statistical distribution of masses. In the absence of free particles, therefore, $G = -1$ and (10) takes the form

$$\frac{d\bar{v}}{dt} = -\frac{Y}{X}(\bar{r} - \bar{v}t). \tag{14}$$

In $\tau$-time this appears for small $\bar{v}$ as

$$\frac{d^2\pi}{d\tau^2} = 0, \tag{15}$$

where the distance $\pi$ in $\tau$-time for nearby objects [$\bar{r}/t$ small] is

$$\pi = \frac{t_0}{t}\bar{r}. \tag{16}$$



Hence for nearby, slowly moving objects, in the absence of other free particles, Newton's first law of motion applies in $\tau$-time. Milne therefore identified $\tau$-time with the time of Newtonian dynamics. Milne and Walker have shown that with appropriate interpretations Eq. (13) can lead to an inverse square law of gravitation. The constant $C$ turns out to give the relation between particle numbers and their effective gravitational masses. Milne and Walker have built up a complete system of dynamics. Most of the results of this scheme are in agreement with Newtonian theory, but there are a few exceptions which though small on the local scale, are of great importance for cosmic problems. A very significant deviation is that angular momentum turns out to be a quantity proportional to the cosmic time $t$. This result greatly aids Milne's theory of galactic structure. Kinematic relativity also leads to new theories of photon and electromagnetic fields and even supplies a new basis for atomic and nuclear theory.

## 9.2 Einstein–Cartan cosmologies

In the Einstein–Cartan gravitation theory, the influence of the intrinsic spin of matter on space-time geometry is considered. The theory was originated by Cartan (1922), independently formulated by Sciama (1962, 1964), Kibble (1961) and Peres (1962) and developed by Hehl and coworkers (Hehl, 1973, 1974; Hehl, von der Heyde and Kerlick (1976) and Trautman (1972).

A justification for the consideration of spin as a source of the gravitational field may be made as follows (Hehl, 1973, 1974). In the classification of elementary particles using the irreducible representations of the Poincaré group, particles are labelled by their mass $m$ and their intrinsic spin $s$. Mass is connected to the translational part of the Poincaré group and spin to the rotational part. In a field theoretic formalism mass corresponds to the energy–momentum tensor and spin to the spin angular momentum tensor. In general relativity, the gravitational field is the dynamical manifestation of energy–momentum to which it is dynamically coupled through the energy–momentum tensor. This is appropriate in macroscopic situations when the spin usually averages out to zero. In microscopic situations, as well as in those macroscopic situations where there is alignment of spin, however, there is no *a priori* reason for relegating to spin a role inferior to that assigned to mass. In such situations the dynamical effects of spin on the geometry of space-time should be considered. This is done in the Einstein–Cartan–Kibble–Sciama ($U_4$) theory by making the affine connection asymmetric and relating its antisymmetric part to the spin angular momentum tensor.

The space-time of the $U_4$ theory is a four-dimensional non-Riemannian manifold on which an asymmetric affine connection $\Gamma^k{}_{ij}$ and a symmetric metric $g_{ij}$ are defined. These are compatible with each other, i.e., the covariant derivative of the metric tensor, defined using the affine connection $\Gamma^k{}_{ij}$, vanishes and the scalar product of any two vectors is preserved under parallel transport. The affine connection is given by

$$\Gamma_{ij}{}^k = \left\{{}_{ij}^{\,k}\right\} + Q_{ij}{}^k - Q_j{}^k{}_i + Q^k{}_{ij}, \tag{17}$$

where $\left\{{}_{ij}^{\,k}\right\}$ is the Christofel symbol of the second kind and $Q^k{}_{ij}$ is the torsion tensor:

$$\left\{{}_{ij}^{\,k}\right\} = \frac{g^{km}}{2}(g_{mi,j} + g_{mj,i} - g_{ij,m}),\ Q_{ij}{}^k = \tfrac{1}{2}(\Gamma^k{}_{ij} - \Gamma^k{}_{ji}). \tag{18}$$

The Riemann and Ricci tensors are defined in the usual manner using the affine connection $\Gamma^k{}_{ij}$, and the scalar curvature $R = g^{ik}R_{ik}$ as in a Riemann manifold.

The field equations for the metric and torsion are

$$R_{ij} - \tfrac{1}{2}g_{ij}R = -\kappa t_{ij} \tag{19}$$

and

$$Q^i{}_{jk} - \delta^i{}_j Q^l{}_{lk} - \delta^i{}_k Q^l{}_{jl} = \kappa S^i{}_{jk}, \tag{20}$$

where $\kappa$ is Einstein's constant of gravitation $8\pi G/c^4$, $t_{ij}$ is the asymmetric canonical energy-momentum tensor and $S^i{}_{jk}$ the spin angular momentum tensor. If $L$ is the Lagrangian scalar density, these tensors are defined as

$$\sqrt{(-g)}S_\lambda{}^{\mu i} = \frac{\delta L}{\delta K_{ij}{}^\lambda},\quad K_{ij}{}^\lambda = -Q_{ij}{}^\lambda + Q_j{}^\lambda{}_i - Q^\lambda{}_{ji}, \tag{21}$$

and

$$t^{ij} = T^{ij} + \nabla_k{}^*(S^{ijk} - S^{jki} + S^{kij}), \tag{22}$$

where $T^{ij}$ is the metric energy-momentum tensor defined by

$$\sqrt{(-g)}T^{ij} = 2\frac{\delta L}{\delta g_{ij}} \tag{23}$$

and $\nabla_k^*$ is the modified covariant derivative

$$\nabla_k{}^* = \nabla_k + 2Q_{kl}{}^l, \tag{24}$$

with $\nabla_k$ the usual covariant derivative defined in terms of $\Gamma^k{}_{ij}$.

The field equation (20) is an algebraic rather than a differential equation relating torsion to spin. The consequence is that in the $U_4$ theory, there can be no torsion of space-time outside spinning matter. Torsion is inextricably bound to matter and cannot propagate through the vacuum as a torsion wave or via any interaction of non-vanishing range. Outside the matter distribution spin makes itself felt only through its influence on the metric tensor.



Because the equation for torsion is algebraic, it is possible everywhere to substitute spin for torsion and effectively eliminate torsion from the formalism. For this purpose the Einstein tensor on the left of Eq. (19) is split into its Riemann part, which is computed using the Christofel symbol $\left\{{}_{ij}^{\,k}\right\}$, and its non-Riemann part, the torsion terms for which are substituted for using (20). The result is the single field equation

$$(R_{ij} - \tfrac{1}{2}g_{ij}R)^{\{\}} = -\kappa\tilde{T}_{ij}, \tag{25}$$

where the symmetric energy-momentum tensor is defined by

$$\tilde{T}^{ij} = T^{ij} + \kappa[-4t^{ik}{}_{[l}t^{jl}{}_{k]} - 2t^{ikl}t^j{}_{kl} + t^{kli}t_{kl}{}^j + \tfrac{1}{2}g^{ij}(4t_{m[l}{}^k t^{ml}{}_{k]} + t^{mkl}t_{mkl})], \tag{26}$$

and obeys the conservation law $\nabla_j^{\{\}}\tilde{T}^{ij} = 0$.

It is possible to estimate the densities at which spin effects are as important as mass effects by considering a continuum made of elementary particles with number density $n$. For particle mass $m$ and spin $\hbar/2$, the mass and spin densities are $\varrho = mn$ and $\sigma = n\hbar/2$ respectively. From Eq. (26) it follows that mass density receives corrections from the spin density which are of the order of $\kappa\sigma^2$, and spin effects are as important as mass effects when the density has the critical value

$$\bar{\varrho} \simeq \frac{m^2}{k\hbar^2}. \tag{27}$$

$\bar{\varrho}$ is of the order of $10^{47}$ gms cm$^{-3}$ for electrons and $10^{54}$ gms cm$^{-3}$ for neutrons. This shows that it is necessary to consider spin effects only in connection with the very high densities which occur in the early epochs of cosmological models or in gravitational collapse, and in the study of quantum gravitational processes.

An interesting feature of the Einstein–Cartan theory is that it admits non-singular cosmological solutions which are physically reasonable in the sense that they can reproduce the observed properties of the universe as well as general relativistic models do. The work of Penrose, Hawking and Geroch (Hawking and Ellis, 1973) has shown that space-time singularities, at least in the sense of geodesic incompleteness, are a general feature of Einstein's theory. In this circumstance a theory which admits non-singular solutions, in situations which are always described as singular in general relativity, is especially significant.

It was first conjectured by Trautman that the singularities of gravitational collapse and cosmology might be prevented by the direct influence of spin on the geometry of space-time. Kopczynski (1972) has obtained a two parameter family of non-singular cosmological solutions for a spherically symmetric distribution of spinning dust. He has also obtained an anisotropic solution

(Kopczynski, 1973) with shear in which there is no singularity when the effect of torsion is greater than that of shear. A special case of this solution in which the spins are aligned along an axis and the shear vanishes has been considered by Trautman (1973).

In Trautman's model, the universe is filled with a spinning dust characterized by a mass density $\varrho$ and four velocity $u^i$. A classical description of spin is assumed, so that the spin angular momentum tensor can be expressed in terms of an antisymmetric tensor $S_{ij}$:

$$S^i{}_{jk} = u^i S_{jk}, \quad S_{jk} u^k = 0. \tag{28}$$

The metric is of the Robertson–Walker form

$$ds^2 = c^2 dt^2 - R^2(t)(dx^2 + dy^2 + dz^2), \tag{29}$$

which is comoving, i.e., $u^i = \delta_0{}^i$, with $x^0 = ct$, $x^1 = x$, $x^2 = y$, $x^3 = z$. It is assumed that the spins are aligned along the $x$-axis, and the only independent non-vanishing component of $S_{ij}$ is $S_{23} = \sigma$. The field equations are reduced to a modified Friedmann form

$$\frac{\dot{R}^2}{2} - \frac{GM}{R} + \frac{3G^2 S^2}{2c^4 R^4} = 0, \tag{30}$$

and the conservation laws for mass and spin are

$$M = \tfrac{4}{3}\pi\varrho R^3 = \text{constant}, \qquad S = \tfrac{4}{3}\pi\sigma R^3 = \text{constant}. \tag{31}$$

The last term on the left hand side of (30) is repulsive. It dominates at small $R$ and prevents $R$ from reducing to zero. Equation (30) can be solved exactly and gives

$$R(t) = \left[\frac{3GS^2}{2Mc^4} + \frac{9GMt^2}{2}\right]^{1/3}. \tag{32}$$

$R$ is minimum at $t = 0$. For a sphere containing $N$ particles with mass $m$ and spin $\hbar/2$, the minimum value of $R$ is

$$R_{\min} = \left(\frac{3GS^2}{2Mc^4}\right)^{1/3} = \left(\frac{3NG\hbar^2}{8mc^4}\right)^{1/3}. \tag{33}$$

If $m$ is the nucleon mass, $R_{\min} \simeq 3 \times 10^{-27} N^{1/3}$ cms. For $N = 10^{80}$, which is representative of the total number of nucleons in the universe, $R_{\min} \simeq 1$ cm. This is a very small size, but it is much larger than the Planck length $(G\hbar/c^3)^{1/2} \simeq 1.6 \times 10^{-33}$ cm, which is considered as providing the only natural limitation on the validity of classical, Einsteinian cosmology. The density of matter at $t = 0$ is $\varrho \simeq m^2 c^4/G\hbar^2 \simeq 10^{55}$ gms cm$^{-3}$. This density is much



smaller than the density $c^5/G^2\hbar \simeq 10^{94}$ gms cm$^{-3}$ at which the quantum effects of the gravitational field are presumed to play a dominant part. Nevertheless this density is so large that the properties of matter near $t = 0$ are entirely unknown, and the merits of considering the state to be formally non-singular may be questioned. For $t \gg \tau = \hbar/mc^2 \simeq 10^{-23}$ secs the model coincides with the Einstein–de Sitter solution of general relativity. Therefore any considerations about the hadronic era of the universe, and its subsequent development are little affected by the introduction of torsion in cosmology. (It is interesting to note that the C-field cosmology of Section 4 yields exactly the same solution (16) when there is no creation of matter.)

It is assumed in this model that spins are aligned along the $x$-axis. Trautman (1973) has suggested that a cosmic magnetic field $B$ may provide a correlating mechanism for the spins. Such a magnetic field may successfully compete with increasing temperature $T$, which would tend to destroy the alignment, provided that the flux is conserved and $BR^2$ = constant. As $RT$ = constant, the ratio $\mu B/kT$ behaves like $1/R$ and may have been so large in the past that the magnetic field aligned all spins in spite of the large temperature.

Isham, Salam and Strathdee (1973) have considered the result on Trautman's model of the interposing $f$ gravity, which is the strong short range component of their two tensor ($f$–$g$) theory of gravity (Hehl, 1973, 1974). Isham *et al.* find that when $f$-gravity is taken into account the minimum radius of Trautman's model is $\simeq 10^{13}$ cms, rather than 1 cm and the density of nucleonic matter at minimum radius is $10^{17}$ gms cm$^{-3}$ rather than $10^{55}$ gms cm$^{-3}$. The matter thus collapses to densities only a few orders of magnitude higher than known nuclear densities. The arguments used on arriving at this result are heuristic in nature and are not based on some exact cosmological solution of the field equations.

When the shear in Kopczynski's anisotropic solution does not vanish, the metric is of the form

$$ds^2 = dt^2 - X^2(t)dx^2 - Y^2(t)(dy^2 + dz^2). \tag{34}$$

The cosmological fluid is considered to be pressure and rotation free. The density of matter $\varrho$, density of spin $\sigma$ and shear scalar $\psi$ are all $\propto R^{-3}(t)$, where $R(t) = (\sqrt{-g})^3 = (XY^2)^{1/2}$ is the average length scale factor (see Section 2.12). As in Eq. (31) constnats $M$, $S$ and $\Sigma$ can be defined with

$$\varrho = \frac{3M}{4\pi R^3}, \quad \sigma = \frac{3S}{4\pi R^3}, \quad \psi = \frac{3\Sigma}{4\pi R^3}. \tag{35}$$

The first integral for $R(t)$ analogous to (30) is

$$\frac{\dot{R}^2}{2} - \frac{GM}{R} + \frac{3G^2 S^2}{2c^4 R^4}(S^2 - \Sigma^2) = 0, \tag{36}$$

which may be integrated to give

$$R(t) = \left[\frac{3G(S^2 - \Sigma^2)}{2Mc^4} + \frac{9GMt^2}{2}\right]^{1/3}. \tag{37}$$

The shear term is always attractive and has the same $R$ dependence as the repulsive spin term. A singularity can only be prevented when $S > \Sigma$.

Stewart and Hajicek (1973) have independently obtained this solution and argued that the singularity is avoided in this model only because of its high symmetry. At large values of $R$, we get from Eq. (37)

$$H^2 = \frac{\dot{R}^2}{R^2} \simeq \frac{2GM}{R^3} = \frac{8\pi G\varrho}{3}, \tag{38}$$

where $H$ is Hubble's constant, and we have from Eqs. (35)

$$\psi^2 = \lambda\sigma^2, \tag{39}$$

where $\lambda$ is a constant. From Eqs. (35), (38), (39),

$$\frac{\psi^2}{H^2} \simeq \frac{3\lambda\sigma^2}{8\pi G\varrho} = \frac{3\lambda}{8\pi G}\left(\frac{\sigma^2}{\varrho^2}\right)\varrho = \frac{3\lambda}{8\pi G}\left(\frac{S^2}{M^2}\right)\varrho, \tag{40}$$

and for a fluid containing $N$ neutrons each with mass $m$ and spin $\hbar/2$,

$$\left(\frac{\psi^2}{H^2}\right) \simeq \frac{3\lambda\hbar^2\varrho}{32\pi Gm^2}. \tag{41}$$

Since the singularity is avoided when $\Sigma < S$, we have $\lambda < 1$, and

$$\left(\frac{\psi^2}{H^2}\right)_{\text{now}} \lesssim \frac{3\hbar^2\varrho_{\text{now}}}{32\pi Gm^2} \simeq 10^{-27}\left(\frac{gm}{cm}\right)^2, \tag{42}$$

for $\varrho_{\text{now}} = 10^{-29}$ gms cm$^{-3}$. This may be expressed in a dimensionless form as

$$\left(\frac{\psi^2}{H^2}\right)_{\text{now}}\left(\frac{G}{c^2}\right)^2 \lesssim 10^{-83}. \tag{43}$$

Comparison of this value with the observational upper limit $\simeq 10^{-6}$ shows that unless the universe is very nearly isotropic a singularity cannot be avoided.

Non-singular cosmological solutions with spins aligned along an axis of symmetry have also been obtained by Kuchowicz (1976a). Non-singular Bianchi type-I cosmological models in which a magnetic field is present have been obtained by Kuchowicz (1976b) and Raychaudhari (1975) for various equations of state. Tafel (1973) has shown that non-singular cosmological models of Bianchi types I–VIII are possible, and he has obtained a class of solutions of Bianchi type V.



In the Einstein–Cartan theory, spinless free massive particles move along timelike geodesics and photons along null geodesics. Geodesics therefore have the same fundamental significance as in general relativity, and geodesic incompleteness of a $U_4$ manifold should be considered an indication that it is singular. The singularity theorems (Hawking and Ellis, 1973) set down a set of conditions under which geodesic incompleteness occurs. At least one of these conditions must be violated in the non-singular cosmological models of the $U_4$ theory. One necessary condition common to the singularity theorems is the strong energy condition

$$R_{ik}\xi^i\xi^k \leqslant 0 \tag{44}$$

for every non-spacelike vector $\xi^i$ (the inequality sign in (21) can be reversed by adopting a convention different from ours in defining various basic tensors). Hehl, von der Heyde and Kerlick (1974) have shown that the strong energy condition is violated under the assumptions which have been made in obtaining the known non-singular models. In all such models the metric energy-momentum tensor is

$$T^{ij} = (\varrho c^2 + p)u^i u^j - pg^{ij}, \tag{45}$$

and the spin angular momentum tensor obeys relations of the form given in equations (28). The symmetric energy-momentum tensor for such a fluid is

$$\begin{aligned}\tilde{T}^{ij} = {} & (\varrho c^2 + p - \tfrac{1}{2}\kappa c^2 s^2)u^i u^j - (p - \tfrac{1}{4}kc^2 s^2)g^{ij} \\ & - 2c(u_k u^l + \delta^l{}_k)\nabla_l^{\{\}}(S^{k(i}u^{j)}),\end{aligned} \tag{46}$$

where $\nabla_l^{\{\}}$ is the covariant derivative corresponding to the Christofel symbol, and $s^2 = 2S_{ik}S^{ik}$.

For the mixed field equation (25) the strong energy condition reduces to

$$(\tilde{T}^{ij} - \tfrac{1}{2}g^{ij}\tilde{T}_k{}^k)\xi_i\xi_j \geqslant 0, \tag{47}$$

which in the special case $\xi_i = u_i$ gives upon using (46)

$$\tfrac{1}{2}\varrho c^2 + \tfrac{3}{2}p - \tfrac{1}{2}\kappa c^2 s^2 + 2cS^{ij}\Delta^{\{\}}{}_{[i}u_{j]} \geqslant 0. \tag{48}$$

For the known cosmological models the last term on the left is zero, and the strong energy condition reduces to

$$\tfrac{1}{2}\varrho c^2 + \tfrac{3}{2}p - \tfrac{1}{2}\kappa c^2 s^2 \geqslant 0. \tag{49}$$

The quantities $\varrho$, $p$ and $s^2$ are all dependent on the time (i.e., on the average length scale factor), and as soon as the spin density $s^2$ reaches a value such that

$$\varrho c^2 + 3p < \kappa c^2 s^2, \tag{50}$$

the strong energy condition is violated, and the possibility of avoiding geodesic incompleteness arises. But geodesic completeness of itself is not sufficient to ensure that there is no space-time singularity in a $U_4$ manifold. In the $U_4$ theory particles with spin do not move along geodesics; their equations of motion are of the Papapetrou type (Hehl, undated). It is therefore necessary to ensure $b$-completeness (Hawking and Ellis, 1973) if it is to be concluded that the space-time is non-singular.

## 9.3 The chronometric cosmology

Recently I. Segal (1976a) has proposed a new kinematic approach to cosmology. The main feature of this cosmology is the observational prediction that the distance is quadratically proportional to redshift. Segal claims that such a relation fits the observed redshift–magnitude diagram better than the linear Hubble law of standard cosmology. Segal calls this theory chronometric cosmology since it involves "time" in two different aspects, local and global. The basic features of this theory are summarized below. The observational validation is deferred to Section 9.

*The fundamental assumptions* Globally the cosmos is taken to be a four dimensional manifold $\tilde{M}$, causality preserving, homogeneous and isotropic. Specifically, we take $\tilde{M}$ as $R' \times S$ where $R'$ is the real line corresponding to a time co-ordinate $t$ and $S$ is the 3-dimensional spherical space (i.e., the surface of a 4-sphere). Locally, however, the space-time is Minkowskian with a time co-ordinate $x_0$ which *is not the same as* $t$. The local measurements measure $x_0$, not $t$. The differences between $x_0$ and $t$ can be apparent only from a cosmological observation connecting the local domain to a distant one, e.g., the measurement of the redshift of a distant galaxy. The theory is based on considerations of symmetry and on the group theoretical properties of Maxwell's equations. The role of Riemannian geometry is only incidental.

The cosmos admits stationary observers. The assumption of stationarity, Segal considers essential since without the notion of stationary states, changes of state cannot be validly described. The physical observers are subjected to restrictions similar to that in the Minkowski space-time. For example any given timelike direction at a point is tangential to the forward direction of some admissible observer. Also, two different observers at the same point see the cosmos in causally compatible ways. That is, causality is preserved when transformation is made between the maps of the cosmos made by the two observers.

The main problem next arises of "connecting" the local Minkowski space-time $M$ with the global space-time $\tilde{M}$ so that non-local observations can be locally interpreted. Although Segal considers a correspondence between the



$(n + 1)$ dimensional Minkowski space and an $(n + 1)$ dimensional spherical space we will describe here the relevant particular case of $n = 3$.

