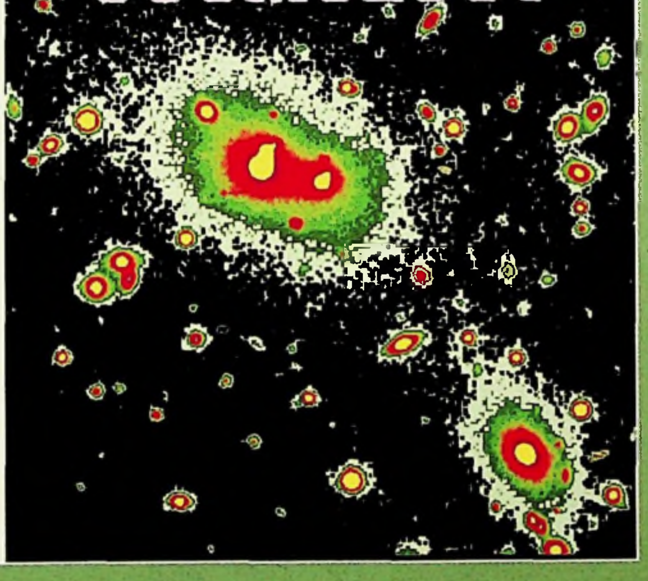
Universities Press

Educational Monographs

Jawaharlal Nehru Centre for
Advanced Scientific Research

ELEMENTS OF COSMOLOGY



JAYANT V NARLIKAR

ELEMENTS OF
COSMOLOGY

The Educational Monographs published by Universities Press in collaboration with JNCASR address the needs of students and the teaching and research community.

Published

Deterministic Chaos: Complex Chance Out of Simple Necessity : N. Kumar

Forthcoming

Dynamic Himalaya : K.S. Valdiya

Cooperation and Conflict in Animal Societies : Raghavendra Gadagkar

Electrostatics of Atoms and Molecules : S.R. Gadre

Titles published earlier by Wiley Eastern

Superconductivity Today : T.V. Ramakrishnan and C.N.R. Rao

Supercomputers : V. Rajaraman

The World of Bohr and Dirac: Images of 20th Century Physics : N. Mukunda

Educational Monographs

ELEMENTS OF COSMOLOGY

Jayant V Narlikar
Inter-University Centre
for Astronomy and Astrophysics
Pune 411 007, India

Jawaharlal Nehru Centre
for Advanced Scientific Research

Universities Press

First published 1996
ISBN 81 7371 043 0

Distributed by
Orient Longman Limited

Registered Office
3-6-272, Himayatnagar, Hyderabad 500 029 (A.P.), India

Other Offices
17, Chittaranjan Avenue, Calcutta 700 072
160, Anna Salai, Madras 600 002
Kamani Marg, Ballard Estate, Mumbai 400 001
1/24, Asaf Ali Road, New Delhi 110 002
80/1, Mahatma Gandhi Road, Bangalore 560 001
365, Shahid Nagar, Bhubaneshwar 751 007
41/316 'Gour Mohan', Ambady Lane, Chittoor Road, Ernakulam 682 011
S.C.Goswami Road, Panbazar, Guwahati 781 001
3-6-272, Himayatnagar, Hyderabad 500 029
28/31, 15 Ashok Marg, Lucknow 226 001
City Centre Ashok, Govind Mitra Road, Patna 800 004

Typeset by
Access Center, Secunderabad 500 003

Printed in India at
Navya Printers, Hyderabad, 500 482

Published by
Universities Press (India) Limited
3-5-820, Hyderguda, Hyderabad 500 029

To

Hermann Bondi

whose classic book *Cosmology*
introduced me to the subject

Contents

Foreword

The Jawaharlal Nehru Centre for Advanced Scientific Research was established by the Government of India in 1989 as part of the centenary celebrations of Pandit Jawaharlal Nehru. Located in Bangalore, it functions in close academic collaboration with the Indian Institute of Science.

The Centre functions as an autonomous institution devoted to advanced scientific research. It promotes programmes in chosen frontier areas of science and engineering and supports workshops and symposia in these areas. It also has programmes to encourage young talent.

In addition to the above activities, the Centre has undertaken a programme of publishing high quality Educational Monographs written by leading scientists and engineers in the country. These are short accounts of interesting areas in science and engineering addressed to students at the graduate and postgraduate levels, and the general research community.