*The $M$–$\tilde{M}$ correspondence* Set up co-ordinates $x_0, x_1, x_2, x_3$ in the Minkowski space $M$ with the usual metric $\eta_{ik}$ (cf. Section 2). Define the following quantities:

$$\left.\begin{aligned} X &\equiv (x_0, x_1, x_2, x_3) \\ X^2 &= x_0^2 - x_1^2 - x_2^2 - x_3^2 \\ \xi_{-1} &= 1 - \tfrac{1}{4}X^2,\ \xi_4 = 1 + \tfrac{1}{4}X^2,\ \xi_1 = x_i, \end{aligned}\right\} \tag{51}$$

where $i = 0, 1, 2, 3$. Let $\Xi = (\xi_{-1}, \xi_0, \xi_1, \xi_2, \xi_3, \xi_4)$, with

$$\Xi^2 = \xi_{-1}^2 + \xi_0^2 - \xi_1^2 - \xi_2^2 - \xi_3^2 - \xi_4^2 \tag{52}$$

be defined for arbitrary $\xi_j$ and suppose that the transformation

$$\Lambda : X \to \Xi \tag{53}$$

is defined such that if $\Xi = \Lambda(X)$ for $X$ in $M$ then $\Xi^2 = 0$. We can look upon $\Xi$ as a point in a general 6 dimensional vector space $W$ in which $\Xi^2 = 0$ defines a quadric $Q$. Let $\tilde{M}$ be the 5-dimensional projective space in rays in $W$, obtained by the identification

$$(\xi_{-1}, \xi_0, \xi_1, \xi_2, \xi_3, \xi_4) \equiv (\alpha\xi_{-1}, \alpha\xi_0, \alpha\xi_1, \alpha\xi_2, \alpha\xi_3, \alpha\xi_4) \text{ for all } \alpha \to 0.$$

For the ray $\Xi$ through $\Xi$, let

$$\tilde{\lambda} : X \to \tilde{\Xi} \tag{54}$$

define a map of $M$ into the quadric $\tilde{M}$ in $\tilde{M}$ corresponding to $Q$ in $W$. The mapping is one to one. It is possible to recover $(x_0, x_1, x_2, x_3)$ from a general point of $\tilde{Q}$ in the following way.

For a non-zero $\Xi$ in $W$ let $u$ denote the vector $(u_{-1}, u_0, u_1, u_2, u_3, u_4)$ with

$$u_j = \xi_j(\xi_{-1}^2 + \xi_0^2)^{-1/2}. \tag{55}$$

Set $\mathbf{u} = u(\Xi)$ and note that $u$ is the same for all vectors on the same ray as $\Xi$. Thus $\tilde{u}(\Xi) = u(\Xi)$ is unique and $\Xi \to \tilde{u}$ is a well-defined mapping from $Q$ into a 5-dimensional space $R^5$. Also, by definition

$$\mathbf{u}_1^2 + u_0^2 = 1 = u_1^2 + u_2^2 + u_3^2 + u_4^2. \tag{56}$$

Thus this mapping, to be denoted by

$$\Gamma : \tilde{M} \to R^4 \tag{57}$$

maps $\tilde{Q}$ into the direct product of a circle $S^1$ co-ordinatized by $(\mathbf{u}_1, u_0)$ and a 4-sphere $S^3$ co-ordinated by $(u_1, u_2, u_3, u_4)$. Conversely any point of

$S^1 \times S^3$ corresponds to a point of $\tilde{M}$ via the mapping $\gamma : \mathbf{u} \to \Xi$, where $\Xi = (u_{-1}, \ldots u_4)$. This mapping is, however, two to one since both $\pm\mathbf{u}$ correspond to the same point of $\tilde{M}$. We say that $\gamma$ is a two-fold covering of $\tilde{M}$ by $S^1 \times S^3$.

Now, given $(u_{-1}, \ldots u_4)$ for any one of the two points of $S^1 \times S^3$, corresponding to a given $X$ in $M$, we have from Eq. (51)

$$u_{-1} = k^{-1}(1 - \tfrac{1}{4}X^2), \quad u_4 = k^{-1}(1 + \tfrac{1}{4}X^2)$$
$$u_j = k^{-1}x_j, \quad k = \pm[(1 - \tfrac{1}{4}X^2)^2 + x_0^2]^{1/2}. \tag{58}$$

The mapping $X \to \mathbf{u}$ is one-to-two but is locally one-to-one. Taking $k > 0$ in the vicinity of $X = 0$, we have uniquely,

$$x_j = 2u_j(u_{-1} + u_4)^{-1} = 2\xi_j(\xi_{-1} + \xi_4)^{-1}. \tag{59}$$

Here we have $\mathbf{u} = (1, 0, 0, 0, 0, 1)$ at $X = 0$. We may now define a new time $t$ by

$$t = \tan^{-1}\frac{u_0}{u_{-1}}. \tag{60}$$

We now have a metric of the following form on $S^1 \times S^3$:

$$d\tilde{s}^2 = dt^2 - ds^2 \tag{61}$$

where $ds^2$ is the line element on $S^3$. The apparent Lorentzian structure on $S^1 \times S^3$ is not invariant under the action of the Lorentz group on $M$; however it is invariant under conformal transformation. Also, as $(u_{-1}, u_0)$ moves on the unit circle, $t$ goes from $-\infty$ to $\infty$ because of the multivaluedness of the transformation (60). This is the universal space $R^1 \times S^3$ representation of Segal's chronometric cosmos $\tilde{M}$. We will now discuss its observable properties.

*The physical implications* The space $S^1 \times S^3$ is not causal in the large, since the substitution

$$u_{-1} = \cos t \quad u_0 = \sin t \quad u_i = c_i(\text{constant}) \tag{62}$$

describes a cyclic curve in forward timelike direction. Is there any observable effect of this peculiar timelike behaviour? Segal points out that the local observer will make his measurements with the Minkowskian co-ordinates $x_i$ in $\tilde{M}$. He cannot measure $t$ in a local experiment. However, a non-local experiment can reveal the effect of the non-linear relationship between the chronometric cosmos $\tilde{M}$ and the local Minkowski space-time.

Consider for example, the redshift of distant galaxies. this arises in this cosmology because frequency is measured on the $t$-scale rather than on the



$u_0$-scale. That is, the operator for frequency is $-i\partial/\partial t$ rather than $-i\partial/\partial x_0$. If $\psi$ is the wavefunction of light of frequency $\nu$ at the point $P_0$ of emission taken as the origin, then initially we should have

$$-i\frac{\partial\psi}{\partial x_0} = \nu\psi = H_0(0)\psi \tag{63}$$

where $H_0(0)$ is the energy operator at the initial stage. (We have taken $\hbar = 1$.) Initially, the $t$-time is also zero. At a later time $t = S > 0$, the energy operator will be $H_0(S)$ at the point $P$ where the frequency is measured. $P$ has co-ordinates $(S, u)$, say, where $u$ is the position in $S^3$. Segal shows that $H_0(S)$ is related to $H_0(0)$ by the unitary transformation

$$H_0(S) = e^{-iSH}H_0(0)e^{iSH}, \tag{64}$$

where $H$ is the "true energy operator". In the chronometric cosmos $H$ stands for $-i\partial/\partial t$. It can be expressed in terms of the Minkowski co-ordinates by using the relation

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial x_0} + \frac{1}{4}(x_0^2 + x_1^2 + x_2^2 + x_3^2)\frac{\partial}{\partial x_0} + \frac{x_0}{2}\left(x_1\frac{\partial}{\partial x_1} + x_2\frac{\partial}{\partial x_2} + x_3\frac{\partial}{\partial x_3}\right) \tag{65}$$

which follows from the transformations given earlier. A lengthy computation then gives

$$H_0(S) = \frac{1 + \cos S}{2i}\frac{\partial}{\partial x_0} + \frac{1 - \cos S}{2i}Q\frac{\partial}{\partial x_0}Q + \frac{\sin S}{2i}K, \tag{66}$$

where

$$Q\frac{\partial}{\partial x_0}Q \equiv -\tfrac{1}{4}(x_0^2 + x_1^2 + x_2^2 + x_3^2)\frac{\partial}{\partial x_0} + \frac{x_0}{2}\left(x_1\frac{\partial}{\partial x_1} + x_2\frac{\partial}{\partial x_2} + x_3\frac{\partial}{\partial x_3}\right),$$
$$K \equiv -\left(x_0\frac{\partial}{\partial x_0} + x_1\frac{\partial}{\partial x_1} + x_2\frac{\partial}{\partial x_2} + x_3\frac{\partial}{\partial x_3}\right). \tag{67}$$

Thus at $P$ we would not expect $H_0/S$ to have the same frequency eigenvalue that $H_0(0)$ had. Segal computes the new eigenvalue and finds that it is given by

$$\nu^1 = \frac{1 + \cos S}{2}\nu, \tag{68}$$

so that at $P$ a redshift $z$ is observed, given by

$$z = \frac{\nu - \nu^1}{\nu^1} = \tan^2 \frac{S}{2}. \tag{69}$$

A rigorous derivation of this formula was given by Segal in a later paper (Segal, 1976b). The quantum dispersion round this frequency shift is too small to be observed.

Since $S$ in (8.19) represents the change in the $t$ co-ordinate, we now need to connect it with the local Lorentz frame at the observer. This Lorentz frame should be *tangential* to the cosmic space-time in the manner described below.

Take, at the point of emission $P_0$, $R(P_0)$ to represent the radius of the space $S^3$. With $P_0$ as origin of the local Minkowski space, and $R(P) = R_0$ (say), the Minkowski co-ordinates are defined in terms of global co-ordinates $u_j$ by

$$x_j^0 = 2u_j R_0 (u_{-1} + u_4)^{-1}. \tag{70}$$

We can similarly set up co-ordinates at the point of observation $P$. Without loss of generality we may take the problem as a two-dimensional one in which the light from $P_0$ leaves in the local $x_1^0$ direction, and arrives at $P$ in the local $x_1$ direction. We now use the parameters $t$ and $\varrho$, for the typical point on $S^1 \times S^3$

$$u_{-1} = \cos t, \quad u_0 = \sin t, \quad u_1 = \sin \varrho, \quad u_2 = \cos \varrho. \tag{71}$$

The tangential Minkowski co-ordinates near the point of emission are

$$x_0^0 = \frac{2R_0 \sin t}{\cos t + \cos \varrho}, \quad x_1^0 = \frac{2R_0 \sin \varrho}{\cos t + \cos \varrho}. \tag{72}$$

Near point $P_1$ which being light-like w.r.t. $P$ must have $t = \varrho = \alpha$ (say), the co-ordinates are

$$x_0 = \frac{2R \sin (t - \alpha)}{\cos (t - \alpha) + \cos (\varrho - \alpha)}, \quad x_1 = \frac{2R_1 \sin (\varrho - \alpha)}{\cos (t - \alpha) + \cos (\varrho - \alpha)}. \tag{73}$$

The transformation $(x_0^0, x_1^0) \rightarrow (x_0, x_1)$ can be interpreted as the product of a scale transformation by the factor $(R_1/R_0) \sec \alpha$ and a Lorentz transformation with a velocity depending on $\alpha$. However, the velocity is only virtual since the chronometric cosmos is stationary.

In the spherical geometry of $S^3$, the distance varies as $\sin \varrho$ and hence the expected flux from a source with luminosity spectral function $f(\nu)$ in the observed frequency range $(\nu_1, \nu_2)$ is

$$l = \frac{1}{\sin^2 \varrho} \cdot \frac{1}{1 + z} \int_{\nu_1(1+z)}^{\nu_2(1+z)} f(\nu)\, d\nu. \tag{74}$$



The derivation of (8.24) is similar to that in standard cosmology. If in the relevant range,

$$f(\nu) \alpha \nu^{-\alpha} \tag{75}$$

then we have in terms of magnitudes

$$m = 2.5 \log z - 2.5(2 - \alpha) \log (1 + z) + \text{constant}, \tag{76}$$

where we have used the relation

$$z = \tan^2 \frac{\varrho}{2} \tag{77}$$

to eliminate $\varrho$. (8.26) is the redshift magnitude relation in Segal's cosmology.

We shall discuss the observational implications of Segal's cosmology in the next section. We mention in passing what Segal has often emphasized: that his velocity–distance relation is *not* linear but quadratic. Thus at large redshifts the curve ($\log z$ against $m$) should rise *faster* than the standard Hubble curve.

## 9.4 McCrea's cosmological uncertainty

In a series of articles W. H. McCrea (1960, 1960–1) has questioned some of the foundations of cosmology as a predictive science. His reasoning is as follows.

Normally in physics we predict the future behaviour of a system on the basis of physical laws and the information available about the system *now*. However, in cosmology, the vastness of the distance scale and the limitations of the finite speed of light impose some restrictions on this procedure. For example, when we look at a remote galaxy $P$ we see it in a state in which it could possibly affect us now. However, we do not know of the state of the parts of the universe that affected $P$ at the time and in the state we are seeing it now. In a static Euclidean universe, suppose $P$ is $10^9$ light years away. Then we see it as it was $10^9$ years ago. Now, it would be affected by other points $Q$, $R$, . . . on *its* past light cone. Thus a galaxy $Q$ located $10^9$ light years away on the opposite side of us with respect to $P$, would be on this light cone $3 \times 10^9$ years in the past—i.e., $2 \times 10^9$ years before the epoch in which it is observed now. In a non-static universe, however, we cannot know (or claim to know) from the state in which we see $Q$ now, about the state in which it was when it affected $P$ at the epoch when $P$ itself could affect us now.

Although standard cosmologies get round this difficulty by *postulation* mathematical models for the universe, this approach, according to McCrea, hides the underlying basic limitation on any cosmological theory making predictions on the basis of observations on the past light cone. Many of these predictions (cf. Section 10) involve factors $(1 + z)$ where $z$ is the redshift of a remote galaxy. Many different cosmological theories differ in their predictions in powers of $(1 + z)$. According to McCrea, it is futile, in view of the above limitation, to try to distinguish between these theories with such tests. Just as quantum uncertainty makes it meaningless to try to distinguish between different paths of an electron in a scattering process, so McCrea's cosmological uncertainty makes it meaningless to talk of a difference between say, the Hubble diagrams of the Einstein–de Sitter cosmology and the steady state cosmology.

Davidson (1960) has criticized this point of view by arguing that once a cosmological model is formulated it makes definite predictions. Thus it should be possible to compare the present and the remote parts of the universe and decide, for example whether it is in steady state. McCrea, on the other hand, believes the proposition undecidable by light cone tests in view of the cosmological uncertainty.

McCrea has not carried this concept any further (in a quantitative manner) and because of this the controversy seems to be unresolved. The concept of limitation of finite light speed is nevertheless interesting and deserves further consideration.

## 9.5 The anthropic principle

The anthropic principle states (Carter, 1974) that what we can expect to observe in the universe must be restricted by the conditions necessary for our presence as observers. It arose as a reaction to the Copernican principle being carried too far in cosmology. We have seen in Section 2 that the cosmological principle makes it impossible to talk of a privileged position in the universe at any given cosmic epoch. In Section 4 we saw that the perfect cosmological principle makes even all epochs alike. It has been argued by Dicke (1961) and Carter (1974) that it would be wrong to believe that our situation cannot be privileged in any sense if one accepts (a) that specially favourable conditions of temperature, chemical environment, etc., are prerequisite for our existence and (b) that the universe evolves and is by no means spatially homogeneous on a local scale.

One aim of the anthropic principle is to explain the various large number coincidences, mentioned in Section 6, without invoking new theories or hypotheses, like Dirac's Large Numbers hypothesis. But it can also be driven further to derive relationships between the fundamental constants which would restrict their range of values, and perhaps even determine them with accuracy. The observed coincidences and relations are of three kinds (Carter, 1974): (1) those which can be understood in terms of conventional physics, without invoking the anthropic principle, (2) those which require the use of a "weak" form of the principle, and (3) those which require an extended



"strong" form of the anthropic principle. We will now consider one example of each kind.

(1) Stellar masses: The mass of the Sun can be elegantly expressed by the relation

$$M \sim \alpha_G^{-3/2} m_p, \tag{78}$$

where $m_p$ is the proton mass and

$$\alpha_G \equiv \frac{G m_p^2}{\hbar c} \simeq 5 \times 10^{-39}$$

is the gravitational coupling constant. This was first noticed by Jordan (1974), who believed on the basis of it that star formation was a cosmological rather than purely astrophysical process, and constructed an elaborate theory to explain the coincidence. But the relation follows from a straight forward application of star formation theory (Carr and Rees, 1979). The virial theorem can be used to show that the maximum temperature reached in the gravitational collapse of a gas cloud is

$$kT_{\max} \sim \left(\frac{N}{N_0}\right)^{4/3} m_e c^2, \tag{79}$$

where $m_e$ is the electron mass, $N$ the number of protons in the cloud and $N_0 \equiv \alpha_G^{-3/2}$. The collapsing cloud becomes a star only if $T_{\max}$ is high enough for nuclear reactions to occur, that is if $kT_{\max} > q m_e c^2$, where $q$ depends on the strong and electromagnetic coupling constants and is $\sim 10^{-2}$. It then follows from Eq. (79) that $N \gtrsim 0.1 N_0$, i.e.

$$M \gtrsim 0.1 \alpha_G^{-3/2} m_p.$$

An upper limit to the mass of a star is obtained from the requirement that it should not be radaition pressure dominated, for such a star would be unstable to pulsations which would probably lead to its disruption. Calculations show that the upper limit is $\sim 10\, N_0 m_p$, so all main sequence stars are in the range

$$0.1 \alpha_G^{-3/2} m_p \lesssim M \lesssim 10 \alpha_G^{-3/2} m_p,$$

which explains relation (78). A detailed review on the scales of various other structures is nature has been given by Carr and Rees (1979).

(2) Coincidences between large numbers: We saw in section 6 that there are several dimensionless numbers in physics and cosmology with magnitude $\sim 10^{40}$. Dirac's Large Numbers hypothesis was invoked to explain the approximate equality between these seemingly unrelated numbers. One such

relationship is

$$\frac{ct_0}{a_0} \sim \frac{\alpha}{\alpha_G}, \tag{80}$$

where $t_0$ is the present age of the universe, $a_0$ is the Bohr radius of the atom $\simeq 10^{-8}$ cm and $\alpha \equiv e^2/\hbar c = 1/137$ is the fine structure constant. (It must be noted that this relation is equivalent to equation (2) of Section 6. We are using a slightly different form here to confirm with popular usage in the present context).

An anthropic explanation for Eq. (80) has been suggested by Dicke (1961). Assuming that life requires elements heavier than hydrogen and helium, Dicke argued that any cognisable universe would be one in which some stars have already completed their main sequence evolution. This is because the heavier elements can only be produced and spread throughout the universe in supernova explosions. The main sequence lifetime of a star whose opacity is dominated by electron scattering, which happens in the large mass stars necessary for generating supernovae is given by

$$t_{ms} \sim \left\{\eta\alpha^2\left(\frac{m_p}{m_e}\right)^2\right\}\alpha_G^{-1}\left(\frac{M}{M_c}\right)^{-2} t_p,$$

where $\eta \lesssim 10^{-2}$ is the efficiency of stellar nuclear burning, $M_c$ is the Chandrasekhar mass $\simeq 1\,M_\odot$, and $t_p \equiv \hbar/m_pc^2 \simeq 10^{-23}$ s. The quantity is braces is of the order of unity, and M is of the order of $M_c$ (see above). Living observers can therefore exist only when

$$t_0 > t_{ms} \sim \alpha_G^{-1} t_p. \tag{81}$$

However $t_0$ cannot be much greater than $t_{ms}$ since most of the universe would otherwise have been processed into white dwarfs, newtron stars or black holes. Observers are therefore most likely to exist at $t_0 \sim t_{ms} \sim \alpha_G^{-1} t_p$, so that the relation (80) is satisfied (to within a few orders of magnitude).

Carr and Rees (1979) have provided an independent argument, based on cosmology and independent of stellar evolution theory, in support of the relation (81). These authors, and others, have also demonstrated how the anthropic principle could be used to explain other large numbers coincidences.

Dicke's argument is an example of the "weaker" form of the anthropic principle, in which it is assumed that we must be prepared to take into account the fact that our location in the universe is necessarily priviledged, to the extent of being compatible with our existence as observers. In using this form of the principle, no attempt is made to explain the actual values of the fundamental constants. All that is done is that some otherwise unexplained features of the universe are understood as features for our very existence, given various laws and constants.



(3) Relation between the gravitational and electromagnetic fine structure constants: Carter (1974) has obtained a connection between $\alpha$ and $\alpha_G$ using the anthropic principle and the existence of convective stars. The mass $M_*$ which divides convective (red dwarf) stars from radiative (blue giant) stars may be shown to be (see Carr and Rees (1979))

$$M_* \sim M_c\alpha_G^{-1/2}\alpha^6\left(\frac{m_e}{m_p}\right)^2 \sim M_c\alpha_G^{-1/2}\alpha^{10}$$

using the relation $m_e/m_p \sim 10\alpha^2$, which results from a coincidence in nuclear physics. The mass $M_*$ lies in the range around the Chandrasekher mass $M_c$ in which main sequence stars exist only because

$$\alpha_G \sim \alpha^{20}. \tag{82}$$

Were $G$ (and hence $\alpha_G$) slightly larger, all stars would be radiative; if it were slightly smaller, all stars would be convective. It is possible to ascribe anthropic significance to (20) if it is assumed that planets form only around convective stars. For then no planets, and hence no life, would form if $\alpha_G$ were much larger than $\alpha^{20}$. If $\alpha_G$ were much smaller than $\alpha^{20}$, all stars would be chemically homogenised due to convection and probably would not form supernovae, thus making it impossible for the heavy elements, which are again necessary for our existence, to be produced. The weakest link in this argument is the supposition that planets form only around convective stars. This must at the moment be considered speculative, but it is supported by the observation that red dwarfs have much less angular momentum than blue giants, and the loss of angular momentum may be a consequence of planet formation.