This monograph is one of the series being brought out as part of the publication activities of the Centre. The Centre pays due attention to the choice of authors and subjects and style of presentation, to make these monographs attractive, interesting and useful to students as well as teachers. It is our hope that these publications will be received well both within and outside India.

C.N.R. Rao
President

Preface

This monograph is based on lectures given by me at university campuses with the aim of introducing the subject of cosmology to students and teachers at the graduate level. Since unfamiliarity with the general theory of relativity hampers the understanding of basic cosmological issues, I used the rather unusual method of discussing the subject within the framework of Newtonian gravity and mechanics. It turns out that Newtonian cosmology can capture essentially all the salient features of relativistic cosmology with the added advantage that it is more readily understood by the students of physics or mathematics at the undergraduate level. For the same reason, I have avoided going into details of statistical mechanics, nuclear physics, particle physics, etc.

The description is up-to-date at the time of writing and includes cosmological models, their physical properties and observational tests. It is hoped that the student or teacher will be sufficiently enthused by the account to tackle the more fuller version provided by relativistic cosmology. A list of references is provided at the end.

Some astronomical details assumed in the text are given in the Appendix in the form of tables at the end. The aim is to make the monograph complete in itself.

I wish to thank the Jawaharlal Nehru Centre for Advanced Scientific Research for including it in the list of its publications.

Jayant V. Narlikar
Pune, 1996

1 *The Large Scale Structure of the Universe*

1.1 Introduction

The universe, by definition, includes everything that is in existence. Man's perception of it has been limited by his ability to observe and interpret what he sees. Inevitably, ever since his quest for understanding the structure of the universe began, the horizons of his perception of the universe have further receded. Table 1.1 illustrates how this has happened over the last two millennia.

Table 1.1 A historical look at the expanding horizons of human understanding of the universe

Period	State of knowledge
1^{st}–15^{th} century	The geocentric theory prevailed, with the Earth fixed at the centre of the universe.
16^{th} century–mid 17^{th} century	Gradual acceptance of the Copernican heliocentric theory.
Late 17^{th} century–mid 19^{th} century	Newtonian law of gravitation well established and used for studies of the solar system.
19^{th} century–early 20^{th} century	Solar system believed to be at the centre of our galaxy. Also all faint nebulae believed to be within our galaxy.
1920s	Both the above beliefs shown to be false. The solar system is now known to be about 10 kpc from the galactic centre; also faint nebulae, now known to be galaxies in their own right, located far away from ours.
1917	A. Einstein proposes a static model of the universe, the first general relativistic model in cosmology. W. de Sitter follows with a model of an expanding but empty universe.

contd.

Table 1.1 *contd.*

Period	State of knowledge
1922–24	A. Friedmann discovers solutions of Einstein's equations describing an expanding universe.
1929	E.P. Hubble discovers a linear relation between the redshift and distance of a typical galaxy. The relation is now understood within the framework of the expanding universe.
1946–50	George Gamow and his colleagues R.A. Alpher and R.C. Herman work out the theory of nucleosynthesis in the early stages of the hot big bang universe.
1948	H. Bondi, T. Gold and F. Hoyle propose the model of the steady state universe.
Late 1950s– early 1960s	First indications of second order clustering of galaxies. Majority of astronomers, however, do not take it seriously and believe the universe to be homogeneous on scales larger than the typical intergalactic distance $\sim$ 1 Mpc.
1965	Discovery of the microwave background radiation by A.A. Penzias and R.W. Wilson. This discovery lends credibility to the hot big bang model.
1967	Extensive calculations of big bang nucleosynthesis by R.V. Wagoner, W.A. Fowler and F. Hoyle to demonstrate that light nuclei in appropriate quantities can be made.
Mid-1960s–1980	Attempts to form large scale structures in the big bang scenario lead to expected fluctuations of microwave background that are far above the limits placed by observations.
1981– present	A. Guth proposes the inflationary model as a consequence of very high energy physical interactions affecting the expansion rate of the very early universe. Also particle physicists propose various kinds of non-baryonic matter as candidates for dark matter suspected by astronomers to exist inside galaxies and also in the intergalactic space. Efforts are on to understand how visible matter formed into various structures at different scales from dwarf galaxies to superclusters.
1992	Discovery of fluctuations of microwave background by COBE satellite.