The anthropic principle has thus provided a relationship between $\alpha$ and $\alpha_G$, and explained why $\alpha_G$ is so much smaller. Carter's argument is an example of an application of the "strong" form of the anthropic principle, which states that the universe, including the fundamental parameters on which it depends, must be such as to admit the creation of observers within it at some stage. This is a stronger assumption than the one Dicke had to make in his explanation of the large numbers' coincidence. There it was merely assumed that observers like ourselves can exist only after the universe has evolved sufficiently, but not too late in its history, i.e. at some privileged epochs.

There are coincidences in nuclear physics as well to which anthropic significance may be ascribed. For instance, Hoyle (1965) has pointed out that life as we know it may turn out to be sensitively dependent on certain energy levels of carbon. In stellar evolution the element carbon is synthesized by the reaction sequence

$$He^4 + He^4 \rightarrow Be^8, \qquad Be^8 + He^4 \rightarrow C^{12}.$$

However, $Be^8$ is unstable and normally $C^{12}$ would not have formed. But there is a resonant level of excited state $C^{*12}$ of Carbon so that the reaction

$$3He^4 \rightarrow C^{*12}$$

proceeds. If this excited state were not possible the universe would have been almost devoid of carbon which is so necessary for life as we know it. There is however no similar favourably placed resonance in $O^{16}$. Otherwise almost all the carbon would be transmuted into oxygen.

An attempt has been made to provide a physical foundation to the anthropic principle by associating it with the "many worlds" interpretation of quantum mechanics (Everett 1957). According to this, at each observation, the universe branches into a number of parallel universes, of which only one is known to any well defined observer. Wheeler (1971) has in accordance with this scheme envisaged an infinite ensemble of universes, all with different coupling constants. Most of these universes are completely uninteresting; only those which start off with the right coupling constants become "aware of themselves" (Carr and Rees, 1979), by producing observers who make the universe cognisable. It is possible that all such cognisable universes have some features in common with our universe.

The anthropic principle has not yet been used to predict any new feature of the universe. It leads to an estimate of the order of magnitude of most of the fundamental constants of physics, but it has not yet been used to derive any exact values. Moreover, it is based on a vary parochial concept of an observer, which assumes that all intelligent living creatures have the same nature as we ourselves do. Nevertheless apart from Dirac's Large Numbers hypothesis (where applicable), it is the only framework we have for explaining many of the remarkable coincidences that are found in nature. And Dirac's ideas are now in some doubt because of the severe restrictions recently placed on the rate of change of the gravitational constant.

## 10. THE VERY EARLY UNIVERSE

In Section 2.10 we described the background of this topic, and the *raison d'etre* for studying it. Until the particle physicists agree upon what is the form of a unified theory of all interactions, the topic of the very early universe will continue to belong to non-standard cosmology. We summarize the highlights of this exciting field *as they are seen at the time of writing*.



### 10.1 The problems of the very early universe

What are the problems of cosmology that necessitate the study of the universe much prior to the epochs discussed by Gamow *et al.*? The outstanding problems are the following.

(i) *The photon–baryon ratio* The microwave background has a Planckian spectrum which enables us to calculate the number density of photons $n_\gamma$ to be

$$n_\gamma = n_{\gamma 0} \equiv \frac{2.404}{\pi^2}\left(\frac{kT_0}{c\hbar}\right)^3, \tag{1}$$

where $T_0$ is background temperature at present.

Suppose that the universe at present has a total matter density of $\Omega_0 \varrho_c$ where $\varrho_c$ is given by Eq. (85) of Section 2. Let a fraction $f_B$ of this be made of baryons. Taking the typical baryonic mass $m$ as the mass of a proton we get the total baryon number density at present as

$$n_{B_0} = \frac{3H_0^2}{8\pi Gm}\Omega_0 f_B. \tag{2}$$

Notice that both $n_\gamma$ and $n_B$ at any epoch in the past (except the very early onces which we shall discuss separately) depend on the scale factor $S$ as $S^{-3}$. Therefore

$$\frac{n_\gamma}{n_B} = \frac{n_{\gamma 0}}{n_{B_0}} = \frac{6.4}{\pi}\frac{Gmk^3T_0^3}{c^3\hbar^3}(H_0^2\Omega_0 f_B)^{-1}. \tag{3}$$

Given the range of values of $H_0$, $\Omega_0$, $T_0$ and $f_B$ based on present extragalactic astronomy, we find that the ratio $n_\gamma/n_B$ lies in the range $\sim 10^8$–$10^{10}$.

Why does this ratio have this particular value? That this value was determined at a very early epoch is clear from the fact that only then was it possible to alter the values of $n_\gamma$ and $n_B$ substantially through (possibly) baryon nonconserving reactions.

(ii) *The horizon problem* The microwave background radiation is extremely isotropic both on large and small angular scales (except for the dipole component which is due to the Earth's motion against the background). The large scale isotropy indicates that the radiation has the same energy density over large volumes of space, of linear dimensions $\sim 10^{28}$ cm. However, when the radiation was made, the particle horizon of the universe was very small, too small to guarantee this degree of homogeneity. For, a horizon necessarily restricts the range of communication and hence the ability of physical processes to homogenize the universe. For example, at the recombination epoch of redshift $\sim 10^3$, the scale factor was smaller than its present value by the factor $\sim 10^3$ and the present size of $\sim 10^{28}$ cm corresponded to $\sim 10^{25}$ cm. However, using the relations (89) and (90) of Section 2.10 we find that the universe at this epoch was only $\sim 3 \times 10^5$ yrs old and had a particle horizon size of $\sim 10^{23}$ cm, i.e., a hundred times smaller than the above value.

If we insist that the isotropy of the universe was established even earlier, then the discrepancy becomes even larger: the horizons are far too small to account for the observed isotropy. How then could the process of homogenization take place?

(iii) *Flatness* This problem was first mentioned by Dicke and Peebles (1979) although it was highlighted by Guth (1981) later. It can be seen from the Eqs. (72) and (73) of Section 2.10. When we talk of the decleration parameter $q_0$ as defined by eqn (81) of Section 2.10, its value in the currently estimated range of say 0.01 to 5 (cf. Section 11.3) implies that *at present* the various terms on the left-hand side of (72) and (73) are comparable to within an order of magnitude. However, the terms $\ddot{S}/S$, $\dot{S}^2/S^2$ and $k/S^2$ do not maintain this comparability at all epochs. In particular, the curvature term declines in relative importance compared to the other terms. Closer to $S = 0$, $k/S^2 \ll \dot{S}^2/S^2$, which is why we set $k = 0$ in Eq. (89), Section 2.10.

Justified through this appears when judged from the *present* stand-point, the reasoning contains a fallacy. For, the state of the universe was decided not at the present epoch but at some early epoch close to $t = 0$. It is at that epoch, say at $t = \tau$, that one would expect the curvature term $k/S^2$ to be comparable to the other two. If that were the case, the characteristic time scale of the universe would have turned out to be $\sim\tau$; that is the universe would either have collapsed to a singularity ($k = +1$) or dispersed to infinity ($k = -1$) in times comparable to $\sim\tau$. Dicke and Peebles took $\tau \sim 1$ sec corresponding to the epoch of nucleosynthesis. If one is talking of the very early universe as Guth did $\tau$ would have to be even smaller.

The only way out of this difficulty is to assume that at $t = \tau$, the universe was very finely tuned close to the state $k = 0$. For only in this "flat" model do we find that the time scale $\tau$ does not enter the functional form $S \propto t^{2/3}$ of the scale factor. Why should the universe have selected the flat mode of expansion—or a mode which was extremely close to the flat mode?

## 10.2 GUTs and cosmology

By writing $kT$ in units of MeV or GeV it is convenient to identify the epochs in the very early universe that are relevant to high energy physics. The relation (90) of Section 2 is thus generalized to

$$t_{\text{second}} = 2.4 g^{-1/2} T_{\text{MeV}}^{-2} = 2.4 \times 10^{-6} g^{-1/3} T_{\text{GeV}}^{-2}, \qquad (4)$$



where

$$g = g_b + \tfrac{7}{8} g_f. \qquad (5)$$

Here $g_b$ = number of spin states of relativistic bosons and $g_f$ = number of spin states of relativistic fermions, it being assumed that the bosons and fermions are in thermodynamic equilibrium. (For details see Narlikar, 1983).

In the mid 1970s it was apparent (cf. Steigman, 1976) that the photon to baryon ratio cannot be explained in terms of a simple model wherein baryons and antibaryons form a primordial mixture with photons, all being in thermodynamic equilibrium. Thus initially, creation/annihilation processes

$$B + \bar{B} \leftrightarrow \gamma + \gamma \qquad (6)$$

go on against a back-drop of the expanding universe until the baryons become nonrelativistic and then drop out of equilibrium. The number $n_\gamma/n_B$ is then frozen from that epoch. Steigman (*op. cit.*) found that the value is too high:

$$\frac{n_\gamma}{n_B} \cong 5 \times 10^{17}. \qquad (7)$$

Moreover, the same ratio applies to antibaryons also, thus leading to zero baryon number.

This result is not surprising since nowhere in this assumption is the particle–antiparticle asymmetry built in. Indeed to understand the observed value of $n_\gamma/n_B$ three ingredients are necessary to be satisfied by the primordial scenario: (i) the baryon number is not conserved (ii) there is a basic asymmetry in the way baryons and antibaryons react (iii) thermodynamic equilibrium is not maintained prior to the freezing of the ratio $n_\gamma/n_B$.

Yoshimura (1978), Weinberg (1979) and others have given scenarios that lead to $n_\gamma/n_B$ in the range $10^4$–$10^{12}$ (which includes the observed range). The large uncertainty of the answer depends on the uncertain parameters of the grand unified theories chosen for the description of particle physics.

Grand unified theories (GUTs) seek to unify the strong interaction, the weak interaction and the electromagnetic interaction. The simplest of GUTs, SU(5) framework (Georgi and Glashow, 1974), includes three colours of quarks and two leptons (in three flavours) and requires 24 vector bosons. Of these the photon ($\gamma$), the W-bosons ($W^\pm$, $W^0$) and the eight gluons were already required by the simple justaposition of the three interactions in a $SU(3) \times SU(2)_L \times U(1)$ framework. (The subscript L stands for left handed leptons.) The other 12 bosons, designated as X bosons can convert leptons to quarks and vice versa and thereby violate the baryon number conservation law.

However, the so called minimal SU(5) framework predicts the decay of the proton with a lifetime of $10^{31}$–$10^{32}$ years. The experiments on proton decay set lower bounds on proton lifetimes that are $\gtrsim 2 \times 10^{31}$ years (Krishnaswamy *et al.*; 1982) for decay of protons in iron and $\gtrsim 2 \times 10^{32}$ years (Park *et al.*, 1985) for decays in water. Physicists consider even more elaborate GUT models including supersymmetric ones. In supersymmetry the distinction between bosons and fermions is removed. Bosons like photons, gravitons, mesons have supersymmetric fermion partners photinos, gravitinos, mesinos, etc. While fermions like electrons, neutrinos etc. have bosonic partners selectrons, sneutrinos etc (for a review of supersymmetric theories see, for example, Gates *et al.* (1983)). Another approach is based on the original Kaluza-Klein theory (Kaluza, 1921; Klein, 1926) which was proposed to unify electromagnetism with gravity in a 4 + 1 dimensional spacetime. In modern K–K theories there are 7 extra dimensions which are compactified to represent internal degrees of freedom of particles. (see Van Nieuwenhuizen, 1981, for a review).

It is expected in all these exotic theories that their plethora of exotic particles had a transient but definitive role to play in the early universe. Although the particles themselves may not be observable today, their early interactions should have left relics which serve as observational checks. One such awkward relic is the magnetic monopole. The monopole should not exist in isolation according to Maxwell's theory. It can exist, as a compact structure in GUTs, with typical mass energy of $\sim 10^{16}$ GeV. Being stable indestructible structures, the monopoles should survive from the early epochs when they were formed. Estimates of their present number density, however, show that their total mass density today should be $\gtrsim 10^{15}$ times the closure density. Clearly, such monopoles should not be allowed to exist!

## 10.3 The inflationary universe

A new input into the early universe scenario was made by Guth (1981) who explored the consequences of the phase transition as the GUT group (SU(5) or any other) breaks into SU(3) × SU(2) × U(1) that is, when the three component interactions separate. Guth pointed out that the change is not achieved instantly, but over an extended period and in this process the temperature of the universe drops in a manner that may be compared to the supercooling of steam and its eventual conversion to water. When steam is supercooled, it retains its gaseous structure even below the existing boiling point. However, this state is not quite stable and in the course of time the steam condenses to droplets of water releasing its latent heat.

In the very early universe the GUT-phase transition produces a shift in the state of the vacuum. What was originally the vacuum (i.e., state of lowest



energy) at high temperature is no longer the true vacuum at low temperature. The change from false to true vacuum releases energy; the false vacuum also generates negative stresses. As first calculated by Callan and Coleman (1980) the extended phase transition would occur in bubble of space, like the droplets of water. However, because of the negative stresses generated by the false vacuum these bubbles would expand until the false vacuum has completely changed over to true vacuum of lower energy.

Recall that in the context of the steady state universe described in Section 4.6.1, McCrea (1951) had explained the expansion of the universe as due to negative stresses in the cosmic fluid. Guth's inflationary universe is nothing but the same idea: the negative stresses arising from the false vacuum blow up the bubble with the exponential scale factor

$$S = \exp Ht, \tag{8}$$

where the negative pressure is given by

$$p = -\frac{3H^2}{8\pi G}. \tag{9}$$

(Refer to Eq. (29) of Section 4.6.1.) This exponential expansion was called "inflation" by Guth. For a general review of the inflationary models see Linde (1984).

It is customary to interpret this effect as a $\lambda$-term. Using (98) from Section 2.11, we get

$$\lambda = 3H^2 = 8\pi G\varrho \tag{10}$$

where $\varrho$ is the energy density due to vacuum. The $\lambda$-term here is in principle calculable from GUTs and is expected to be quite large compared to the cosmical constant $\lambda$ discussed in Section 2.11. However, the solution here has formal similarity with the de Sitter and the steady state models.

The inflation, according to Guth's model did not last longer than $\sim 10^{-32}$ sec. The value of $\lambda$ was such as to increase the scale factor by as much as $\sim 10^{50}$. When the expansion was complete, the energy released by the change over of vacuum states was expected to reheat the universe to $\sim 10^{27}$ K. (In the steam analogy the condensation of supercooled steam releases latent heat.) The energy released was supposed to be thermalized by collisions of different bubbles. Thereafter the bubble expanded as the standard radiation universe with $S \propto t^{1/2}$. What we consider as the observable universe is a subset of this bubble. It is this feature that helps the inflationary model to get round the horizon and flatness problems.

The exponential function (8) has the property that it does not produce a particle horizon. During the inflationary phase therefore it is possible to argue that the ease of communication over large distances would tend to establish homogeneity. It could also be shown that the curvature term $k/S^2$ became negligible during the exponential expansion so that the flatness problem does not exist.

The similarity of the inflationary model with other ideas proposed for the steady state model is obvious. The concept of the bubble universe discussed in Section 4.6.3 finds an echo here. So does the work described in Section 4.5.2. For, it is argued that the exponential inflation more or less eliminates the fine details of whatever initial structure that existed before inflation. In modern terminology this result is called the cosmic baldness hypothesis (cf. Barrow, 1984, and references therein).

The original Guth version of inflation, however, ran into difficulties largely because the phase transition required quantum tunneling that led to large inhomogeneities in the universe — inhomogeneities too large to be consistent with the cosmological principle. A new way of phase transition was proposed by Linde (1982) and independently by Albrecht and Steinhardt (1981). These authors studied the consequences of phase transition as discussed by Coleman and Weinberg (1973). Here the change over from false to true vacuum did not require quantum tunneling and the resulting phase transition was slow and smooth.

The associated inflationary model called the *new* inflationary model ran into fresh difficulties, however. The scenarios of galaxy formation and the small scale fluctuations of microwave background radiation based on the new inflationary model lead to temperature fluctuations at least $\sim 10^4$ times those quoted as limits of the present observations.

The fact that $\lambda$ (inflation) exceeds Einstein's $\lambda$ term by a factor $\sim 10^{108}$ is also a cause for concern. For, as seen by observational tests at present (cf. Section 11.3) the Einstein $\lambda$-term is quite small if at all non-zero. How was it possible for the universe to switch over to a small value of $\lambda$ after the inflationary epoch? Clearly a fine tuning of $\lambda$ seems to be unavoidable.

Such difficulties have led workers in the area to more esoteric particle theories like supersymmetry and Kaluza–Klein theories. There are scenaries which discuss inflation in these frameworks. (See for example, Linde, 1982 and Van Nieuwenhuizen, 1981.) Since supersymmetry introduces more adjustible parameters, the fluctuation problem is resolved by supersymmetric inflation. However, it is still too early to assess the merits of these latest versions of the inflationary universe.

Another feature which is still unsolved is the nature and extent of dark matter in the universe. The rotation curves of galaxies and the observed velocity dispersion in clusters of galaxies suggest the presence of unseen matter considerably in excess of the visible component. The unseen matter,



if it exists, must have played a crucial role in the formation of galaxies, their distribution in clusters and superclusters etc. A major research industry has grown round this problem in which the dark matter is supposed to be made of any of these combinations: low mass stars, black holes, massive neutrinos, gravitinos, photinos, axions and other exotic particles depending on the research workers' prejudices. So far no directly observable prediction has emerged. (See Section 11 for the discussion of dark matter.)

## 10.4 Quantum cosmology

The standard big bang picture gets modified if we recognize that the Einstein equations of classical gravity break down when the characteristic radius of curvature of spacetime becomes comparable to or less than the Planck length

$$L_p = \left(\frac{G\hbar}{c^3}\right)^{1/2}. \tag{11}$$

In the very early universe this happened prior to the cosmic time

$$t_p = \left(\frac{G\hbar}{c^5}\right)^{1/2} \cong 5 \times 10^{-43}\,\text{sec}. \tag{12}$$

In energy units this corresponds to $\sim 10^{19}$ GeV. A correct picture of the universe beyond these limits will emerge from quantum gravity.

There is no universally accepted formally complete and directly applicable theory of quantum gravity yet to answer the question whether the standard model needs to be substantially modified because of quantum inputs. There are many approaches by different people (DeWitt, 1967a, b, c; Gibbons and Hawking, 1977; Arnowitt *et al.*, 1962; Penrose, 1975; to name a few) and there is no common agreement between these approaches even on the fundamental question as to whether the universe had originated in the big bang.

A somewhat simplified approach that addresses the questions of spacetime singularity, horizon and flatness has come from Narlikar and Padmanabhan (1983). In this approach only the conformal degree of freedom is quantized. Thus, given the classical spacetime with singularities as a manifold $\bar{m}$ with metric $\bar{g}_{ik}$, the conformal quantization considers metrics of the kind

$$g_{ik} = \Omega^2 \bar{g}_{ik}, \tag{13}$$

where $\Omega$ is a function of spacetime coordinates. $\Omega$ can be quantized by the Feynman path integral method and it leads to the conclusion that amongst such spacetimes those with singularity and horizon have probability measure zero. In other words the standard big bang singularity is an exception rather than the rule.

The conformal fluctuations of empty flat Minkowski spacetime can generate nontrivial world models (Brout *et al.*, 1980; Padmanabhan, 1983a). In particular, Narlikar and Padmanabhan (*op. cit.*) find that of the various conformally flat models, the $k = 0$ Robertson–Walker model is overwhelmingly preferred by probabilistic considerations. Thus the so called flatness problem is also resolved.

Padmanabhan (1983b) has also shown that a semi classical approach to cosmology can be developed in which the back ground metric $\bar{g}_{ik}$ is not a solution of Einstein's equations but satisfies a modified set of equations. The modifying terms arise from taking expectation values of the quantum operators $\Omega^2$, $\Omega_i\Omega^i$, $\Omega\,\Box\,\Omega$, etc. It is interesting that the modified equations resemble those of the HN Cosmology discussed in Section 7.

In general, conformal fluctuations produce effective contributions to the energy tensor which resemble the negative energy or negative stress contributions of energy tensors in the steady state theory (cf. Section 4).

While in this approach the conformal degrees of freedom are treated in most general fashion and this is the only degree of freedom relevant to the Robertson–Walker line element, the question as to what will happen if non-conformal degrees of freedom are included remains unsettled.

## 11. THE OBSERVATIONAL TESTS

The test of the pudding lies in its eating. For all its conceptual elegance a scientific theory is judged in the last analysis by its ability to measure up to the experimental/observational tests. In this final chapter we will discuss the observational situation *vis-a-vis* the non-standard models presented in the preceding chapters. Some tests peculiar to specific models have been discussed already along with those models and to avoid repetition we will not mention those. Rather, we shall be concerned here with those tests which can be applied to a wide range of cosmological models. These fall into tests of several types, outlined below in brief.

(A) *Local tests of weak gravity* These mainly refer to the experimental tests in the Solar system where the gravitational fields are weak. Since non-standard cosmologies employ some gravitation theory or other, these tests are also indirect tests of these cosmologies.

(B) *Local tests involving astrophysics* The state of the universe in our local neighbourhood (i.e., up to say the local cluster of galaxies which includes our



own) can tell us whether it is consistent with the large scale structure of the universe as predicted by the cosmological model.

(C) *Tests of the large scale structure of the universe* These depend on the extragalactic observations through optical and radio telescopes with possible inputs from space astronomy.

(D) *Tests of fundamental constants* Some non-standard cosmologies directly or indirectly imply the variation of some of the so-called fundamental constants. We shall be concerned here mainly with the variation of the gravitational constant $G$.

In these discussions we refer from time to time to the predictions of standard cosmology to enable the reader to compare the performance of various theories with the standard one. The type of test (in the above four categories) will be indicated by an appropriate letter in brackets following it.