In this book we will present the state of the art in theory and observations in cosmology. We will deal qualitatively with the very technical aspects but build a quantitative base using Newtonian gravitation and dynamics. The reader wishing to read more advanced texts will find this elementary treatment helpful as a starting point.

Before proceeding further, however, it will be useful to define a few units and quantities often used in astronomy and cosmology. We begin with the units of length, mass and time. For expressing large numbers we will use million (10^6), billion (10^9), etc. Large units of a quantity will be expressed by kilo(10^3), mega(10^6), giga(10^9), etc., while small fractions by milli(10^{-3}), micro(10^{-6}), etc.

Length : The c.g.s. unit of length is centimetre (cm). For astronomical distances *light year* (distance travelled by light in vacuum in one year) is more useful. A simple calculation gives

$$1 \text{ light year (ly)} = 9.4605 \times 10^{17} \text{ cm}.$$

However, the astronomer prefers to use the distance unit of *parsec* which is the distance at which half the diameter of the Earth's orbit around the Sun subtends an angle of 1 arc second. Measurements give

$$1 \text{ parsec (pc)} = 3.0856 \times 10^{18} \text{ cm} \cong 3.26 \text{ ly}.$$

It is customary to use kiloparsec (kpc $\equiv 10^3$ pc), megaparsec (Mpc $\equiv 10^6$ pc) and gigaparsec (Gpc $\equiv 10^9$ pc) as appropriate.

Mass : The c.g.s. unit of mass is gram (g); but it is too small for astronomical masses. More convenient is one solar mass unit denoted by $M_\odot$. Astronomical estimates give

$$M_\odot = 1.989 \times 10^{33} \text{ g}.$$

Time : The c.g.s. unit of time, second (s), is used in astronomy, although for longer time scales, year (yr) or even giga year (Gyr $\equiv 10^9$ yr) are used. Remember that

$$1 \text{ yr} \cong 3 \times 10^7 \text{ s}.$$

Magnitude : This term denotes a measure of light received from an astronomical object. If the source of light has luminosity L, that is, if it is emitting L units of energy per unit of time, then the flux crossing unit area held normal to the direction of propagation at a distance D is

$$l = \frac{L}{4\pi D^2}. \tag{1.1}$$

The *apparent magnitude* of the source is then defined by

$$m = -2.5 \log l + \text{ constant.} \quad (1.2)$$

The constant is so adjusted that $m = 0$ corresponds to a flux $l = l_0 \equiv 2.48 \times 10^{-5}$ erg cm^{-2} s^{-1}.

The *absolute magnitude* M of the source is defined as its apparent magnitude at a distance $D = 10$ pc. From (1.2) we therefore have $M = -2.5 \log L +$ constant, and

$$m = M + 5 \log D_{\text{pc}} - 5. \quad (1.3)$$

Here D_{pc} = distance of the source measured in parsecs. Thus if we are looking at a cluster of galaxies of approximately the same L, then m for each gives an estimate of its distance on a logarithmic scale.

1.2 Structural hierarchy

The cosmologist is concerned only with the large scale structure of the universe. So, in the first approximation he concentrates his attention on structures larger than a given size. That 'starting point' happens to be 'galaxy'. For example, our Milky Way has a mass $\sim 2 \times 10^{11} M_\odot$, a disc shape with a visible diameter $\sim$ 30 kpc and has more than 10^{11} stars. Yet, we may treat the Milky Way as a point for the purpose of cosmology.

This approximation is justified when we consider the larger cosmological scales. These are given in the following sequence :

galaxy $\rightarrow$ group $\rightarrow$ cluster $\rightarrow$ supercluster $\rightarrow$ Hubble radius

A group may consist of 10–50 galaxies. The Milky Way belongs to the Local Group (L.G.) of some 20 galaxies. The L.G. has two dominant members, our galaxy and Andromeda. The rest are smaller galaxies.

A cluster may have upto $\sim$ 1000 galaxies with a mass $\sim 10^{14} M_\odot$, and linear extent $\sim$ 5 Mpc. A supercluster may be ten times as massive, extending upto 50 Mpc in diameter.

Structures of even larger scales are also suspected to exist. The so called 'Great Wall' extends as a linear structure of size 60 Mpc × 157 Mpc and is made up of several superclusters. The filamentary nature of matter distribution is shown in Fig.1.1. Concentrations along filaments are contrasted with giant 'voids' which apparently contain very few galaxies. These voids may easily extend to $\sim$ 100 Mpc in size.

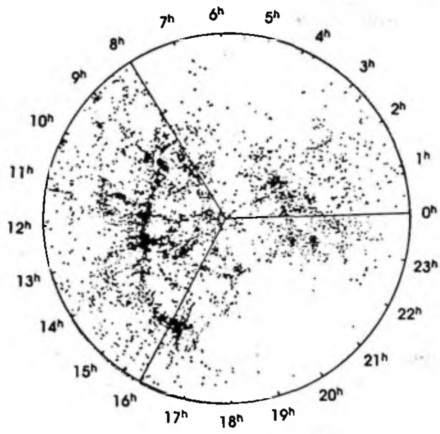


Fig.1.1 A large scale structure has a filamentary distribution of galaxies; the one shown here is called the 'Great Wall'. Picture adapted from M.J. Geller and H.P. Huchra, *Science*, **246**, 897 (1989).

Indirect studies of galaxies and clusters have further pointed to the possible existence of *dark matter* which, as its name implies, does not radiate electromagnetic waves. Its existence can be inferred by its gravitational pull on visible matter. Estimates based on such observations suggest that the density of dark matter may far exceed the density of visible matter described above.

Going to the maximum length scale in the above sequence, we arrive at the 'Hubble radius', which we will discuss in the following section. It also happens to be the limit of what is currently observable — a range of ~ 3000 Mpc. Thus, compared to the largest scale, our galaxy size of 30 kpc is a very tiny fraction of $\sim 10^{-5}$ (a comparison may be made between the size of a pea and the distance of 1 kilometre).

1.3 Hubble's law

What is the large scale dynamics of galaxies like? There are two ways in which an astronomer measures motion. Let us discuss his limitations in this field.

a) Transverse Motion : Suppose we have a galaxy moving with a speed v perpendicular to the line of sight. If the galaxy is at a distance r, its angular position will appear to change over time T by

$$\theta = \frac{vT}{r}. \tag{1.4}$$

Expressing r in Mpc, T in years and v in kms^{-1} we get θ in arc seconds as

$$\begin{aligned} \theta &= 2 \times 10^5 \times \frac{v_{\mathrm{kms}^{-1}} \times 3 \times 10^7 T_{\mathrm{yr}}}{r_{\mathrm{Mpc}} \times 3 \times 10^{19}} \\ &= 2 \times 10^{-7} \times \frac{v_{\mathrm{kms}^{-1}} \times T_{\mathrm{yr}}}{r_{\mathrm{Mpc}}}. \end{aligned} \tag{1.5}$$

Thus, even over an observation of 30 yrs, a transverse speed of $\sim$ 300 kms^{-1} at a distance of 1 Mpc will produce an angular displacement $< 2 \times 10^{-3}$ arc sec. Current angular resolutions in optical astronomy do not go beyond 10^{-1} arc sec. Thus detection of a transverse shift is beyond its capabilities. In radio astronomy, however, the very long baseline interferometry (VLBI) is capable of milli-arc second resolutions. Thus one can hope to make progress in this direction.

b) Radial Motion : The Doppler effect has been a useful tool for measuring radial motions in astronomy. If the spectrum of a galaxy shows a well identified line at wavelength λ instead of its laboratory measured wavelength λ_0, the line is said to be spectral shifted by a fraction z of the original wavelength, i.e.,

$$z = \frac{\lambda - \lambda_0}{\lambda_0}. \tag{1.6}$$

The line is said to be *redshifted* if $z > 0(\lambda > \lambda_0)$ and *blueshifted* if $z < 0(\lambda < \lambda_0)$.

According to the Newtonian Doppler effect formula, z is related to the radial velocity of recession v by

$$z = \frac{v}{c}, \tag{1.7}$$

c being the speed of light. With modern techniques of spectroscopy, z as small as 10^{-6} can be measured. That is, radial velocities of the order of $\sim$ 10 kms^{-1} can be measured.