### 11.1 The weak field tests (A)

There are three classical tests of general relativity which have been discussed extensively since the early days of the theory:

(i) the gravitational redshift
(ii) the perihelion precession for planet Mercury
(iii) the bending of light by the Sun's gravitational field.
Recent improvements in technology have added three more tests to this list:
(iv) the equality of inertial and gravitational mass
(v) the radar echo delay
(vi) the precession of gyroscopes.

These tests with the exception of (vi) use the spherically symmetric solution produced by a gravitating mass $M$. In the Newtonian theory the solution is described by a potential

$$\phi = \frac{GM}{R} \tag{1}$$

at a distance $R$ from the centre of the mass. In general relativity the spacetime metric exterior to $M$ is given by the Schwarzschild solution. In isotropic co-ordinates this line element takes the form

$$ds^2 = \left(\frac{1 - GM/2c^2r}{1 + GM/2c^2r}\right)^2 dT^2 - \left(1 + \frac{GM}{2c^2r}\right)^4 [dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)]. \tag{2}$$

In other gravitation theories using the space-time metric in the same way as general relativity, the solution in the case of weak gravitational fields is usually expressible in the form (Weinberg, 1972) (with $\eta = GM/c^2r$):

$$ds^2 = (1 - 2\alpha\eta + 2\beta\eta^2 + \ldots)dT^2 - (1 + 2\gamma\eta + \ldots)[dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)] \quad (3)$$

where $\alpha, \beta, \gamma$ are constants. A comparison of (2) and (3) gives $\alpha = 1, \beta = 1, \gamma = 1$ for general relativity. For the Brans–Dicke theory we have (Brans and Dicke, 1961)

$$\alpha = 1, \quad \beta = 1, \quad \gamma = \frac{\omega + 1}{\omega + 2} \quad (4)$$

where $\omega$ is the parameter discussed in Chapter V. For the Hoyle–Narlikar cosmology (Hoyle and Narlikar, 1974) with variable $G$ (cf. Section 7.7)

$$\alpha = 1, \quad \beta = 1, \quad \gamma = \tfrac{1}{2}; \quad (5)$$

the Hoyle–Narlikar cosmology without variable $G$ gives the same values for $\alpha, \beta, \gamma$ as general relativity. The same applies to the Dirac cosmology. We now consider these tests briefly.

(i) The gravitational redshift test is not a sensitive one for distinguishing between different theories. Even in Newtonian gravity it is possible to ascribe a photon of frequency $\nu$ with a gravitational mass $h\nu/c^2$ and to arrive at the same prediction as Einstein's general relativity.

(ii) The precession rate for the perihelion of Mercury gives a remarkably close agreement with the prediction of general relativity. The predicted rate for an orbit of semi-latus rectum $L$ is

$$\Omega_{GR} = \frac{6\pi GM}{L} \cong 43.03'' \text{ per century} \quad (6)$$

which may be compared with the observed rate of 43.11″ ± 0.45″ per century. The above observed rate is arrived at by eliminating the Newtonian contributions to the precession rate produced by the gravitational perturbation of other planets. This effect is as large as 532″ per century. Thus the residual Newtonian contribution $\Omega_N$ is zero by definition. In the general case of the line element (3) the predicted rate is

$$\Omega = \frac{(2 + 2\gamma - \beta)}{3}\Omega_{GR}. \quad (7)$$

We therefore get for the Brans–Dicke theory

$$\Omega_{BD} = \frac{3\omega + 4}{3\omega + 6}\Omega_{GR}. \quad (8)$$



If we assume that there is no *other* source of precession, e.g., a gravitational quadrupole moment contribution due to the Sun, then the Brans–Dicke theory would be ruled out unless $\omega$ is very large, say $\omega \geqslant 30$. Earlier Dicke (1974) had argued that there is evidence for solar oblateness which could produce a precession rate of $\sim 0.08\,\Omega_{GR}$. To account for the remaining 90 percent it is the necessary to have $\omega = 6.4$. However, more recent observations due to Hills *et al.* (1982) and Duvall *et al.* (1984) show that the quadrupole moment parameter $J_2$ is so small that it makes a negligible contribution to any solar system test (see Section 5.4). It is worth emphasizing, however, that *should* it turn out in the future that there is some cause for the precession rate even as small as ~5% of $\Omega_{GR}$, the remarkable agreement between GR and theory would disappear.

The Hoyle–Narlikar cosmology (1974) with variable $G$ would also seem to be ruled out by this test because it gives a much smaller precession rate:

$$\Omega_{HN} = \tfrac{2}{3}\Omega_{GR}. \quad (9)$$

Here again another source of precession rate ~14″/century is required to make the theory survive. The earlier version (1972b), however, survives.

As a postscript on behalf of classical general relativity one may add that the residual 43″/century mentioned in Section 2.9(ii) was recomputed by Narlikar and Rana (1985) by evaluating the Newtonian contribution to the perihelion precession of Mercury by other planets. In contrast to the traditional method of celestial mechanics, this approach was based on the N-body programme developed by S. J. Aarseth for use on a fast computer. The new method is more accurate and it shows that the residual precession rate is exactly accounted for by general relativity.

(iii) The bending of light test has recently been made much more accurate than the early attempts with visual light, thanks to the sophistication of microwave technology. The bending of an electromagnetic wave produced by the metric (3) is given by

$$\theta = \frac{4MG}{c^2R_0}\left(\frac{1 + \gamma}{2}\right) \quad (10)$$

where $R_0$ = radius of the Sun—or more precisely the distance of closest approach of the ray to the centre of the Sun. The bending produced by the four theories described above is as follows:

$$\theta_{GR} = \frac{4GM}{c^2R_0} = 1.75''$$

$$\theta_N = \tfrac{1}{2}\theta_{GR}, \quad \theta_{BD} = \frac{2\omega + 3}{2\omega + 4}\theta_{GR}, \quad \theta_{HN} = \tfrac{3}{4}\theta_{GR}.$$

Since the present accuracy of the experiments (Fomalent and Sramek, 1975) admits an experimental error of $\sim 5\%$ and since they are in good agreement with $\theta_{GR}$, the other three theories seem to be ruled out (except possibly the Brans–Dicke theory for $\omega \gtrsim 30$). The Dirac cosmology, the $C$-field cosmology of Section 4 and the Hoyle–Narlikar cosmology of constant $G$ give the same answer as general relativity.

(iv) The equality of gravitational and inertial mass is implied by the principle of equivalence on which general relativity is based. The Hoyle–Narlikar cosmology and the Dirac cosmology also imply the same principle. However, the Brans–Dicke theory departs from it in the following sense. For one massive object moving in the field of another the gravitational potential energy could change with space and time and this could lead to a difference in the gravitational and inertial mass. Such an effect would show up in the equations of motion of the object. Nordvedt (1968) calculated such an effect for example, for the Earth–Moon–Sun system. The result is a change in the Earth–Moon distance by an amount

$$\delta r = c_0 \eta \cos (L - L')$$

where $c_0$ is a constant $\approx 10$ metres, $L$ = mean longitude of the Moon and $L'$ = mean longitude of the Sun. The parameter $\eta$ is related to $\alpha$, $\beta$, $\gamma$ and to other weak field parameters. (For an exact and detailed discussion of $\eta$ see Nordvedt, 1968 and the general literature on parametrized post-Newtonian approximation—Misner, Thorne and Wheeler, 1973.) For the present purpose it is sufficient to approximate $\eta$ by

$$\eta = 4\beta - \gamma - 3. \tag{11}$$

Laser ranging of the Moon by two groups of observers (Williams *et al.*, 1976; Shapiro, Counselman and King, 1976) leads to $\eta = 0$ to a high degree of accuracy ($\lesssim 5\%$). For the Brans–Dicke theory this again implies a large $\omega \gtrsim 29$.

(v) *Radar echo delay* General relativity predicts that the round-trip echo delay for light signals travelling between the Earth and another planet, and grazing the limb of the Sun, will have an extra contribution relative to the situation when the Sun is not near the track. The general expression for the delay is

$$\tau = \frac{(1 + \gamma)}{2} \tau_{GR}$$

where $\tau_{GR}$ is the delay predicted by general relativily.



Recently Reasenberg *et al.* (1979) have analysed 14 months of data obtained from radio ranging to the Viking landers on Mars and the Viking spacecraft in orbit around it. They find that

$$\gamma = 1.000 \pm 0.002.$$

The uncertainty is based on the spread obtained in the estimates of $\gamma$ from many test solutions, and on a (subjective) judgement of the many procedures used. This value of $\gamma$ again rules out the second version of the Hoyle–Narlikar cosmology based on variable $G$; (1974 *op. cit.*) and restricts the Brans–Dicke parameter $\omega$ to $\geqslant 500$.

(vi) The gyroscope precession test (Everitt, 1974; Richard, 1975) is still to be made at the time of writing this article and it will be some time before its implications for the various gravitation theories become known.

To summarize, of the gravitation theories discussed here, those which make substantially different predictions from general relativity are the Bran–Dicke theory and the Hoyle–Narlikar cosmology with variable $G$. The rest of the theories do not seem to differ critically from general relativity in so far as the local Solar system tests are involved. These tests, especially (ii)–(v) have now become accurate enough to rule out the two rival theories mentioned above. The Brans–Dicke theory can survive with a large $\omega(\gtrsim 30)$ in which case it is very similar to general relativity for weak gravity.

We will now review tests of type (B).

## 11.2 Local astrophysical tests (B)

Under this category the following tests may be mentioned: (i) the age of the universe, (ii) the microwave background, (iii) the abundance of helium and deuterium.

(i) The age of the standard Friedmann cosmology is expressible in terms of two parameters $H_0$ and $q_0$, the present value of the Hubble constant and the deceleration parameter (see Section 2):

$$\tau(q_0, H_0) = H_0^{-1} \times \begin{cases} \dfrac{q_0}{(2q_0 - 1)^{3/2}} \left\{ \sin^{-1} \dfrac{q_0 - 1}{q_0} + \dfrac{\pi}{2} \right\} - \dfrac{1}{2q_0 - 1} & (k = 1) \\ \frac{2}{3} & (k = 0) \\ \dfrac{1}{(1 - 2q_0)} - \dfrac{q_0}{(1 - 2q_0)^{3/2}} \ln \dfrac{1 - q_0 + \sqrt{(1 - 2q_0)}}{q_0} & (k = -1). \end{cases} \tag{13}$$

In all cases $\tau \ngtr H_0^{-1}$. With the value of Hubble constant around 75 km s$^{-1}$ Mpc$^{-1}$, $H_0^{-1} \cong 13 \times 10^9$ years. This poses a problem of "accommodating" old objects in a time-scale of this order since the big bang. The age of our Galaxy (Hoyle and Fowler, 1960) is estimated at a figure lying in the range $(1-1.5) \times 10^{10}$ years. Galaxies older than ours are also likely to be in existence.

It is worth remarking that the estimates of both the Hubble constant and the age of the Galaxy have fluctuated over the past two decades. In the mid-1960s, $H_0^{-1} \cong 10^{10}$ years posed severe problems for standard cosmologies. However, the age of the Galaxy was revised down (Dicke, 1962b) to $\sim 7 \times 10^9$ years. Subsequently $H_0^{-1}$ was revised up to as high as $18 \times 10^9$ years and it became embarrassing to argue that our Galaxy was formed so much later than the big bang. This led to an upward estimate of galactic age (Smoot, Gorenstein and Muller, 1977) to $\sim 15 \times 10^9$ years, only to be faced with a general consensus that $H_0^{-1}$ should be revised down to $\sim 13 \times 10^9$ years!

Clearly the age problem will be with the standard models for some time to come and its implications can only be realized after $H_0$, $q_0$ and the age of the Galaxy settle down to stable values. Nevertheless it is safe to argue that models with $q_0 > 0.5$ give fairly low ages (see Figure 11.1) and are therefore more vulnerable to this test.

Most of the non-standard models discussed here are likely to face the same problems, with the possible exception of the steady state theory. Models with variable $G$ need to be looked at more carefully, however. For, stellar evolution rate is sensitively dependent on $G$ and with higher values of $G$ in the past stars must have evolved faster than in the $G$ = constant case. Consequently the

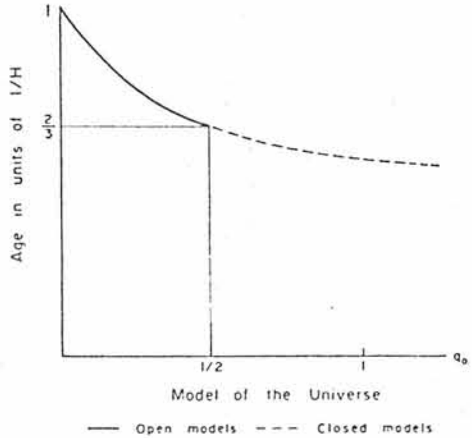



FIGURE 11.1 The age of the Friedmann universe is expressed in units of $H^{-1}$ for different values of $q_0$. The maximum possible age is $H^{-1}$ for an empty open universe.



age of the Galaxy as determined by stellar evolution must be lowered for such models. This may make the test easier to satisfy for these models than for the standard cosmologies.

In the steady state theory the average age of galaxies is $1/3H_0 \cong 4.3 \times 10^9$ years. Thus, although "infinitely old" galaxies exist they are very rare. We should be seeing more of the very young galaxies. This will pose a problem if galaxies formed one at a time. In the hot universe model (cf. Section 4.6) this difficulty can be avoided by forming galaxies in large groups.

The inflationary universe or any other cosmological theory that claims to resolve the flatness problem predicts the age of the universe to be $2/3H_0$. This age is too low to accommodate the astrophysically determined ages of stars, globular clusters and galaxies if $H_0^{-1} = 10^{10}$ years. Even with $H_0^{-1} = 18 \times 10^9$ years it is not possible to accommodate globular cluster ages $\gtrsim 12 \times 10^9$ years.

(ii) The microwave background is claimed to be strongest evidence to date for the standard cosmology. There are, however, certain aspects of this data which raise doubts whether everything is right with this picture. These are as follows:

First, the energy density *in the observed part* of the microwave radiation spectrum is $\sim 10^{-13}$ erg cm$^{-3}$. This is comparable to energy densities of many other astrophysical quantities, e.g. star light, cosmic rays and galactic magnetic field. Yet, according to the big-bang explanation, this coincidence of energy densities is accidental: it just so happens that at the epoch of observation the microwave background energy density happens to be at the value comparable to the energy densities of the other processes.

The second problem was referred to as the horizon problem in Section 10.1. Another problem is posed by the observed high degree of isotropy (temperature variation $\Delta T/T \lesssim 10^{-5}$ of the radiation background on the small angular scale of a few arc seconds to an arc minute). Such anisotropes are expected by galaxy formation scenarios if the galaxies formed after the epoch of recombination when radiation decoupled from matter.

A fourth difficulty springs from the interaction of cosmic rays of very high energies ($> 10^{18}$ eV) with the microwave background. As shown by Greisen (1966), Zetsepin and Kuzmin (1966), if the very high energy cosmic rays are extragalactic, they should lose energy while travelling intergalactic distances by collisions with the microwave photons. This would lead to a cut off or a steepening of the cosmic ray spectrum at energies $\gtrsim 10^{18}$ eV. No cut off has been noticed in the air shower primaries with energies $\gtrsim 10^{20}$ eV (Cunningham *et al.*, 1980; Linsley, 1980). Instead, flattening of the spectrum above $10^{19}$ eV has been reported (Horton *et al.*, 1983).

We should mention that in the 1970s there were two more difficulties with the standard interpretation of the microwave background that have now gone away. The so called dipole anisotropy of the microwave background found by two groups (Smoot *et al.*, 1977; Cheng *et al.*, 1979) could be explained by ascribing the Galaxy a velocity relative to the cosmological substratum. This would lead to an anisotropy in the observed Hubble law. Such anisotropy had been noted earlier by Rubin and Ford (Rubin *et al.*, 1976) from optical observations but its directionality and magnitude were different. However, later work of Hart and Davies (1982) using 21 cm velocity measurements did show Hubble flow anisotropy in agreement with observations of the microwave background.

The second difficulty was related to the spectrum of the microwave background. The direct measurements of the radiation background at frequencies around and beyond the peak frequency of the blackbody radiation curve showed significant ($\gtrsim 3\sigma$) departures of the spectrum from the best-fit Planckian curve (Woody and Richards, 1981). Several interpretations were advanced to explain the effect. However later studies seem to indicate that the spectrum is Planckian with temperature $\simeq 2.75$ K (see p. 326 for references).

We will now review the performance of the non-standard cosmologies with respect to microwave background.

The steady state theory does not have a "hot" era and so it has to find a new explanation for the microwave background. Such an explanation, if it is successful, will be astrophysical in origin and show how the existing energy density in other forms could be thermalized. Attempts that were made to explain the radiation in terms of microwave sources (Sciama, 1966; Gold and Pacini, 1968; Narlikar and Wickramsinghe, 1968; Wolfe and Burbidge, 1969) run into the difficulty of how to account for the fine scale isotropy of the radiation (Conklin and Bracewell, 1967). It turns out that the sources must be much more numerous than galaxies although they need not be intrinsically very energetic (Smith and Partridge, 1970). Going to the other extreme there have been attempts to produce the microwave radiation by thermalizing starlight with intergalactic dust grains in the form of long needles (Narlikar, Edmunds and Wickramsinghe, 1975). Such dust grains are made of graphite whiskers and may have cross-sectional radius of $\sim 10^{-5}$ cm but length of $\sim 1$ mm. This model has not been fully investigated. Its main problems are that it seems to need highly efficient conversion mechanism and that as yet there is no independent evidence for such grains in the intergalactic space.

Misner (1969) sought to solve the particle horizon problem by starting the universe from a homogeneous and anisotropic cosmology and later approaching the homogeneous and isotropic standard model. In this model



the universe originates with a big bang but has shear and rotation. The anisotropy is such that at any epoch there is at least one direction in which there is no particle horizon. However, this direction changes with time in a random fashion generating what is called the *mixmaster* universe. Misner hoped that by "mixing" in this fashion the information could be conveyed to arbitrarily large distances. However, Chitre (1972) has shown that this process does not work: the information does not mix enough to make the radiation background eventually homogeneous and isotropic.

In Section 7 (Section 7.5) we saw that in the conformal gravity theory of Hoyle and Narlikar it was possible to extend the past of the standard model to epochs *prior* to the space-time singularity. In the conformal frame in which the Einstein–de Sitter model changes to the Minkowski space-time the singularity is replaced by a zero mass surface. Hoyle (1975) used this extension to explain the origin of the cosmic microwave background. In this theory the starlight and other forms of radiation generated in the past half of the universe, i.e., in the $\tau < 0$ part of the Minkowski space-time are thermalized by electron scattering close to the zero mass surface $\tau = 0$. Near this surface, the scattering cross-section diverges leading to an efficient thermalization. The advantage of this model is that the coincidence of the energy density of the microwave background with the energy densities of other forms of radiation is explained.

In Segal's chronometric cosmology (see Section 9) the background radiation is divided into two classes: (a) the "pristine" which has made less than half a circuit of the universe since emission and (b) the "residual" part which may have made many circuits before interacting with matter. This latter radiation is expected to have thermalized in the infinite time available for its circulation. The energy density of this radiation (which is supposed to have originated in discrete sources) cannot be directly estimated but is expected to be comparable to other energy densities, e.g., of starlight. Thus a correct order of magnitude black body temperature would result. Segal (1976) identifies the residual radiation with the cosmic microwave background radiation and the pristine radiation with the light from stars, galaxies, etc. He estimates the ratio of energy of residual to pristine radiation to be $\sim 0.4\,\mu^{-1}r^{-1}$ where $\mu$ = number density of bright galaxies and $r$ = the radius of a typical galaxy treated as an absorbing sphere. Segal finds from this a value of background temperature nearly twice that observed but is not worried by the discrepancy in view of the uncertainties of various parameters in his estimation process.

Dirac's cosmology runs into difficulties in explaining the origin of the microwave radiation, if it really has the black-body spectrum (Dirac, 1975). If the characteristic energy $kT$ of the present day radiation is expressed in terms of a natural energy unit, say the electron rest energy, we get

$$\frac{m_e c^2}{kT} \cong 2 \times 10^9 \cong \left[\frac{t_0}{t_e}\right]^{1/4}, \tag{14}$$

where $t_0$ is the age of the universe and $t_e$ the time taken by light to travel across the classical electron radius. If follows from the Large Numbers Hypothesis (see Section 6) that $T \propto t^{-1/4}$, i.e., $T \propto S^{-1/4}$, since $S(t) \sim t$ in Dirac's cosmology. This is much slower than the rate $T \propto S^{-1}$ expected in the standard models. The slow rate of cooling requires that the microwave radiation is not decoupled from matter very long ago, for if it were it would cool as $\sim S^{-1}(t)$. We have to assume that the radiation is in interaction with some form of matter which is cooling at a rate $\sim t^{-1/4}$. Dirac suggests that the matter may be in the form of a tenuous intergalactic gas.

If it is assumed that the mean free path of the photons through the intergalactic gas is a distance corresponding to $z \cong 1$ (this is only a representative number), it follows from $1 + z = t_p/t_{emi}$ that the photons originate at the epoch $t_p/2$, where $t_p$ is the epoch of observation. The temperature of the intergalactic gas at the epoch of emission $t_{emi}$ is $T_e = 2^{1/4} T_p$, where $T_p$ is the present temperature of the gas. The temperature of the radiation at $t_{emi}$ is also $T_e$, and if it is assumed that the radiation propagates freely after $z = 1$, it cools as $t^{-1}$ thereafter and its present temperature is $2^{-2/3} T_p$.

In Dirac's cosmology the black-body nature of the spectrum is expected to be distorted for two reasons. Firstly, low frequency photons have a higher probability of scattering because of stimulated emission effects. The low frequency photons therefore originate closer to us than high frequency photons, and therefore at a temperature a little less than the temperature at which high frequency photons originate. But the low frequency photons cool less because they propagate freely for less time. The net effect is to distort the spectrum at the low frequency end. The second effect for distortion occurs when there is multiplicative creation taking place. In this case there is multiplication of photons in a beam and this causes the spectrum to deviate from the black-body law.