Although stars in our galaxy show both redshifts and blueshifts, the population of galaxies show (with very few exceptions) redshifts only. This trend began to be noticed in the early observations of Slipher during 1912–25, and became well established thanks to the extensive work of E.P. Hubble and M. Humason.

In 1929, Hubble found a linear relation between the measured redshift z and distance D of a typical galaxy. The relation is expressed in the form

$$z = \frac{H_0}{c} D, \tag{1.8}$$

where H_0 is now known as Hubble's constant. Hubble estimated the value of this constant as 530 $\mathrm{kms^{-1}\ Mpc^{-1}}$. The revised measurements today give

$$H_0 = 100 h_0 \ \mathrm{kms^{-1}\ Mpc^{-1}}, \qquad 0.5 \leq h_0 \leq 1. \tag{1.9}$$

The range of uncertainty still persisting is indicated by the parameter h. There are several observational issues still to be resolved before one can claim to know the 'true' value of H. Hubble's plot of 1929 is reproduced in Fig.1.2.

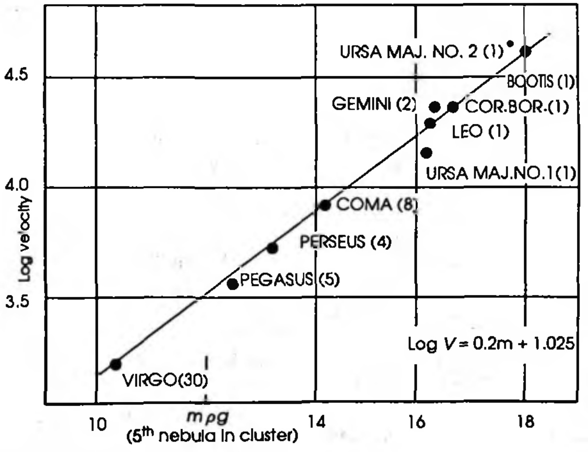


Fig.1.2 Hubble's plot of radial velocity (on the ordinate, on a logarithmic scale) against apparent magnitude of the galaxy. Since apparent magnitude measures distance on a logarithmic scale, the plot is an expression of the linear relation equation (1.8).

If, however, z is large enough to be close to unity, formula (1.7) is modified to the special relativistic result

$$1+z=\sqrt{\frac{1+v/c}{1-v/c}}. \tag{1.10}$$

Thus $z=2$ implies $v=0.8c$.

In the next chapter we will give a separate 'cosmological' interpretation to the redshift. It may well happen that Hubble's constant changes with epoch and may not have the same value H_0 (measured today) at other epochs. c/H_0 is called the Hubble radius at the present epoch.

However, taken at its face value the result implies that the entire population of galaxies is receding from our own. Does it mean that we are located in a special position? The answer turns out to be in the negative. As we shall see in the following chapter, the real interpretation is quite different. Instead of being in a special location we are typical members of a general class of observers.

1.4 Radiation backgrounds

Galaxies, clusters, etc. are manifestations of matter. The universe also contains radiation backgrounds of different wavelengths. Table 1.2 gives the distribution of different types of radiation.

Table 1.2 Radiation backgrounds at different wavelengths

Type of radiation	Wavelength λ Frequency ν Energy range E	Energy density (erg cm^{-3})
Radio	$\nu \leq 4080$ MHZ	$\leq 10^{-18}$
Microwaves	λ: 80 cm–1 mm	$\sim 4 \times 10^{-13}$
Optical	λ: 4000 Å–8000 Å	$\sim 3.5 \times 10^{-15}$
X-rays	E: 1–40 keV	$\sim 10^{-16}$
γ-rays	$E \geq 100$ MeV	$\leq 2 \times 10^{-17}$

We see that the radiation energy density is maximum in the microwave band. However, even taking the largest component at 4×10^{-13} erg cm^{-3}, it corresponds (via Einstein's mass–energy equivalence, $E = Mc^2$) to a matter density of $\sim 4 \times 10^{-34}$ g cm^{-3}. By way of comparison, the density of visible matter in the universe is $\sim 3 \times 10^{-31}$ g cm^{-3}, i.e., higher by some three orders of magnitude. This fact is often expressed through the remark that the universe at present is *matter dominated.*

In Chapter 2 we will proceed with our model building on the assumption that the effect of radiation may be neglected in comparison with matter. Later we will take note of radiation also and see how the microwave background observed today can be linked to the primordial history of the universe.