Recently Steigman (1978) has given an argument to show that a black-body nature of the spectrum is incompatible with particle creation, which is an essential feature of Dirac's cosmology. Steigman's argument is as follows: If a distribution of photons in the universe with average frequency $\nu$ and proper number density $n_\gamma(\nu)$ is to maintain a black-body spectrum as the universe expands, it is necessary that

$$\nu(t) n_\gamma^{-1/3}(t) = \text{constant.}$$



Using $\mathcal{N}_\gamma(t) = n_\gamma(t) S^3(t)$, where $\mathcal{N}_\gamma(t)$ is the comoving number density of photons, and $\nu(t)S(t) = $ constant which is valid for freely propagating photons, it follows from above that

$$\mathcal{N}_\gamma(t) = \text{constant} \tag{15}$$

if the black-body nature of the spectrum is to be preserved.

Now it is estimated that at the present moment the number of photons is some eight to ten orders of magnitude greater than the number of baryons, i.e., $(N_\gamma/N)_0 \cong 10^{10} \cong (t_0/t_e)^{1/4}$. It follows from the Large Numbers Hypothesis that

$$\frac{N_\gamma}{N} = \frac{\mathcal{N}_\gamma}{\mathcal{N}} \sim t^{1/4},$$

where $\mathcal{N}_\gamma$ and $\mathcal{N}$ are the comoving number densities of photons and matter respectively. According to Dirac's hypothesis (see Section 6) $\mathcal{N} \sim t^2$, i.e., $\mathcal{N}_\gamma$ is not a constant and an initial black-body nature of the photon spectrum will not be preserved as the universe expands. It follows from these considerations that if it is firmly established that the microwave background radiation has a black-body spectrum, then Dirac's cosmology in its present form would be incompatible with observation.

Canuto and Hsieh (1978) have argued that the above criticism can be answered in their formulation of Dirac cosmology as follows. From the Einstein metric $ds_E$ they construct another by a gauge (i.e., conformal) transformation

$$ds = \beta\, ds_E \tag{16}$$

where $\beta$ is a function of space and time. In the homogeneous isotropic case $\beta$ is a function of time only. In an arbitrary gauge a comoving volume of mass $M$ gives a generalized conservation law in the form

$$GM\beta = \text{constant.}$$

The 1937 matter conserving Dirac cosmology has $\beta G = $ constant while his 1973 version of the LNH has $G \propto \beta$. Canuto and Hsieh (1978) show that the choice of the gauge

$$G\beta^2 = \text{constant} \tag{17}$$

does preserve the condition (15) as the universe expands. Thus once the radiation is decoupled from matter it continues to remain in equilibrium. Its spectral function differs from that of a black-body by a $\beta$-factor which can be normalized to unity at the present epoch.

While this approach resolves the major difficulty of spectral shape being different, it still becomes necessary to give cogent reasons why the normalization to the black-body form is being achieved at the present epoch.

Recently Rees (1978) has proposed that the pregalactic radiation in a cold big bang universe can also be thermalized and would look like the present day 2.7°K radiation. The merit of the idea proposed by Rees is that it explains in astrophysical terms why the ratio of photons to baryons essentially stays constant during expansion at the value

$$\frac{n_\gamma}{n_b} \cong 3.6 \times 10^7 (\Omega h^2)^{-1}, \tag{18}$$

where $h$ = Hubble constant in units of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\Omega = 8\pi G\varrho_0/H_0^2$, the cosmological density parameter. In the Rees model it is assumed that in the early universe there were supermassive stars which formed when the age of the universe was $\lesssim 10^6$ years. These stars would evolve quickly and be past their bright phase long before galaxies formed. However, their radiation would be thermalized at redshifts $\gtrsim 100$ by the intergalactic gas and dust which would contain heavy elements, molecules etc. from nucleosynthesis of these supermassive stars. The thermalized background then cools down to the present 2.7°K temperature. Detailed calculations (Dirac, 1975) give a value for $n_\gamma/n_b$ close to the observed one. The full implications of this idea are still to be investigated.

(iii) Finally we consider the third test involving the abundance of light nuclei. The study of the relative abundance of chemical elements in various objects in the universe has revealed a "cosmic" distribution of abundances, with hydrogen and helium by far the most abundant, followed by the group C–N–O–Ne and with the group Li–Be–B and all elements heavier than nickel scarce. This raises the possibility that the elements were all generated in a single cosmic process. Detailed study has shown that hydrogen and helium, and small fractions of the heavier elements, can indeed be produced in the early, hot, dense stages of an evolving universe. We will consider here the production of the elements in the standard big bang models of the universe, and see what difference would be made to the abundances produced when non-standard cosmological models are used. We will consider mainly the cases of helium and deuterium, for these are the most important probes for what must have been happening in the primeval fireball.

In general the abundance determinations for $He^4$ give mass fractions $x$ of $He^4$ in the range $0.22 < x < 0.34$ for young stars in our Galaxy, in the interstellar medium in our Galaxy, and in other nearby galaxies. Considered especially interesting are the cases of dwarf blue galaxies in which the abundance of $He^4$ is normal, while the abundances of $O^{16}$ and $Ne^{20}$ are only 20% of their normal value. These properties suggest that these are young galaxies in which stars have produced few heavy elements while the $He^4$ is of primeval origin (for a review of abundances and references to the literature, see Wagoner, 1973, 1974; Reeves, Audoze, Fowler and Schramm, 1973).



Helium production in ordinary stars is inadequate to explain the observed values. If it is believed that the Galaxy has always had the same luminosity as at present, it is expected from most models of the Galaxy (for an estimate, see Hoyle and Taylor, 1964) that the production of $He^4$ by burning of hydrogen in stars can only contribute a mass fraction $\sim 0.01$–$0.04$. This fact together with the uniformity of $He^4$ abundance raises the possibility that the $He^4$ was produced in a cosmic process. If the abundance of $He^4$ is indeed universal, then allowing for its production in stars, its pregalactic abundance is expected to lie in the range $0.22 \leqslant x(He^4) \leqslant 0.32$.

Observations indicate (see Rees, 1978, and the references given there) that the mass fractional interstellar abundance of deuterium ($H^2$) is $2 \times 10^{-5}$. If this deuterium is of cosmological origin, then its pregalactic abundance would have been higher due to subsequent stellar destruction, but the factor is difficult to estimate reliably. We summarize in Table 11.1 the present day estimates of the abundances of the light elements in the interstellar medium.

It was first suggested by Gamow and his collaborators that the chemical elements could have been produced by nucleosynthesis in the early universe (Gamow, 1946; Alpher, Bethe and Gamow, 1948). We will now consider briefly the more recent detailed studies of nucleosynthesis in the fireball.

The theory of nucleosynthesis in the primordial universe was worked out in great detail by Peebles (1966) and by Wagoner, Fowler and Hoyle (1967). More recently Wagoner (1973) has carried out a calculation taking into account the improved knowledge of many of the cross-sections of the nuclear reactions involved in the processes and the neutron half-life.

The assumptions which define the standard big-bang picture are: (1) The universe was reasonably homogeneous and isotropic during the epoch of nucleosynthesis. (2) The temperature was once high enough for statistical equilibrium among all particles present, (3) only known particles were

TABLE 11.1

| Element | $H^2$ | $He^3$ | $He^4$ | $Li^6$ | $Li^7$ | $Be^9$ | $B^{10}$ | $B^{11}$ |
|---|---|---|---|---|---|---|---|---|
| Mass fraction | $2\times10^{-5}$ | $3\times10^{-5}$ | 0.22–0.34 | $4\times10^{-10}$ | $6\times10^{-9}$ | $1\times10^{-10}$ | $5\times10^{-10}$ – $5\times10^{-9}$ | $2\times10^{-9}$ $2\times10^{-8}$ |

(Taken from Wagoner, 1979)

present, (4) all particles were non-degenerate and (5) general relativity is valid. Excellent reviews of the scenarios considered by Peebles and Wagoner *et al.* have been given by Peebles (1971) and Weinberg (1972).

At very early times, when the temperature $T$ of the universe is above $10^{12}$ °K, it contains a great variety of particles in thermal equilibrium, including photons ($\gamma$), leptons ($e^{\pm}$, $\mu^{\pm}$), mesons ($\pi^{\pm}$, $\pi^{0}$) and nucleons and anti-nucleons ($n$, $\bar{n}$, $p$, $\bar{p}$). As the universe expands, its temperature drops and various particle annihilations and creations take place. At the time when the temperature is about $\sim 10^{12}$ °K, the universe contains $\gamma$, $\mu^{\pm}$, $e^{\pm}$, $\nu$, $\bar{\nu}$ (electron and muon neutrinos and antineutrinos). There is also a very small nucleonic contamination, and all the particles are in thermal equilibrium. The $\mu^{\pm}$ begin to annihilate at temperatures below $10^{12}$ °K, and are all gone at $T \cong 1.3 \times 10^{11}$ °K. The $\nu$, $\bar{\nu}$ decouple from the other particles and propagate freely, leaving behind $e^{\pm}$, $\gamma$ and a few nucleons in thermal equilibrium, with $T \alpha S^{-1}$, $S$ being the scale factor. As the temperature drops below $5 \times 10^{9}$ °K (which happens at about $t \cong 4$ secs, with $t = 0$ as the moment of creation), the $e^{\pm}$ pairs begin to annihilate, leaving as dominant constituents of the universe only $\gamma$, $\nu$, $\bar{\nu}$ in essentially free expansion, with the photon temperature higher than the neutrino temperature by a factor of $(11/4)^{1/3} \cong 1.4$. The neutron–proton ratio is now frozen at about 1 : 5. At a temperature of about $10^{9}$ °K ($t \cong 180$ secs) the neutrons rapidly begin to fuse with protons to form heavier nuclei, leaving an ionized gas of $H^1$ and $He^4$, with about 27% $He^4$ by mass, and traces of $H^2$, $He^3$ and other elements. As the universe continues to expand, the $\gamma$ and $\nu$ propagate freely, with the ionized gas temperature locked to the photon temperature till hydrogen recombines at $T \cong 4000$°K. At some temperature between $10^{3}$ °K and $10^{5}$ °K the energy density of the $\gamma$ and $\nu$ drops below the rest-mass density of hydrogen and helium and the universe enters the matter dominated era.

The mass-fractional abundances of various elements synthesized have been calculated after detailed consideration of the above scenario by Peebles (1966) and Wagoner and coworkers (Wagoner, 1973; Wagoner, Fowler and Hoyle, 1967). We reproduce in Table 11.2 some results from Wagoner (1973).

The $He^4$ abundance expected from the standard big-bang models compares well with the lower bounds of the abundance for $5 \times 10^{-32} \lesssim \varrho_0 \lesssim 10^{-28}$ gm cm$^{-3}$. A universe with $\varrho_0 \lesssim 6 \times 10^{-31}$ gm cm$^{-3}$ can also produce the required pregalactic deuterium abundance. It is seen that the abundance of $He^4$ produced does not depend very sensitively on the present density of matter. This means that any conclusion drawn regarding the nature of the $He^4$ production would not depend upon $\varrho_0$ which is poorly known. Only very small mass-fractions of nuclei heavier than $He^4$ are produced. This is principally because there are no stable nuclei with A = 3 or A = 8, which makes



TABLE 11.2

| A / $\varrho_0$ | $H^1$ | $H^2$ | $He^3$ | $He^4$ | $Li^7$ | Others |
|---|---|---|---|---|---|---|
| $10^{-31}$ | 0.763 | $6.2 \times 10^{-4}$ | $6.3 \times 10^{-5}$ | 0.236 | $5.2 \times 10^{-10}$ | $< 10^{-12}$ |
| $10^{-30}$ | 0.737 | $2.3 \times 10^{-5}$ | $2.1 \times 10^{-5}$ | 0.263 | $4.4 \times 10^{-9}$ | $< 10^{-12}$ |
| $10^{-29}$ | 0.719 | $2.5 \times 10^{-12}$ | $5.6 \times 10^{-6}$ | 0.281 | $4.3 \times 10^{-8}$ | $< 10^{-12}$ |
| $10^{-28}$ | 0.701 | $< 10^{-12}$ | $3.5 \times 10^{-6}$ | 0.299 | $2.9 \times 10^{-7}$ | $1.0 \times 10^{-10}$ |

A: Mass fractional abundance
$\varrho_0$: Present mass density in gms/cm$^3$

it difficult to build elements heavier than $He^4$ by $p$–$a$, $n$–$a$ or $a$–$a$ collisions. The low abundance of the heavier elements means that the production or destruction of these nuclei in stellar interiors can drastically alter their primordial abundance. It is for this reason that the cosmic abundance of $He^4$ and $H^2$ serve as the primary check on models of the early universe.

The abundances of elements produced in the fireball depend sensitively on the expansion rate $\dot{S}/S$ (Peebles, 1966). This is especially the case with $He^4$ for which a change in the expansion time scale by a factor only of two is sufficient to move the abundance outside the observed range. Any speed up in the expansion rate tends to increase the frozen-in abundance of neutrons at $T \cong 10^{9}$ °K, as there is less time for conversion of neutrons to protons. There are more neutrons available and therefore more $He^4$ is produced. But if the speed up factor $\geqslant 10^{4}$, there would not be enough time for $He^4$ to be produced before the density and temperature fall too low. On the other hand if the expansion time scale is lengthened by reducing $\dot{S}/S$, more neutrons would decay into protons before nucleosynthesis occurs, and less $He^4$ is produced. We show in Figure 11.2 how various abundances are affected by a speed up factor for a particular value of the present density.

Of the non-standard cosmologies only in the Brans–Dicke cosmologies there are detailed calculations of abundances.to compare with the predictions of the standard big bang models. We outline the main features of these below.

Dicke (1968) has investigated the problem of primordial nucleosynthesis in the representation of the Brans–Dicke theory in which $G$ is constant and particle masses vary. Greenstein (1968) has obtained the same results in the representation of the theory in which $G$ varies. We will consider Greenstein's approach as it is the simpler of the two.

In the representation of the Brans–Dicke theory in which particle masses are constant, physical processes are identical to those in general relativity (this is because the matter Lagrangian is independent of the $\phi$-field and all field equations except those for the gravitational field are identical to the corresponding

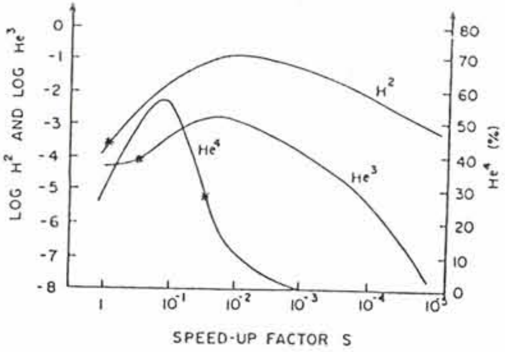



FIGURE 11.2 Mass fractional abundances of elements produced in the primordial fireball, for a present mass density of $\varrho_0 = 3 \times 10^{-31}$ gms cm$^{-3}$, as a function of an arbitrary speed up factor $S$. Solar system abundances are indicated by asterisks.

equations in general relativity). The only difference that the $\phi$-field makes is that it alters the rate of expansion from what it would be if the $\phi$-field were a constant. We have seen earlier that the abundances of elements synthesized are sensitively dependent on the expansion rate, i.e., on the time taken by the universe to reach a given stage in its evolution. This means that if the $\phi$-field can alter considerably the expansion rate, the abundances predicted for various elements would be very different from their standard big bang values.

Greenstein (1968) has investigated homogeneous and isotropic models under the same physical assumptions that Peebles and Wagoner *et al.* make. He has integrated the field equations (5.34) and (5.35) numerically in the case of "matter-dominated models" for which the integration constant $C = \phi(0)S^3(0) = 0$ (see Eq. 5.37). The solutions obtained have to be iterated to obtain solutions which satisfy the proper boundary conditions, $\phi S^3 \to 0$ as $t \to 0$ and $\phi \to [(2\omega + 4)/(2\omega + 3)]G_0^{-1}$ as $t \to t_0$, where $G_0$ is the gravitational constant at the present epoch $t = t_0$. Some of the results obtained by Greenstein are shown in Table 11.3.

$\omega = \infty$ is the general relativistic case; the abundances for this value of $\omega$ compare well with the values of Wagoner *et al.* For $\omega = 5$ there are no essential differences at low densities between the abundances obtained in the scalar–tensor and the general relativistic models. At high densities more $H^2$ and $He^4$ are formed in the scalar–tensor theory. It may be estimated from Greenstein's results that for $\omega \cong 30$, $\varrho_0 \lesssim 5 \times 10^{-30}$ gm would be allowed if the abundances produced are to lie in the limits set by observation. The upper bound on $\varrho_0$ is significantly less than the standard big bang upper bound of $\varrho_0 \lesssim 10^{-28}$ gm.

TABLE 11.3

| Present mass density in gms cm$^{-3}$ | $10^{-31}$ | | | $10^{-30}$ | | | $10^{-29}$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\omega$ | 5 | 10 | $\infty$ | 5 | 10 | $\infty$ | 5 | 10 | $\infty$ |
| Mass fraction of $H^2$ | $7.6 \times 10^{-4}$ | – | $6.6 \times 10^{-4}$ | $2.6 \times 10^{-5}$ | $2.1 \times 10^{-5}$ | $1.3 \times 10^{-5}$ | $3.4 \times 10^{-8}$ | $\sim 10^{-9}$ | $\sim 10^{-11}$ |
| Mass fraction of $He^4$ | 0.26 | – | 0.25 | 0.33 | 0.30 | 0.27 | 0.40 | 0.35 | 0.29 |

In the "$\phi$-dominated models" for which the integration constant $C = \phi(0)S^3(0) \neq 0$ (see Section 5.5.1B), the expansion rate can be made as large as one likes by making $C$ sufficiently large in absolute value. By adjusting $\varrho_0$ and $C$ the abundances can be made as high or as low as necessary. Defining

$$x = \frac{t \text{ (scalar tensor theory)}}{t \text{ (general relativity)}} \text{ at } T = 10^9 \,{}^\circ\mathrm{K}, \qquad (19)$$

there is only a small range of values of the present matter density $\varrho_0$ and $x$, which correctly reproduce the observed abundances: $x \cong 1$ and $\varrho_0 \cong 3 \times 10^{-31}$ gms/cm$^3$. If the expansion rate is increased significantly beyond this value the abundances of either $He^4$ or $He^3$ and $H^2$ become first too great and then too small. If $x \ll 1$, $H^2$ and $He^3$ are no longer too abundant but the abundance of $He^4$ is negligible. If $|S^3(0)\phi(0)|$ is so large as to have appreciable effects at the present epoch, $x \cong 0$ and no elements whatsoever are produced. If observations show that the "cosmic abundances" are much lower than now believed, scalar–tensor models with suitably low values of $C$ would provide possibilities for making theory compatible with observation.

## 11.3 Cosmological tests (C)

Under this heading the following tests are usually discussed: (i) the redshift-magnitude relation, (ii) the counting of radio sources, (iii) the variation of apparent angular size with distance.

We will discuss these tests in that order.

(i) The redshift magnitude relation is simply an extension of Hubble's law to large redshifts. The procedure is to find the theoretical relation between the apparent magnitude $m$ and the redshift $z$ of a particular class of galaxies in a given cosmological model and then to compare this relation with observations of that class of galaxies.

The relation in the Friedmann cosmologies for galaxies of luminosity $L$ and normalized intensity function $I(\lambda)$ is given by

$$m(z) = -2.5 \log L - 2.5 \log I[\lambda_0/(1 + z)] + 5 \log D(z) + 7.5 \log(1 + z) + \text{constant} \qquad (20)$$

where $D(z)$ is given by (Mattig, 1958, 1959)

$$D(z) = \frac{c}{H_0}[q_0 z + (q_0 - 1)(\sqrt{(1 + 2q_0 z)} - 1)](1 + z)^{-1}. \qquad (21)$$

In (15), $\lambda_0$ is the observed wavelength. Normally the astronomer measures the flux of radiation in wavelength band of specified width $\Delta$ around $\lambda_0$.



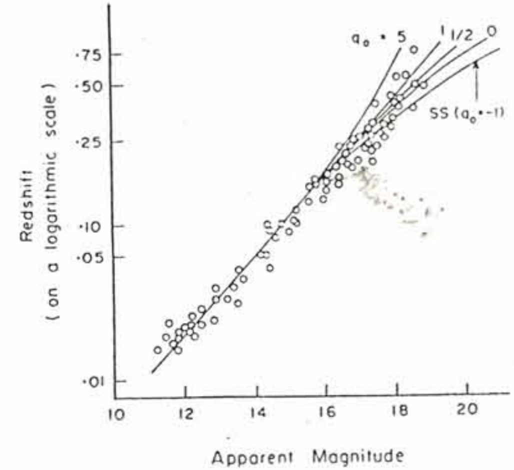



FIGURE 11.3 The redshift magnitude relation for brightest cluster members based on the work of Kristian, Sandage and Westphal (1978, Figure 3). Theoretical curves have been superposed on the data points.

The galaxies chosen for this test must have a small dispersion in their $L$-values. In practice this is difficult to achieve. The best so far has been the method adopted by Sandage — of picking the bright and massive elliptical which seems to dominate in a typical cluster of galaxies. Figure 11.3 shows the observed data with a superposition of theoretical curves. The $q_0 = -1$ curve is that for the steady state model. Although Sandage and his co-workers (Sandage, 1972; Kristian, Sandage and Westphal, 1978) often quote a "formal" value of $q_0 \cong 1$ the scatter in the data and the uncertainties are such as to keep the issue wide open.

The difficult issues concerned with this test have been reviewed by Gunn and Tinsley (1975) who conclude that it is generally difficult to fit all available information (e.g., nucleosynthesis, formation of galaxies, density of matter in the universe etc.) in a consistent way for any standard cosmology. The choice — or the room for manoeuvre is increased if we admit the $\lambda$-term (see Section 2).

The uncertainties in the $m$–$z$ relation arise from many reasons. The chief among them are (i) the dispersion in the luminosity of galaxies (ii) a possible epoch dependence of $L$ as the galaxies evolve (Tinsley, 1974) (iii) the possible absorption by intergalactic dust (Chitre and Narlikar, 1976) and (iv) a selection effect in favour of brighter galaxies at faint magnitudes.

The $m$–$z$ relation for Dirac cosmology has been studied extensively by Canuto (Canuto, Hsieh and Owen, 1979). In this case, in addition to any explicit stellar and galactic evolution which may be present, it is necessary to allow for various time dependencies which arise as a consequence of the Large Numbers Hypothesis. The gravitation "constant" $G \sim t^{-1}$, and the luminosity of a star can depend on a high power of $G$. If multiplicative creation is present, the mass of every star increases $\sim t^2$ and this too effects the luminosities. Canuto has taken into consideration all these effects in arriving at an $m$–$z$ relation valid for the case of multiplicative creation. It is not possible to arrive at a firm value of $q_0$ by comparing the theoretical prediction with the available data.

Segal's chronometric cosmology differs from all other cosmologies in predicting a non-linear Hubble law even at small redshifts. The redshift magnitude relation for this cosmology is

$$m = 2.5 \log z - 2.5\,(2 - \alpha) \log (1 + z) + M \qquad (22)$$

where $M$ is the absolute magnitude of the galaxy. Nicoll and Segal (1978) have made an extensive statistical analysis of dats on QSOs because they emphasize that the difference between (20) and (22) will be negligible for $z \lesssim 0.2$. The redshifts of galaxies exceeding 0.2 are very few while QSOs have substantially large redshifts (up to $z \cong 3.5$). It is well-known that the QSOs give a complete scatter diagram when $\log z$ is plotted against $m$; and the best-fit line does not have the Hubble slope. In order to sustain the Hubble interpretation for QSO redshifts many "ifs" and "buts" have to be introduced into the framework. The results of such efforts is that it is possible to argue that the $m$–$z$ diagram for QSOs is not inconsistent with the hypothesis that their redshifts are of cosmological nature (Bahcall and Hills, 1973; Burbidge and O'Dell, 1973; Kembhavi and Kulkarni, 1977). However this diagram cannot be suggested as a "proof" of Hubble's law.

Against this background Nicoll and Segal show that the quadratic relation (22) fits better than the linear relation (20), the observed $m$–$z$ plot for a sample of 592 QSOs. They base their conclusions on the least square fits of bright envelopes of the QSO data. The "bright envelope" technique involves (i) the arrangement of the sample in order of increasing $z$, (ii) the division of sample into bins of equal size $n$, (iii) the selection of the $k^{th}$ brightest object in the bin for $k$ usually $\lesssim n/2$. This subsample consists of the bright envelope. For a full statistical discussion we refer the reader to the paper by Nicoll and Segal (1978).

(ii) In the cosmological test involving radio-source count the relevant relation is between $N$, the number of radio-sources brighter than a flux density† $S$, and the flux density $S$. In a Euclidean universe with a uniform

†Flux density is a measure of the amount of radiation crossing normally unit area in unit time in a unit wavelength range.



distribution of radio-sources we expect

$$N(S) = \text{constant} \times S^{-3/2},$$

i.e.,

$$\frac{d \log N}{d \log S} = -1.5. \qquad (23)$$

In a standard cosmology the log $N$–log $S$ curve should start off with a slope of $-1.5$ and then its slope should flatten with decreasing $S$.

As a result of several radio-source surveys an overall log $N$–log $S$ curve is beginning to emerge. At high $S$ it starts off by being steeper than the Euclidean line – with a slope of $-1.8$. In the middle $S$-range the slope of the curve is $-1.5$ and then it flattens as required qualitatively by the standard cosmologies. We do not wish to enter into the controversy surrounding the interpretation of the log $N$–log $S$ curve. One thing is clear. *The standard models do not satisfy it.* However, they can be made to do so by ascribing evolutionary properties to radio sources, e.g., making them more numerous in the past up to say $z = 3$. This type of evolutionary scenario brings in considerations extraneous to cosmology and thereby weakens the data as a source of *cosmological* test.

Hoyle and Narlikar (1961) showed that the large scale units resulting from the hot universe model of the steady state universe (cf. Section 4) introduce inhomogeneity which can result in the log $N$–log $S$ being steeper than the Euclidean curve for small redshifts ($z \lesssim 0.2$]. This seems one way of salvaging the steady state theory which cannot use the loophole of evolution (of radio-sources] that works for the standard cosmologies.

It is also significant that the observed steepness of the log $N$–log $S$ curve is mainly due to a group of unidentified radio-sources. Until further information can be obtained about the nature of these sources and their distances, the interpretation of the log $N$–log $S$ curve remains largely a parameter fitting exercise.

(iii) The angular size test provides in principle a beautiful manifestation of the non-Euclidean geometry of space-time. If we calculate the angular size $\alpha$ of a radio source of linear size l (projected perpendicular to the line of sight) and redshift $z$, in a standard cosmology, we get a relation of the following type

$$\alpha = l\, f(z). \qquad (24)$$

The function $f(z)$ for small $z$ behaves as if $z$ were proportional to distance in a Euclidean universe: $f(z) \propto z^{-1}$. However, at large $z$ the rate of decrease of $f$ is slower and beyond a certain critical value of $z = z_c$, $f(z)$ increases with

$z$. This critical redshift $z_c$ is a function of $q_0$. For Einstein–de Sitter cosmology $z_c = 1.25$. In 1958 Hoyle had suggested that radio astronomy could be used to verify such a relation.

Unfortunately, as in the log $N$–log $S$ test, the actual application of this test tells us very little about cosmology since it brings in many intrinsic properties of the radio sources. Chitre and Narlikar (1977) have discussed these issues in their critique of the work of Swarup (1975) and Kapahi (1975) who had attempted to draw cosmological conclusions from their data of angular sizes and flux densities of radio sources. The test is therefore unlikely to be very significant in deciding between standard or non-standard cosmologies.

Roeder (1975) has pointed out how sensitive this test is to the assumption of homogeneity in the universe. If there are large scale clumps of matter the theoretical conclusions drawn by Hoyle are significantly altered. Indeed, the plot of $\alpha$ against $z$ will tell us more about the inhomogeneity parameter $\delta\varrho/\varrho$ (= fractional departure from the smoothed out density of the universe) than about the large scale geometry of space-time.

It is possible to work out the various theoretical predictions for these three tests in the non-standard cosmologies discussed in this article. However, these predictions either do not differ from those of standard cosmology; or, if they do, the differences are usually masked by the observational uncertainties. We do not therefore consider it worthwhile at this stage to go into further details.

## 11.4 Variation of fundamental constants

Here we shall consider three possible tests: (i) the variation of the gravitational constant, $G$ (ii) the variation of masses of fundamental particles (iii) the variation of the fine structure constant $e^2/\hbar c$.

It is of course possible to imagine variations of many other basic quantities — but we consider here only the above three because for these some observational inputs are available.

(i) In the gravitation theories of Newton and Einstein, the gravitational field couples to matter through the constant of gravitation $G$. In more recent theories like those of Brans–Dicke and Hoyle–Narlikar, a constant $G$ is replaced by a scalar field. This means that in the homogeneous cosmological models based on these theories, $G$ should be a function of the cosmological time $t$. Though there is no theory of gravitation to provide a framework for Dirac's cosmology, the Large Numbers Hypothesis predicts in a straightforward manner that $G \sim t^{-1}$.

The time variation of $G$ predicted in all these cosmological models provides a potentially very strong test of the theories on which the models are based. If it can be shown experimentally that $G$ is constant with sufficient accuracy, then all the theories which have a non-constant $G$ would be effectively ruled out. On the other hand if experiment does point towards a non-constant $G$, it will be necessary to replace Einstein's theory, *whatever* may be the outcome of any other exxperiment to test gravitation theories. We will consider in this section the direct and indirect evidence currently available on a non-zero rate of change of $G$.



*Direct evidence* Most cosmological models with a changing $G$ predict that $G$ should be decreasing with time. An immediate consequence of decreasing $G$ is a slow adiabatic expansion of all orbits in the Solar system about their primaries. To see how this comes about, we will consider a planet in a circular orbit about a central mass $M$. If $v$ is the velocity of the planet in its orbit of radius $r$, we have

$$GM = v^2 r. \tag{25}$$

If it is assumed that as $G$ varies, the central mass $M$, the mass of the planet $m$, and its angular momentum $L = mvr$ are all constant, it is apparent from (1) that

$$r \propto v^{-1} \propto G^{-1}, \tag{26}$$

and we have for the rate of change of the period $T = 2\pi r v^{-1}$ of the planet,

$$\frac{\dot G}{G} = -\frac{\dot T}{2T}, \tag{27}$$

with the dot indicating differentiation with respect to the time $t$. In a more rigorous calculation, Vinti (1974) has shown that for two-body motion with $G$ varying inversely with time and averaging over a complete revolution,

$$\frac{\dot G}{G} = -\frac{\dot T}{2T} = \frac{\dot n}{2n} = -\frac{\dot a}{a}, \tag{28}$$

where $a$ is the semi-major axis and $n$ the mean motion. It is assumed in this derivation that all the masses involved are constant. This is no longer true if the creation of matter implied by Dirac's Large Numbers hypothesis is also taken into account. The general expression for $\dot G/G$ is

$$\frac{\dot G}{G} = f\frac{\dot n}{n}, \tag{29}$$

where $f = \frac{1}{2}$ for the case of constant masses, $f = 1$ and $-1$ for Dirac's cosmology with multiplicative and additive creation respectively.

Van Flandern (1981) has made use of the change in the motion of the moon about the Earth to determine $\dot G/G$. Since the Earth is slowed down in its orbit around the Sun if $G$ decreases, an astronomical time-scale derived from the observed motion of the Sun around the Earth, such as ephemeris time, will suffer a slowing down at the same rate. The motion of the moon around the Earth, on the other hand, is affected by varying $G$, as well as other by causes such as tidal friction. If the acceleration of the mean longitude of the moon is measured using atomic time, which is presumed to be uniform, the total acceleration from all causes is obtained. When the same acceleration is measured using ephemeris time, any contribution due to varying $G$ is absorbed into the time scale and excluded. The difference between the two acceleration gives the amount exclusively due to decreasing $G$.

Using all the available data, Van Flandern has determined a weighted mean for acceleration in ephemeris time to be

$$\dot{n}^{et} = (-28.8 \pm 1.5)''/(\mathrm{cy})^2 \tag{30}$$

where cy stands for century. The experience with determination of $\dot{n}^{et}$ shows that there could be unknown sources of errors in this result.

Lunar occultation timings since 1955 (when atomic time first became available), can be used to determine the acceleration as measured in atomic time. Combining his observations with preliminary results from Lunar laser ranging observations, Van Flandern (1981) has determined that

$$\dot{n}^{at} = (-23.2 \pm 1.2)''/(\mathrm{cy})^2. \tag{31}$$

Here again the situation regarding systematic errors is uncertain, and the result must be viewed with caution. In fact $\dot{n}^{at}$ as determined here is on the opposite side of $\dot{n}^{\mathrm{tidal}}$ as compared with earlier results (Van Flandern, 1975). A differencing of Eqs (31) and (30) yields

$$\frac{\dot{n}}{n} = (3.2 \pm 1.1) \times 10^{-11}/\mathrm{yr}, \tag{32}$$

using $n = 1.73 \times 10^{9''}/\mathrm{cy}$ for the Moon. In a theory with constant masses for astronomical bodies, this would lead to a gravitational constant *increasing* with time. This would also be the case for Dirac's cosmology with multiplicative creation, whereas for additive creation we have

$$\frac{\dot{G}}{G} = -(3.2 \pm 1.1) \times 10^{-11}/\mathrm{yr}.$$

The most accurate measurement to date of the rate of change of the gravitational constant has been made from an analysis of the range data to the Viking landers on Mars (Hellings *et al.*, 1984). The data used in the analyis include (1) range measurements to the Viking landers, (2) to the Mariner 9 spacecraft in orbit around Mars, (3) radar bounce range measurements from the surfaces of Mercury and Venus, (4) lunar laser range



measurements and (5) optical position measurements of the Sun and planets. A least-square fit of the parameters of the solar system model to the data limits a simple variation in the Newtonian gravitation constant to

$$\frac{\dot{G}}{G} = (0.2 \pm 0.4) \times 10^{-11}/\mathrm{yr}.$$

The quoted errors are much larger than the formal standard deviations, and represents uncertainties is the masses of asteroids which are used in the analysis. The result is clearly consistent with zero variation in the gravitation constant. This cannot be considered to be entirely inconsistent with the result in Eq. (32) because of the many uncertainties involved there.

*Indirect evidence* A decrease in the gravitational constant at the rate obtained by Van Flandern can have drastic effects on terrestrial and astrophysical phenomena over geologic and cosmologic time scales. Some of the effects which may be expected to be present have been discussed by Jordan (1971), Dicke (1972c), Hoyle (1972) and others. We will briefly consider some of the more interesting of these effects.

The interior of the Earth is compressed by the weight of the overlying layers. If $G$ is decreasing, the weight of the overlying layers should be decreasing, and this should result in the expansion of the Earth. Murphy and Dicke (1964) have estimated from calculations on Earth models that the rate of expansion is given by

$$\frac{\delta r}{r} = -0.1\,\frac{\delta G}{G}. \tag{33}$$

For $\dot{G}/G = -8 \times 10^{-11}/\mathrm{yr}$, this gives $\delta r = 0.005\,\mathrm{cms/yr}$.

Hoyle and Narlikar (1972b) have also calculated the expansion rate of the Earth due to decreasing $G$. They use a three zone model of the Earth, consisting of a fluid core of radius 3500 kms, a non-liquid mantle extending to about 6300 kms and a thin crust of a few tens of kilometers. Assuming that the system undergoes a homologous evolution as $G$ decreases, they calculate a rate of increase of the radius of the Earth of about 0.01 cm/yr. Hoyle and Narlikar estimate that the horizontal pressures which may arise during the complicated changes which the Earth undergoes because of its tendency to expand can move portions of the Earth's crust horizontally, providing a mechanism for continental drift. Lyttleton and Fitch (1977) have questioned the validity of Hoyle and Narlikar's approach. Proceeding numerically with the system of differential equations, they have obtained a rate of increase of the Earth's radius which is one-fourth the Hoyle–Narlikar value. Lyttleton and Fitch have pointed out that for their value of the rate of change of the radius, a standard length on the surface of the Earth would increase at the rate of $2 \times 10^{-3}$ cm yr, which is only one thousandth of the widely believed rate of continental drift.

From the data currently available, Wesson (1973) estimates that the Earth must be expanding at the rate of about 0.05 cm/yr. Wesson points out that such an expansion rate is useful in explaining many geophysical phenomena. Expansion from a completely sial-covered globe of about 3700 kms radius at a constant rate of about 0.06 cm/yr would give the continents their present configuration, perhaps modified by continental drift. The mid-oceanic ridge system which completely surrounds Antarctica and the tectonic fabric of the sea floor are better explained by expansion than by convection currents in the mantle.

McElhinny, Taylor and Stevenson (1978) have used palaeo-magnetic data to estimate the rate of expansion of the Earth. They arrive at the conclusion that any expansion is limited to less than 0.8% in the last 400 M yrs. This corresponds to a maximum possible expansion rate of 0.013 cm yr$^{-1}$, which is about one-fourth the value obtained by Wesson. McElhinny *et al.* have used the estimated expansion rate of the Earth to obtain upper bounds on the rate of change of $G$. For this they write an equation of the form $\dot{R}/R = \alpha \dot{G}/G$, where $R$ is the radius of the Earth, and use various methods to estimate the value of $\alpha$. The result is that if planetary masses are constant (in Dirac's cosmology this would correspond to additive creation), the data for Earth gives $|\dot{G}/G| \lesssim 3 \times 10^{-10}$ yr$^{-1}$. Similar analysis for Moon, Mercury and Mars lead to limits which are at least one order of magnitude less than this. If it is assumed that planetary masses increase as $\sim t^2$ (in Dirac's cosmology this corresponds to multiplicative creation), the Earth give $|\dot{G}/G| \lesssim 5 \times 10^{-9}$ yr$^{-1}$ and the data for Mercury give $|\dot{G}/G| \lesssim 2.5 \times 10^{-11}$ yr$^{-1}$.

The theoretical predictions of the rate of expansion of the Earth are less than the observed values reported by Wesson by nearly (or more than) an order of magnitude. Even after allowing for the fact that the theoretical estimates are model dependent and possibly subject to upward revision, this means that a decreasing gravitational constant makes only a minor contribution to the total expansion. It will be difficult in the present state of geophysics to isolate the contribution due to decreasing $G$, if any. Nevertheless, it may be safely assumed thta the hypothesis of a decreasing gravitational constant is consistent with geophysical phenomena.

The luminosities of stars vary as a high power of $G$. For the Sun the dependence on $G$ is about $G^{6.5}$, Hoyle (1972) has estimated that at the time of the origin of the Solar system, about $4.5 \times 10^9$ years ago, the Sun must have been brighter than at present by a factor of about 3. Allowing for the



fact that the Earth must have been closer to the Sun in the past ($r \propto G^{-1}$), this means that the Solar constant 4.5 billion years ago must have had 5 times the present value. Hoyle has estimated that the temperature on the surface of the Earth in any epoch is given by $\sim 280\eta^{1/4}$°K, where $\eta$ is the factor by which the Solar constant is increased. On this reckoning 4.5 billion years ago the temperature should have been around 150°C.

The oldest fossil evidence for life on Earth goes back to about $3 \times 10^9$ years ago. The Solar constant then had about 3 times the present value, and the temperature would be about 100°C. It is known that the oldest known life forms, bacteria, can survive at this temperature. As $\eta$ decreased to the present day value of unity, the average temperature would have fallen from $\sim 100$°C to the present day value of $\sim 10$°C. If the hypothesis that $G$ is decreasing is true, there should be a correspondence between these temperatures and the maximum fossil ages of life forms which can exist at these temperatures. This expectation is not contrary to known facts. It is of course possible that the maximum temperatures on the Earth could have decreased in the same way due to effects which are entirely independent of a decreasing gravitational constant.

We have been considering the effects of a higher Solar luminosity in the past on conditions on the Earth. The consequences of such a higher luminosity on the Sun itself can also be important. If 4.5 billion years ago (at the time of the origin of the Solar system) the Sun in its initial main sequence state had a luminosity substantially higher than at the present, and if the luminosity only slowly decreased to its present value, then the Sun must have burned a substantially larger fraction of its hydrogen store than is computed under the assumption of a constant luminosity. If the rate of hydrogen burning had been too large, the Sun would have turned off the main sequence by now and would have been on its way to becoming a red giant.

Pochoda and Schwarzschild (1964) have carried out detailed computations to set limits to the degree to which $G$ can vary without producing unacceptable consequences. They find that $G \sim t^{-0.2}$ produces no difficulty for the representation of the observed Sun as the end product of an evolution starting with the initial main sequence rate and lasting for 4.5 billion years. For $G \sim t^{-1}$ a satisfactory representation of the present Sun is obtained only if the age of the universe is at least about 15 billion years. This limitation arises because with a shorter age of the universe the initial main sequence state of the Sun falls relatively earlier in the history of the universe when $G$ would have been rather high. This causes so large an initial luminosity that the Sun would turn off the main sequence before 4.5 billion years had elapsed. When Pochoda and Schwarzschild did their calculations, it was believed that the universe was only $\sim 10$ billion years old, and so the variation of $G$ as fast as $t^{-1}$ seemed to be ruled out. But the present value of Hubble's constant shows that the universe must be at least ~20 billion years old, and the evolution of the Sun is consistent with $G \sim t^{-1}$.

Hoyle has pointed out some cosmological effects which may be expected if $G$ is decreasing. Stars in galaxies formed at early epochs would be subject to extremely rapid evolution. The initial main sequence would soon be eaten away to considerably smaller masses than is normally taken to be the case. As time increases towards the present epoch the luminosity of the remaining evolved stars would fall to low values. Consequently such systems would appear at the present epoch as galaxies of very high mass-to-light ratios. Elliptical galaxies have such high mass-to-light ratios, a circumstance which is difficult to understand on the basis of conventional theory; see Barnothy and Tinsley (1973) for a critique of this point of view and a refutation of their criticism by Canuto and Narlikar (1980) who have shown that the present cosmological observations are consistent with the Hoyle–Narlikar cosmology with variable $G$ (Hoyle and Narlikar, 1972b). By considering the effects of decreasing $G$ on spiral galaxies consisting of stars in the range of 0.5 to 0.75 $M_{\odot}$, Hoyle has shown that these should appear at high redshift as blue, almost star-like objects. The quasi-stellar-galaxies discovered by Sandage could possibly be such objects.

*Comparison between theories and observations* We now consider the non-standard cosmologies which predict $|\dot{G}| > 0$ in the light of the available data.

The comparison is most easily made in the case of Dirac's cosmology (Section 6), for the Large Numbers Hypothesis leads uniquely to $G \sim t^{-1}$, $S(t) \sim t$, i.e., $H \sim t^{-1}$, where $S(t)$ is the scale factor of the universe and $H = \dot{S}/S$ is Hubble's "constant" at the epoch $t$. It follows that

$$\left.\frac{\dot{G}}{G}\right|_{0(\text{Dirac})} = -\frac{1}{t_0} = -H_0. \tag{34}$$

For $H_0 = 50\,\text{kms sec}^{-1}\,\text{Mpc}^{-1}$, the currently favoured value, we have $(\dot{G}/G)_{0(\text{Dirac})} = -5.1 \times 10^{-11}\,\text{yr}^{-1}$, which is an order of magnitude larger than the value favoured by the Viking observations.

In the case of the Brans–Dicke theory the situation is more complicated as there is no unique cosmological model allowed by the theory. In the simplest cases where an analytical solution is available, it is possible to relate $\dot{G}/G$ at the present epoch to parameters like $H_0$ and $\varrho_0$, $\varrho_0$ being the present matter density. In all other cases it is necessary to resort to numerical integration in order to express $(\dot{G}/G)_0$ in terms of the parameters measurable at the present epoch.