1.5 Exercises

1.1 Calculate the absolute and apparent magnitudes of the Sun, given that its luminosity $L_\odot = 4 \times 10^{33}$ $\mathrm{erg s^{-1}}$ and its distance from the Earth is 1.5×10^8 km.

1.2 A star is inclined at 60° to the plane of the Earth's orbit, assumed circular, and the maximum angle subtended by an orbital diameter at the star is 0.5 arc sec. What is the minimum angle subtended at the star by an orbital diameter? Estimate also the star's distance in parsecs.

1.3 The spectral line $H\alpha$ of a galaxy satisfying Hubble's law has an observed wavelength of 7218Å. The laboratory wavelength of the $H\alpha$ line is 5662Å. If the galaxy is at 400 Mpc, calculate Hubble's constant.

1.4 A galaxy has an apparent magnitude 18 and absolute magnitude −17. Estimate its distance.

1.5 Show that in terms of apparent magnitude m and redshift z, Hubble's law takes the form :

$$m = M + 5 \ \log z + \text{constant}.$$

2 *Newtonian Cosmology: Theoretical Models*

2.1 Introduction

Given the large scale structure on the lines described in Chapter 1, how do we go about modelling the universe? What physical interactions determine the dynamics of galaxies, clusters and superclusters? What is the framework in which the Hubble law emerges naturally? To answer these questions we first note the following four basic interactions of physics :

(i) Gravitational interaction
(ii) Electromagnetic interaction
(iii) Weak interaction
(iv) Strong interaction

Only (i) and (ii) above are of long enough range to be of relevance to cosmology. Of these again, (ii) is not likely to be important because galaxies and intergalactic matter are electrically neutral (at least there is no evidence to the contrary). Nor are there any large scale electric currents. Thus we are left only with (i).

Of the two popular theories of gravitation — by Newton and Einstein — the former is simpler but with the conceptual defect that it is inconsistent with special relativity. The fact that Newtonian gravity is instantaneous in its action causes no serious embarrassment for interactions within the galaxy. On the universal level, where very large distances ($\gtrsim 10^9$ light years) are involved, the use of Newtonian gravity is suspect. Likewise the Newtonian Doppler shift formula (1.7) becomes suspect for $z \gtrsim 1$.

For these reasons, cosmologists have preferred using relativity as the basis of cosmology. Indeed, pioneering work in theoretical cosmology by Einstein, de Sitter, Friedmann, Lemaitre, Eddington, etc. was done within the relativistic framework. However, the level at which this text is aimed precludes the use of general relativity. We will therefore revert to Newtonian gravity on grounds of simplicity. Moreover, in 1935, E.A. Milne and W.H. McCrea showed that with suitable reinterpretation, Newtonian gravity does yield models very similar to those of relativistic cosmology. We will follow the treatment of Milne and McCrea in this chapter.

2.2 Simplifying postulates

We shall use two postulates to simplify the above model construction. The first is known as the Weyl postulate and the second, the cosmological principle.

(i) The Weyl postulate : Proposed by Hermann Weyl in the early days of relativistic cosmology, this postulate states that the trajectories of a special class of observers, to be identified with galaxies (now treated as points!), form a bundle of non-intersecting lines in space–time so that there is a unique line passing through each point in space at any given time.

Figure 2.1 illustrates the special kind of motion implied by Weyl's postulate. In the space–time diagram shown in Fig.2.1(b), we see the trajectories distributed in a streamlined fashion. No two members intersect. Thus there is a unique member of the set passing through any given point in space–time. In Fig.2.1(a) on the other hand, the trajectories are in disorder with intersections permitted. In this case it is not possible to identify a unique trajectory through each point. Galactic motion approximates to the idealized case of Fig.2.1(b). We may identify a unique observer for each galaxy. Such observers are called *fundamental observers.*