In the case of the pressure free cosmological model with flat space-sections ($k = 0$) (see Section 5.5.1.A) we have $S \sim t^{2\omega+2/3\omega+4}$, $\phi \sim t^{2/3\omega+4}$. Since $G \propto \phi^{-1}$, we get



$$\left(\frac{\dot{G}}{G}\right)_0 = -\frac{2}{(3\omega + 4)t_0} = -\frac{H_0}{(\omega + 1)}. \tag{35}$$

Using $H_0 = 50\,\text{kms sec}^{-1}\,\text{Mpc}^{-1}$, we get $(\dot{G}/G)_0 \cong -7 \times 10^{-11}$/yr for $\omega = 6$, which was the value of $\omega$ favoured by Dicke, and $(\dot{G}/G)_0 \cong -9.72 \times 10^{-13}$ for $\omega = 500$, which is the lower limit put by the Viking experiments of Reasenberg and others (see Section 5.4). The value of $|(\dot{G}/G)_0|_{k=0, p=0}$ predicted by the Brans–Dicke theory is therefore at least an order of magnitude smaller than $|\dot{G}/G|$ obtained by Hellings *et al.*.

The solutions of Nariai (Eqs. 5.39–42) for the case $k = 0$, $p \neq 0$, give

$$\left.\frac{\dot{G}}{G}\right|_0 = -\frac{B}{A}H_0 = -\frac{(1 - 3n)H_0}{\omega(1 - n) + 1}. \tag{36}$$

For $n = 0$, which is the pressure free case, we recover Eq. (35). $|(\dot{G}/G)_0|_{k=0, p\neq 0}$ decreases as $n$ increases to 1/3 from 0, and $H_0/(\omega + 1)$ is the upper limit on $|\dot{G}/G|_0$ for models with flat space-sections and equation of state $p = n\varrho$. The introduction of pressure does not alleviate the difficulty faced by the $k = 0$, $p = 0$ model *vis-a-vis* the value of Hellings *et al.* This was first pointed out by Morganstern (1979), who has also given upper bounds for $|\dot{G}/G|_0$ for cosmological models with $k \neq 0$. He finds, that the bounds are model dependent, and that the curvature contribution to $|\dot{G}/G|_0$ is of the same order of magnitude as the flat space contribution, and adds to it or subtracts from it for positive or negative curvatures respectively.

For the Hoyle–Narlikar cosmology with variable gravity (Section 7.7) we have $(\dot{G}/G)_0 = -H_0$. For $H_0 = 50\,\text{kms sec}^{-1}\,\text{Mpc}^{-1}$ this gives $(\dot{G}/G)_0 = -5.1 \times 10^{-11}$/yr, which is the same as in Dirac's cosmology. Using various independent lines of reasoning, Muller (1976) has made a detailed comparison of the astronomical data, relevant to the rate of change of the gravitational constant, and the predictions of various theories, current at that time.

(ii) In the discussion of Hoyle–Narlikar cosmology (cf. Section 7) we saw that the phenomenon of redshift should be explained in terms of a secular increase in particle masses. In Dirac cosmology also the two time scales can be interpreted to imply an increase of elementary particle masses on the global time scale. Are these possible tests of this hypothesis of epoch dependent particle mass?

Unfortunately, a straightforward comparison of this effect in a homogeneous and isotropic universe does not seem possible. This is because the mass variation produces (in most cases) observable effects similar to those of the expansion of the universe. There are, however, cases involving local inhomogeneities which promise a possible test of such hypotheses.

In Section 7.6 we saw how in the modified Hoyle–Narlikar cosmology, it is possible for two objects $p$, $q$ situated close to each other, to exhibit markedly different redshifts. The prediction in this case was that of the pair of such objects the one with the larger redshift should be younger. Are there observations to test such a prediction?

Since the discovery of the quasi-stellar objects (the QSOs) one remarkable feature to emerge is the phenomenon of *anomalous redshifts*. Basically an anomalous redshift implies the possibility of an extragalactic object possessing a redshift *not* in accordance with Hubble's law. In the case of QSOs it has been found that a galaxy and a QSO with small angular separation have markedly different redshifts. The galaxy redshift is usually small ($\lesssim 0.5$) and obeys Hubble's law. The QSO redshift is, however, large ($\gtrsim 1$). We have already commented on the fact that the redshift–magnitude relation for QSOs is a scatter diagram with hardly any Hubble type correlation. Could it be that the redshifts of QSOs do *not* follow Hubble's law? In the above example, we would then have to say that the large redshifts of QSOs are *not* due to the expansion of the universe.

In recent years many such pairs of close companions with different redshifts have come to light (Arp, 1977). In a typical pair ($p$, $q$), $p$ is a QSO with large redshift and $q$ a galaxy with moderate or small redshift. The first question is whether the two members of the pair are physically close to each other. Is it likely that in reality they are far apart but that their lines of sight happen to lie close to each other?

To answer this question we have to know what are the number distributions of QSOs and galaxies across the sky so that we may estimate the change of their lines of sight coming within the observed small angle of separation. Only recently this question could be answered with some degree of confidence and it is beginning to look highly improbable that we should have seen so many "chance" cases of contiguous lines of sight to objects of differing redshifts.

Clearly considerable caution needs to be exercised to arrive at a conclusion like this — for it hits at the very basis of the idea of cosmological redshift. Even granted that a redshift anomaly exists, it becomes necessary first to try the conventional explanations of the Doppler or the gravitational redshifts. Only if these fail would we be forced to consider other explanations. The explanation advanced in Section 7.6 falls in this last category. More observational and theoretical work on the idea was done by Narlikar and Das (1980) who have shown that it offer an explanation of the present data on anomalous redshifts. The accumulation of the data has indeed come to a stage where the conventional "standard" cosmologist must start taking the anomalous redshifts seriously.



(iii) We saw in Section 7 that Dirac's Large Numbers Hypothesis predicts that $e^2/Gm_p m_e \sim t$, where $t$ is the cosmic time expressed in natural units. Dirac suggested that this implies $G \sim t^{-1}$, the other quantities being assumed to be constant. Chiefly in order to avoid the difficulties which were once believed to occur if $G \sim t^{-1}$, Gamow (1967a) suggested that the above relation ought to be interpreted to mean that $e^2 \sim t$, with the other quantities remaining constant. If $h$ and $c$ are assumed to be constant, this means that the fine structure constant $\alpha = e^2/hc$ is proportional to the cosmic time.

Bahcall and Schmidt (1967) have measured any possible variation of the fine structure constant with time using the spectra of radio galaxies with redshifts of about 0.2. The light travel time for these galaxies is believed to be about $2 \times 10^9$ years. Bahcall and Schmidt use the O III multiplet line, wavelengths 5007 and 4959 Å, in the emission spectra of five radio galaxies. Given the observed wavelengths, the ratio $\alpha(z)/\alpha(\text{lab})$ can be computed from the relation

$$\left[\frac{\alpha(z)}{\alpha(\text{lab})}\right]^2 = \left[\frac{\delta\lambda}{\lambda}\right]_{(\text{observed})} \cdot \left[\frac{\delta\lambda}{\lambda}\right]^{-1}_{(\text{lab})}, \tag{37}$$

where $\delta\lambda$ is the fine-structure splitting and $\lambda$ is the weighted mean wavelength, weighted according to $2J + 1$. It is found that

$$\frac{\alpha(z = 0.2)}{\alpha(\text{lab})} = 1.001 \pm 0.002 \text{ (probable error)}. \tag{38}$$

The hypothesis that $\alpha$ is proportional to cosmic time requires $\alpha(z = 0.2)/\alpha(\text{lab}) = 0.8$ and is ruled out.

Roberts (1977) has compared the redshifts measured at optical wavelengths and at 21 cms in extragalactic sources to find that $|\dot{\alpha}/\alpha| \leqslant 4 \times 10^{-12}\,\text{yr}^{-1}$. If it is assumed that $\alpha \sim t$, we get in Dirac's cosmology $\dot{\alpha}/\alpha = t^{-1} = H^{-1}$, i.e., $(\dot{\alpha}/\alpha)$. (Dirac) $= 5.1 \times 10^{-11}\,\text{yr}^{-1}$ for $H_0 = 50\,\text{kms sec}^{-1}\,\text{Mpc}^{-1}$.

## 11.5 Concluding remarks

The above review, although necessarily incomplete, gives some indication of the prevailing situation in observational cosmology. Recalling Feynman's comment quoted in the beginning of this work we emphasize the various uncertainties of the many tests discussed here. Until the observational tests become more definitive many of the cosmological arguments will continue to centre round the conceptual questions such as the internal consistency of the theory, its width of applicability, its interaction with the rest of physics and last but, not the least, how it fares under Occam's razor. Viewed from these angles the non-standard cosmologies discussed here have many commendable features. In any case, by providing possible rival alternatives to the standard cosmology, these theories serve the useful purpose of maintaining cosmology a science and not a dogma.

## Acknowledgements

The authors record their grateful thanks to Mr. T. K. Abraham and Mr. P. Joseph for efficient typing of the manuscript, to Mr. S. R. Palekar and his colleagues in the drawing office for assistance with the diagrams and to Mrs. M. M. Vajifdar and her staff for the reproduction facilities.

## References

Abell, G. O. (1974). In M. S. Longair (Ed), *Confrontation of Cosmological Theories with Observational Data*, IAU Symposium No. 63, Riedel, Dordrecht.
Albrecht, A., and P. J. Steinhardt (1982). *Phys. Rev. Lett.*, 48, 1220.
Alfven, H. (1965). *Rev. Mod. Phys.*, 37, 652.
Alfven, H., and O. Klein (1963). *Ark. Fys.*, 23, 187.
Alpher, R. A., H. A. Bethe, and G. Gammow (1948). *Phys. Rev.*, 73, 80.
Alpher, R. A., and R. Herman (1958). *Science*, 128, 904.
Anderson, J. L. (1967). *Principles of Relativity Physics*, Academic, New York.
Anderson, J. L., *et al.* (1975). *Astrophys. J.*, 200, 221.
Arnowitt, R., S. Deser, and C. W. Misner (1962). In L. Witten (Ed), *Gravitation — An Introduction to Current Research*, John Wiley, New York.
Arp, H. (1977). In *Decalages vers le Rouge et Expansion de l'Universe*, IAU Colloquium No. 37 and CNRS Colloquium No. 263, Paris.
Atkatz, D., and H. Page (1982). *Phys. Rev.*, D25, 2065.
Bahcall, J. N., and R. E. Hills (1973). *Astrophys. J.*, 179, 699.
Bahcall, J. N., and M. Schmidt (1967). *Phys. Rev. Lett.*, 19, 1294.
Barnothy, J. M., and B. M. Tinsley (1973). *Astrophys. J.*, 182, 343.
Barrow, J. D. (1984). In X. Fustero and E. Verdaguer (Eds), *Relativistic Astrophysics and Cosmology*, p. 137, World Scientific, Singapore.
Barrow, J. D., and J. Stein-Schabes (1984) Preprint.
Belinskii, V. A., and I. M. Khalatnikov (1973). *Sov. Phys.-JETP*, 36, 591.
Belinskii, V. A., I. M. Khalatnikov, and E. M. Lifshitz (1970). *Adv. Phys.*, 19, 525.
Bianchi, L. (1918). *Lezioni sulla Teoria dei Gruppi Continui Finiti Transformazion*, Spoeri, Pisa.
Bondi, H. (1960). *Cosmology*, Cambridge University Press.
Bondi, H. (1961). *Cosmology*, 2nd ed., Cambridge University Press.
Bondi, H., and T. Gold (1948). *Mon. Not. R. Astron. Soc.*, 108, 252.
Bonnevier, B. (1965). *Ark. Fys.*, 27, 35.
Born, M. (1943). *Experiment and Theory in Physics*, Cambridge University Press.
Brans, C., and R. H. Dicke (1961). *Phys. Rev.*, 124, 925.
Brot, R. *et al.* (1980). *Nuc. Phys.*, B170, 228.
Brown, R. W., and F. W. Stecker (1979). *Phys. Rev. Lett.*, 43, 315.
Burbidge, G. R., E. M. Burbidge, R. A. Fowler, and F. Hoyle (1957). *Rev. Mod. Phys.*, 29, 547.
Burbidge, G. R., and S. L. O'Dell (1973). *Astrophys. J.*, 183, 759.
Callan, C. G., and S. Coleman (1977). *Phys. Rev.*, D16, 1762.
Canuto, V. (undated). *Observational Tests of Dirac's Cosmology*, Institute of Space Studies reprint.
Canuto, V., and S. H. Hsieh (1978). *Astrophys. J.*, 224, 302.
Canuto, V., S. H. Hsieh, and J. R. Owen (1979). *Scale Covariance and G-varying Cosmology. III*, Institute of Space Studies reprint.



Canuto, V., and J. Londenquai (1977). *Astrophys. J.*, 211, 342.
Canuto, V., and J. V. Narlikar (1979). *Cosmological Tests of the Hoyle-Narlikar Cosmology*, Preprint.
Carr, B. J., and M. J. Rees (1979). *Nature*, 278, 605.
Cartan, E. (1922). *Compt. Rend.*, 174, 593.
Carter, B. (1974). In M. S. Longair (Ed). *Confrontation of Cosmological Theories with Observational Data*, IAU Symposium No. 63, Reidel, Dordrecht.
Cheng, E. S., P. R. Saulson, D. T. Wilkinsen, and B. E. Corey (1979). *Astrophys. J.*, 232, L139.
Chitre, D. M. (1972). *Investigations of the Vanishong of a Horizon for Bianchi Type IX (Mixmaster) Universe*, Doctoral Dissertation, University of Maryland.
Chitre, S. M., and J. V. Narlikar (1976). *Astrophys. Sp. Sci.*, 44, 101.
Chitre, S. M., and J. V. Narlikar (1977). *Mon. Not. R. Astron. Soc.*, 180, 525.
Colman, S., and E. Weinberg (1973). *Phys. Rev.*, D7, 1888.
Combes, F., O. Fassi-Fehri, and B. Leroy (1975). *Astrophys. Sp. Sci.*, 37, 151.
Conklin, E. K., and R. N. Bracewell (1967). *Nature*, 216, 777.
Counselman, C. C., *et al.* (1974). *Phys. Rev. Lett.*, 33, 1621.
Cunningham, G., *et al.* (1977). In *Proceedings of the Fifteenth International Cosmic Ray Conference*, Bulgarian Academy of Sciences, Plodiv.
Cunningham, G., *et al.* (1980). *Astrophys. J.*, 236, L71.
Davidson, W. (1960). *Nature*, 187, 583.
Davidson, W., and A. B. Evans (1973). *Int. J. Theoret. Phys.*, 7, 353.
Davidson, W., and A. B. Evans (1977). *Comm. Roy. Soc. Edin. (Phys. Sci.)*, 10, 123.
Deser, S., and F. A. E. Pirani (1965). *Proc. Roy. Soc. Lond. A*, 288, 133.
de Sitter (1917). *Proc. Akad. Weteusch. Amsterdam*, 19, 1217.
DeWitt, B. S. (1967a). *Phys. Rev.*, 160, 1113.
DeWitt, B. S. (1967b). *Phys. Rev.*, 162, 1195.
DeWitt, B. S. (1967c). *Phys. Rev.*, 162, 1239.
Dicke, R. H. (1961). *Nature*, 192, 440.
Dicke, R. H. (1962a). *Phys. Rev.*, 125, 2163.
Dicke, R. H. (1962b). *Nature*, 194, 329.
Dicke, R. H. (1962c). *Rev. Mod. Phys.*, 34, 110; *Science*, 138, 653.
Dicke, R. H. (1968). *Astrology. J.*, 152, 1.
Dicke, R. H. (1974). *Science*, 184, 419.
Dicke, R. H., and H. M. Goldenberg (1967). *Phys. Rev. Lett.*, 18, 313.
Dicke, R. H., and P. J. E. Peebles (1979). In S. W. Hawking and W. Israel (Eds), *General Relativity: A Survey*, p. 504, Cambridge University Press.
Dirac, P. A. M. (1937). *Nature*, 139, 323.
Dirac, P. A. M. (1938). *Proc. Roy. Soc. Lond. A*, 165, 199.
Dirac, P. A. M. (1973a). *Proc. Roy. Soc. Lond. A*, 333, 403.
Dirac, P. A. M. (1973b). *Pont. Acad. Comentarii*, 11, 46.
Dirac, P. A. M. (1974). *Proc. Roy. Soc. Lond. A*, 338, 439.
Dirac, P. A. M. (1975a). In A. Perlmutter and S. M. Widmayer (Eds), *Theories and Experiments in High Energy Physics*, Plenum, New York.
Dirac, P. A. M. (1975b). In J. Mehra (Ed), *Physicists' Conception of Nature*, Reidel, Dordrecht.
Duvall, T. L. Jr., W. A. Dziembowski, P. R. Goode, D. O. Gough, J. W. Harvey, and J. W. Leibacher (1974). *Nature*, 310, 22.
Eddington, A. S. (1930). *Mon. Not. R. Astron. Soc.*, 90, 668.
Eddington, A. S. (1953). *Fundamental Theory*, Cambridge University Press.
Einstein, A. (1915). *Preuss. Akad. Wiss. Berlin, Sitzber.*, 844.
Einstein, A. (1917). *Preuss. Akad. Wiss.. Sitzber.*, 142. (English translation in H. A. Lorentz *et al.*, *The Principle of Relativity*, Dover).
Einstein, A. (1949). In P. A. Schilpp (Ed). *Albert Einstein: Philosopher–Scientist.*
Eisenhart, L. P. (1949). *Riemannian Geometry*, Princeton University Press.
Ellis, G. F. R. (1967). *J. Math. Phys.*, 8, 1171.

Ellis, G. F. R. (1971). In R. K. Sachs (Ed), *General Relativity and Cosmology*, Academic, New York.
Elvius, A., E. T. Karlson, and B. E. Laurent (1974). In M. S. Longair (Ed), *Confrontation of Cosmological Theories with Observational Data*, IAU Symposium No. 63, Reidel, Dordrecht.
Evans, A. B., and W. Davidson (1969). *Einstein Spaces*, Pergamon, Oxford.
Evans, A. B., and W. Davidson (1973). *Mon. Not. R. Astron. Soc.*, 165, 323.
Everett, H. (1957). *Rev. Mod. Phys.*, 29, 454.
Everitt, C. W. F. (1974). In B. Bertotti (Ed), *Experimental Gravitation, Enrico Fermi Course LVI*, Academic, New York.
Fichtel, C. E., *et al.* (1975). *Astrophys. J.*, 177, 341.
Fichtel, C. E., *et al.* (1977). In R. D. Wills and B. Battrick (Eds), *Recent Advances in Gamma-Ray Astronomy*, Twelfth ESLAB Symposium.
Field, G. B. (1972). In L. Goldberg *et al.* (Eds), *Annual Review of Astronomy and Astrophysics*, Vol. 10, Annual Review Inc., Palo Alto.
Fomalont, E. B., and R. A. Sramek (1975). *Astrophys. J.*, 199, 749.
Friedmann, A. (1922). *Z. Phys.*, 10, 377.
Gamow, G. (1946). *Phys. Rev.*, 70, 572.
Gamow, G. (1948). *Phys. Rev.*, 74, 505.
Gamow, G. (1967a). *Phys. Rev. Lett.*, 19, 759 (with erratum on p. 1000).
Gamow, G. (1967b). *Proc. Nat. Acad. Sci. USA*, 57, 187.
Gates, S. J. *et al.* (1983). *"Superspace", Frontiers in Physics*, Vol. 58, Benjamin, New York.
Georgi, H., and S. L. Glashow (1974). *Phys. Rev. Lett.*, 32, 438.
Gibbons, G. W., and S. W. Hawking (1977). *Phys. Rev.*, D15, 2752.
Gittus, J. H. (1975). *Proc. Roy. Soc. Lond. A*, 343, 155.
Godel, K. (1949). *Rev. Mod. Phys.*, 21, 447.
Gold, T., and F. Hoyle (1958). In *IAU Symposium No. 4 on Radio Astronomy*, University of Stanford Press.
Gold, T., and F. Pacini (1968). *Astrophys. J.*, 152, L115.
Goldhaber, M. (1956). *Science*, 124, 218.
Gould, R. J., and G. F. Burbidge (1963). *Astrophys. J.*, 138, 969.
Greenstein, G. S. (1968). *Astrophys. Sp. Sci.*, 2, 155.
Greisen, K. (1966). *Phys. Rev. Lett.*, 16, 748.
Gunn, J. E. (1977). In *Decalages vers le Rouge et Expansion de l'Universe*, IAU colloquium No. 37, Paris.
Gunn, J. E., and J. R. Gott (1972). *Astrophys. J.*, 176, 1.
Gunn, J. E., and B. M. Tinsley (1975). *Nature*, 257, 454.
Gurevich, L. E., A. M. Finkelstein, and V. A. Robson (1972). *Astrophys. Sp. Sci.*, 22, 231.
Guth, A. (1981). *Phys. Rev.*, D23, 347.
Hagstrom, R. (1977). *Phys. Rev. Lett.*, 38, 729.
Harrison, E. R. (1967). *Phys. Rev. Lett.*, 18, 1011.
Harrison, E. R. (1973). *Comm. Astrophys. Sp. Phys.*, 5, 109.
Hart, L., and R. D. Davies (1982). *Nature*, 297, 191.
Hawking, S. W. (1965). *Proc. Roy. Soc. Lond. A*, 286, 313.
Hawking, S. W., and G. F. R. Ellis (1973). *Large Scale Structure of Space-Time*, Cambridge University Press.
Heckman, O., and E. L. Schucking (1955). *Z. Astrophys.*, 38, 95.
Heckman, O., and E. L. Schucking (1956). *Z. Astrophys.*, 40, 81.
Hehl, F. W. (1973). *Gen. Rel. Grav.*, 4, 333.
Hehl, F. W. (1974). *Gen. Rel. Grav.*, 5, 491.
Hehl, F. W. (undated). *Phys. Lett. A*, 36, 225.
Hehl, F. W., P. Von der Heyde, and G. D. Kerlick (1974). *Phys. Rev. D*, 10, 1066.
Hehl, F. W., P. von der Heyde, and G. D. Kerlick (1976). *Rev. Mod. Phys.*, 48, 393.
Hellings, R. W., P. J. Adams, J. D. Anderson, M. S. Keesey, E. L. Lau, E. M. Standish, V. M. Canuto, and I. Goldman (1983). *Phys. Rev. Lett.*, 51, 1609.



Hilbert, D. (1915). *Grundlagen der Physik*, 1 Mitt., Nachr. Ges. Wiss. Gottingen.
Hill, H. A., *et al.* (1974). *Phys. Rev. Lett.*, 33, 1497.
Hill, H. A., and R. T. Stebbins (1975). *Astrophys. J.*, 200, 471.
Hill, H. A., J. B. Randal, and P. R. Goode (1982). *Phys. Rev. Lett.*, 49, 1794.
Hoyle, F. (1947). *Mon. Not. R. Astron. Soc.*, 107, 334.
Hoyle, F. (1948). *Mon. Not. R. Astron. Soc.*, 108, 372.
Hoyle, F. (1959). In R. N. Bracewell (Ed), *Paris Symposium on Radio Astronomy*, IAU Symposium No. 9.
Hoyle, F. (1965). *Galaxies, Nuclei and Cosmology*, Heinemann, London.
Hoyle, F. (1972). *Quart. J. Roy. Astron. Soc.*, 13, 328.
Hoyle, F. (1975). *Astrophys. J.*, 196, 661.
Hoyle, F., and W. A. Fowler (1960). *Ann. Phys.*, 10, 280.
Hoyle, F., and J. V. Narlikar (1961). *Mon. Not. R. Astron. Soc.*, 123, 133; 125, 13.
Hoyle, F., and J. V. Narlikar (1962a). *Observatory*, 82, 13.
Hoyle, F., and J. V. Narlikar (1962b). *Mon. Not. R. Astron. Soc.*, 125, 131.
Hoyle, F., and J. V. Narlikar (1963). *Proc. Roy. Soc. Lond. A*, 273, 1.
Hoyle, F., and J. V. Narlikar (1964a). *Proc. Roy. Soc. Lond. A*, 278, 465.
Hoyle, F., and J. V. Narlikar (1964b). *Proc. Roy. Soc. Lond. A*, 282, 1964.
Hoyle, F., and J. V. Narlikar (1966a). *Proc. Roy. Soc. Lond. A*, 290, 143.
Hoyle, F., and J. V. Narlikar (1966b). *Proc. Roy. Soc. Lond. A*, 290, 166.
Hoyle, F., and J. V. Narlikar (1966c). *Proc. Roy. Soc. Lond. A*, 294, 138.
Hoyle, F., and J. V. Narlikar (1972a). *Mon. Not. R. Astron. Soc.*, 155, 305.
Hoyle, F., and J. V. Narlikar (1972b). *Mon. Not. R. Astron. Soc.*, 155, 323.
Hoyle, F., and J. V. Narlikar (1972c). In F. Reines (Ed), *Cosmology, Fusion and Other Matters*, Colorado Associated University Press.
Hoyle, F., and J. V. Narlikar (1972d). *Nuovo Cim. A*, 7, 242.
Hoyle, F., and J. V. Narlikar (1974). *Action at a Distance in Physics and Cosmology*, Freeman & Co., San Francisco.
Hoyle, F., and R. J. Taylor (1964). *Nature*, 203, 1108.
Horton, L., *et al.* (1983). In *Proc. 18th ICCRC*, P. V. Ramana Murthy (Pub.), Tata Institute of Fundamental Res, Bombay.
Hubble, E. P. (1929). *Proc. Nat. Acad. Sci. USA*, 15, 169.
Hughes, V. W. (1964). In H. Y. Chiu (Ed), *Gravitation and Relativity*, Benjamin, New York.
Isham, C. J., A. Salam, and J. Strathdee (1971). *Phys. Rev. D*, 3, 867.
Isham, C. J., A. Salam, and J. Strathdee (1973). *Nature Phys Sci.*, 244, 82.
Islan, J. N. (1968). *Proc. Roy. Soc. Lond. A*, 306, 487.
Jeffereys, H. (1948). *Nature*, 162, 822.
Johnson, D. G., and D. T. Wilkinson, (1986). *Astrophys. J.*, 313, L1.
Jordan, P. (1971). *The Expanding Earth*, International Series of Monographs in Natural Philosophy, Vol. 37, Pergamon, Oxford.
Jordan, P. (1974). *Die Herkunft der Sterne*, Stuttgart.
Kaluza, Th. (1921). *Sitzungsber. Preuss. Akad Wiss. Berlin, Math. Phys.*, K1, 966.
Kapahi, V. K. (1975). *Mon. Not. R. Astron. Soc.*, 172, 513.
Kasner, E. (1921). *Am. J. Math.*, 43, 217.
Kembhavi, A. K. (1977). *Pramana*, 7, 344.
Kembhavi, A. K. (1978). *Mon. Not. R. Astron. Soc.*, 185, 807.
Kembhavi, A. K., and V. K. Kulkarni (1977). *Mon. Not. R. Astron. Soc.*, 181, 19.
Kibble, T. W. B. (1961). *J. Math. Phys.*, 2, 212.
King I. (1961). *Observatory*, 81, 128.
Klein, O. (1926). *Z. Phys.*, 37, 895.
Klein, O. (1958). In R. Stoops (Ed), *La Structure et l'Evolution de l'Univers*.
Kopczynski, W. (1972). *Phys. Lett. A*, 39, 219.
Kopczynski, W. (1973). *Phys. Lett. A*, 43, 63.
Krishnaswamy, M. R., *et al.* (1982). *Phys. Lett.*, 115B, 349.

Kristian, J., A. Sandage, and J. A. Westphal (1978). *Astrophys. J.*, 221, 383.
Krushaar, W. L., *et al.* (1972). *Astrophys. J.*, 177, 341.
Kuchowicz, B. (1976a). *Astrophys. Sp. Sci.*, 39, 157.
Kuchowicz, B. (1976b). *Astrophys. Sp. Sci.*, 40, 167.
Landau, L. D., and E. M. Lifshitz (1975). *Classical Theory of Fields*, 4th rev. ed, Pergamon, Oxford.
Laurent, B. E., and L. Soderholm (1969). *Astron. Astrophys.*, 3, 197.
Layzer, D. (1954). *Astron. J.*, 59, 268.
Lehnert, B. (1977). *Astrophys. Sp. Sci.*, 46, 61.
Lemaitre, G. (1927). *Ann. Soc. Sci. Bruxelles*, 47A, 49; English translation in *Mon. Not. R. Astron. Soc.*, 91, 483 (1931).
Leroy, B., J. P. Nicholle, and G. Schatzman (1973). In F. W. Stecker and J. L. Trombka (Eds), *Gamma-Ray Astrophysics*.
Lifshitz, E. M., and I. M. Khalatnikov (1963). *Adv. Phys.*, 12, 185.
Linde, A. D. (1982). *Phys. Lett.*, 108B, 389.
Linde, A. D. (1984). *Rep. Prog. Phys.*, 47, 925.
Luisley, J. (1980). In M. Wada (Ed), *Catalogue of Highest Energy Cosmic Rays*, Inst. of Physical and Chemical Research, Word Data Center C2, No. 1, 5.
Lyttleton, R. A., and H. Bondi (1959). *Proc. Roy. Soc. Lond. A*, 252, 313.
Lyttleton, R. A., and J. P. Fitch (1977). *Mon. Not. R. Astron. Soc.*, 180, 471.
Mach, E. (1960). *The Science of Mechanics*, Open Court, La Salle.
MacCallum, M. A. H. (1973). In E. Schatzman (Ed), *Cargese Lectures in Physics*, Vol. 6, Gordon and Breach, New York.
McCrea, W. H. (1951). *Proc. Roy. Soc. Lond. A*, 206, 562.
McCrea, W. H. (1954). *Astron. J.*, 60, 271.
McCrea, W. H. (1960). *Nature*, 186, 1035.
McCrea, W. H. (1960–1). *La Nuovo Critica*, Serie III, 11 (de Luca Editore in Roma).
McElhinny, M. W., S. R. Taylor, and D. J. Stevenson (1978). *Nature*, 271, 316.
Marzke, R. F., and J. A. Wheeler (1964). In H. Y. Chiu and W. F. Hoffmann (Eds), *Gravitation and Relativity*, Benjamin, New York.
Melchiorri, F., and L. Maiani (Eds) (1985). *Cosmic Background Radiation and Fundamental Physics*. Italian Physical Society, Bologna.
Mattig, W. (1958). *A. N.*, 284, 109; 285, 1.
Milne, E. A. (1935). *Relativity, Gravitation and World Structure*, Clarendon, Oxford.
Milne, E. A. (1948). *Kinematic Relativity*, Clarendon, Oxford.
Milne, E. A., and W. H. McCrea (1934). *Quart. J. Math.*, 5, 73.
Misner, C. W. (1969). *Phys. Rev. Lett.*, 22, 1071.
Misner, C. W., and A. H. Taub (1969). *Sov. Phys.-JETP*, 28, 122.
Misner, C. W., K. P. Thorne, and J. A. Wheeler (1973). *Gravitation*, Freeman, San Francisco.
Morganstern, R. E. (1971). *Nature*, 232, 109.
Morganstein, R. E. (1972). *Nature Phys. Sci.*, 237, 70.
Muller, P. M. (1976). *Determination of the Cosmological Rate of Change of G and the Tidal Acceleration of Earth and Moon from Ancient and Modern Data*, JPL Reprint.
Murphy, C. T., and R. H. Dicke (1964). *Proc. Am. Phil. Soc.*, 108, 224.
Nariai, H. (1968). *Prog. Theor. Phys.*, 40, 49.
Nanz, Jr., R. H. (1953). *J. Geol.*, 61, 51.
Narlikar, J. V. (1963). *Mon. Not. R. Astron. Soc.*, 126, 203.
Narlikar, J. V. (1968). *Proc. Camb. Phil. Soc.*, 64, 1071.
Narlikar, J. V. (1973). *Nature Phys. Sci.*, 242, 135.
Narlikar, J. V. (1974). *Pramana*, 2, 158.
Narlikar, J. V. (1977). *Ann. Phys. (NY)*, 107, 325.
Narlikar, J. V. (1978). *Lectures on General Relativity and Cosmology*, Macmillan, New Delhi.
Narlikar, J. V. (1983). *Introduction to Cosmology*, Jones and Bartlett, Boston.
Narlikar, J. V. (1984). *J. Astrophys. Astronon.*, 5, 67.



Narlikar, J. V., M. G. Edmunds, and N. C. Wickramsinghe (1975). In M. Rowan Robinson (Ed), *Far Infrared Astronomy*, Proceedings of Windsor Conference, Pergamon, Oxford.
Narlikar, J. V., and T. Padmanabhan (1983). *Phys. Rep.*, 100, 151.
Narlikar, J. V., and N. C. Rana (1985). *Mon. Not. R. Astron. Soc.*, 213, 657.
Narlikar, J. V., and N. C. Wickramsinghe (1968). *Nature*, 216, 43.
Nelson, A. H., and G. Rowlands (1975). *Astrophys. Sp. Sci.*, 33, L1.
Newman, E., L. Tamburino, and T. Unit (1963). *J. Math. Phys.*, 4, 915.
Nicoll, J. F., and I. E. Segal (1978). *Proc. Nat. Acad. Sci. (USA)*, 75, 535.
Mordvedt, Jr., K. (1968). *Phys. Rev.*, 169, 1014, 1017; 170, 1186.
North, J. D. (1965). *The Measure of the Universe*, Clarendon, Oxford.
Omnes, R. (1972). *Phys. Rep.*, 3C, 1.
Omnes, R., and J. L. Puget (1974). In M. S. Longair (Ed), *Confrontation of Cosmological Theories with Observational Data*, IAU Symposium No. 63, Reidel, Dordrecht.
Oppenheimer, J. R., and H. Snyder (1939). *Phys. Rev.*, 56, 455.
Ozsvath, L., and E. Schucking (1962). In *Recent Developments in General Relativity*, Pergamon, Oxford.
Padmanabhan (1983a). *Phys. Lett.*, 93A, 116.
Padmanabhan (1983b). *Phys. Rev.*, D28, 745, 756.
Papapetrou, A. (1974). *Lectures on General Relativity*, Reidel, Dordrecht.
Park, H. S., *et al.* (1985). *Phys. Rev. Lett.*, 54, 22.
Peebles, P. J. E. (1966). *Astrophys. J.*, 146, 542.
Peebles, P. J. E. (1971). *Physical Cosmology*, Princeton University Press.
Penrose, R. (1975). In C. J. Isham, R. Penrose, and D. W. Schiama (Eds), *Quantum Gravity – An Oxford Symposium*, Oxford University Press, Oxford.
Penzias, A. A., and R. W. Wilson (1965). *Astrophys. J.*, 142, 419.
Peres, A. (1962). *Nuovo Cim. Supp.*, 24, 389.
Petrov, A. Z. (1969). *Einstein Spaces*, Pergamon, Oxford.
Pochoda, P., and M. Schwarzschild (1964). *Astrophys. J.*, 139, 587.
Pound, R. V., and G. A. Rebka (1960). *Phys. Rev. Lett.*, 4, 337.
Ramani, A., and J. L. Puget (1973). *Astron. Astrophys.*, 51, 411.
Raychaudhari, A. K. (1955). *Phys. Rev.*, 98, 1123.
Raychaudhari, A. K. (1975). *Phys. Rev. D*, 12, 952.
Reasenberg, R. D., I. I. Shapiro, J. C. MacNeil, R. B. Goedstein, J. C. Bridenthal, J. P. Brenkle, D. L. Cain, T. M. Kaufman, T. A. Komarek, and A. I. Zygielbaum (1979). *Astrophys J.*, 234, L219.
Rees, M. J. (1978). *Nature*, 275, 35.
Reeves, H., J. Audize, W. A. Fowler, and D. N. Schramm (1973). *Astrophys. J.*, 179, 909.
Reinhardt, M., and M. A. F. Thiel (1970). *Astrophys. Lett.*, 7, 101.
Richard, J. (1975). In E. Shaviv and J. Rosen (Eds), *General Relativity and Gravitation (GR 7)*, Wiley, New York.
Roberts, M. S. (1977). In *Decalages vers le Rouge et Expansion de l'Univers, l'Evolution des Galazies et ses Implications Cosmologiques*, IAU Colloquium No. 37 and CNRS Colloquium No. 263, Paris.
Robertson, H. P. (1935). *Astrophys. J.*, 82, 248.
Robertson, H. P. (1936). *Astrophys. J.*, 83, 187, 257.
Roeder, R. C. (1975). *Nature*, 255, 124.
Rowan-Robinson, M. G. (Ed) (1976). *Far Infrared Astronomy*, Pergamon, Oxford.
Rowan-Robinson, M. G., J. Negroponte and J. Silk (1979). *Nature*, 281, 635.
Roxburgh, I. W. (1977). *Nature*, 265, 763.
Rubin, V. C., N. Thonnard, and W. K. Fold, Jr. (1976). *Astron. J.*, 81, 687, 719.
Salam, A. (1968). In N. Swartholm (Ed), *Elementary Particle Physics*, p. 367, Almquist and Wiksells.
Sandage, A. (1972). *Astrophys. J.*, 178, 1.
Sandage, A. R., and G. A. Tamman (1975). *Astrophys. J.*, 197, 265.

Schwarzschild, K. (1916). *Sitzber. Deut. Akad. Wiss. Berlin, Kl. Math.-Phys. Tech.*, 424.
Sciama, D. W. (1953). *Mon. Not. R. Astron. Soc.*, **113**, 34.
Sciama, D. W. (1962). In *Recent Developments in General Relativity*, Pergamon, Oxford.
Sciama, D. W. (1964). *Rev. Mod. Phys.*, **36**, 463, 1103.
Sciama, D. W. (1966). *Nature*, **211**, 277.
Segal, I. E. (1976a). *Mathematical Cosmology and Extragalactic Astronomy*, Academic, New York.
Segal, I. E. (1976b). *Proc. Nat. Acad. Sci. USA*, **73**, 669.
Senjanovic, G., and F. W. Stecker (1980). *Phys. Lett.*, **B96**, 285.
Shapiro, I. I., C. C. Counselman, and R. W. King (1976). *Phys. Rev. Lett.*, **36**, 555.
Shapiro, I. I., *et al.* (1971). **26**, 1132.
Shapiro, I. I., *et al.* (1976). *Bull. Am. Astron. Soc.*, **8**, 308.
Smith, M. G., and R. B. Partridge (1970). *Astrophys. J.*, **159**, 743.
Smoot, G. F., M. V. Gorenstein, and R. A. Muller (1977). *Phys. Rev. Lett.*, **39**, 898.
Stecker, F. W. (1978). *Nature*, **273**, 493.
Stecker, F. W. (1981). In R. Ramaty and F. C. Jones (Eds), *Tenth Texas Symposium on Relativistic Astrophysics*, Ann. N.Y. Acad. Sci., **375**.
Stecker, F. W., D. L. Morgan, and J. Bredekamp (1971). *Phys. Rev. Lett.*, **27**, 1969.
Stecker, F. W., and J. C. Puget (1972). *Astrophys. J.*, **178**, 57.
Stecker, F. W., and J. L. Puget (1973). In F. W. Stecker and I. J. Trombka (Eds), *Gamma-Ray Astrophysics*.
Steigman, G. (1973). In E. Schatzman (Ed), *Cargese Lectures in Physics*, Vol. 6, Gordon and Breach, New York.
Steigman, G. (1976). *Ann. Rev. Astron. Astrophys.*, **14**, 339.
Steigman, G. (1978). *Astrophys. J.*, **221**, 407.
Stewart, J., and P. Hajicek (1973). *Nature Phys. Sci.*, **244**, 96.
Sunyaev, R. A. (1974). In M. S. Longair (Ed), *Confrontation of Cosmological Theories with Observational Data*, IAU Symposium No. 63, Reidel, Dordrecht.
Sunyaev, R. A., and Ya. B. Zeldovich (1970a). *Astrophys. Sp. Sci.*, **7**, 20.
Sunyaev, R. A., and Ya. B. Zeldovich (1970b). *Astrophys. Sp. Sci.*, **9**, 368.
Swarup, G. (1975). *Mon. Not. R. Astron. Soc.*, **172**, 501.
Synge, J. L. (1960). *Relativity: The General Theory*, North-Holland, Amsterdam.
Tafel, J. (1973). *Phys. Lett. A*, **45**, 341.
Tammann, G. A. (1977). In *Decalages vers le Rouge et Expansion de l'Univers*, IAU Colloquium No. 37, Paris.
Taub, A. H. (1951). *Ann. Math.*, **53**, 472.
Tinsley, B. M. (1974). In L. Motz (Ed), *Seventh Texas Symposium on Relativistic Astrophysics*, Ann. N.Y. Acad. Sci., Vol. 262.
Towe, K. M. (1975). *Nature*, **257**, 115.
Trautman, A. (1972). *Bull. Acad. Pol. Sci., Ser. Sci. Math. Astron. Phys.*, **20**, 185, 503, 895.
Trautman, A. (1973a). *Bull. Acad. Pol. Sci., Ser. Math. Astron. Phys.*, **21**, 345.
Trautman, A. (1973b). *Nature Phys. Sci.*, **242**, 7.
Trombka, I. J., *et al.* (1977). *Astrophys. J.*, **212**, 925.
Van Flandern, T. C. (1975). *Mon. Not. R. Astron. Soc.*, **170**, 333.
Van Flandern, T. C. (1981). *Astrophys. J.*, **248**, 813.
Van Nieuwenhuizen, P. (1981). *Phys. Rep.*, **68**, 191.
Vilenkin, A. (1982). *Phys. Lett.*, **117**, 25.
Vinti, J. P. (1974). *Mon. Not. R. Astron. Soc.*, **169**, 417.
Wagoner, R. V. (1973). *Astrophys. J.*, **179**, 343.
Wagoner, R. V. (1974). In M. S. Longair (Ed), *Confrontation of Cosmological Theories with Observational Data*, IAU Symposium No. 63, Reidel, Dordrecht.
Wagoner, R. W., W. A. Fowler, and F. Hoyle (1967). *Astrophys. J.*, **148**, 3.
Wald, R. W. (1984). *General Relativity*, University of Chicago Press.
Walker, A. G. (1936). *Proc. Lond. Math. Soc.* (2), **42**, 90.



Weiler, K. W., *et al.* (1975). *Phys. Rev. Lett.*, **35**, 134.
Weinberg, S. (1967). *Phys. Rev. Lett.*, **19**, 1264.
Weinberg, S. (1972). *Gravitation and Cosmology*, Wiley, New York.
Weinberg, S. (1979). *Phys. Rev. Lett.*, **43**, 1566.
Wesson, P. S. (1973). *Quart. J. Roy. Astron. Soc.*, **14**, 9.
Wheeler, J. A. (1979). In *Gravitation*, W. H. Freeman & Co.
Wheeler, J. A., and R. P. Feynman (1945). *Rev. Mod. Phys.*, **17**, 157.
Williams, J. G., *et al.* (1976). *Phys. Rev. Lett.*, **36**, 551.
Wolfe, A. M., and G. R. Burbridge (1969). *Astrophys. J.*, **156**, 345.
Woltjer, L. (1967). In H. van Woerden (Ed), *Ratio Astronomy and the Galactic System*, IAU Symposium No. 31, Academic, London.
Woody, D. P., and P. L. Richards (1979). *Phys. Rev. Lett.*, **42**, 295.
Woody, D. P., and P. L. Richards (1981). *Astrophys. J.*, **248**, 18.
Yoshimura, M. (1978). *Phys. Rev. Lett.*, **41**, 281.
Zatsepin, G. T., and V. A. Kuzmin (1966). *Sov. Phys. JETP Lett.*, **4**, 78.