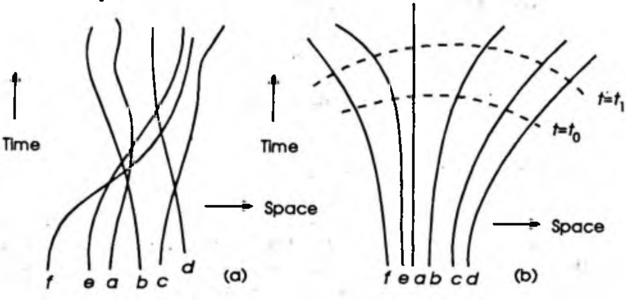


Fig.2.1 In (a) the motion is chaotic with intersecting trajectories indicating collisions. Here we cannot identify any cosmic epochs. In (b) we see streamlined motion with non-intersecting trajectories as per the Weyl postulate. This enables identification of 'epochs' $t = t_0, t_1$, etc. such that the universe is homogeneous and isotropic at any given epoch. These epochs are called 'cosmic time'.

Thus, we may have a continuum of such trajectories of fundamental observers given in the space–time plot with cartesian coordinates $(\mathbf{r}, t)$ as

$$\mathbf{r} = \mathbf{F}(t, \mathbf{r}_0). \tag{2.1}$$

That is, at any given epoch t, a galaxy identified by the triplet of coordinates $\mathbf{r}_0$ is at $\mathbf{r}$ given by (2.1). The vector function $\mathbf{F}$ is still to be determined, but it satisfies the non-intersection condition, i.e.,

$$\mathbf{F}(t, \mathbf{r}_0) = \mathbf{F}(t, \mathbf{r}_0') \Rightarrow \mathbf{r}_0 = \mathbf{r}_0'. \tag{2.2}$$

(ii) The cosmological principle : This principle states that at any epoch t, the universe is homogeneous and isotropic. That is, given any position in the universe and any direction in which it is viewed from that position, the large scale aspect of the universe is the same for all fundamental observers. Let us explore one immediate consequence of this principle. At any position $\mathbf{r}$, the fundamental observer located there moves with a definite velocity given by

$$\mathbf{v} = \left.\frac{d\mathbf{r}}{dt}\right|_{\mathbf{r}_0} = \frac{\partial F(t, \mathbf{r}_0)}{\partial t} \equiv \mathbf{G}(t, \mathbf{r}), \quad \text{say.} \tag{2.3}$$

At any epoch, $\mathbf{v}$ can be a function of $\mathbf{r}$ only because, at each point of space there is a unique fundamental observer. Now imagine three observers at $\mathbf{r}_1$, $\mathbf{r}_2$ and at $\mathbf{r} = 0$. The observer at $\mathbf{r} = 0$ finds that the velocities of the first two observers are

$$\mathbf{v}_1 = \mathbf{G}(t, \mathbf{r}_1), \quad \mathbf{v}_2 = \mathbf{G}(t, \mathbf{r}_2). \tag{2.4}$$

Hence, viewed by the first of these observers, the second has the velocity

$$\mathbf{v}_2 - \mathbf{v}_1 = \mathbf{G}(t, \mathbf{r}_2) - \mathbf{G}(t, \mathbf{r}_1) \tag{2.5}$$

with respect to him. However, by the cosmological principle, the observer at $\mathbf{r} = 0$ has no special status. Thus seen by the observer at $\mathbf{r}_1$, the velocity of the second observer should be the *same function* of their relative vector $(\mathbf{r}_2 - \mathbf{r}_1)$ as in (2.1). That is,

$$\mathbf{v}_2 - \mathbf{v}_1 = \mathbf{G}(t, \mathbf{r}_2 - \mathbf{r}_1). \tag{2.6}$$

Combining (2.5) and (2.6) we get

$$\mathbf{G}(t, \mathbf{r}_2 - \mathbf{r}_1) = \mathbf{G}(t, \mathbf{r}_2) - \mathbf{G}(t, \mathbf{r}_1). \tag{2.7}$$

Problem 2.1 : If a function $G(x)$ satisfies the relation

$$G(x+y) = G(x) + G(y),$$

show that the most general form for G is $G(x) = Ax$, A = constant.

Solution : Differentiate the above defining relation partially with respect to x and y to get

$$G'(x+y) = G'(x) = G'(y).$$

Since $G'(x)$ is a function of x only and $G'(y)$ is a function of y, the two can be equal only if they are constants. Hence the result follows.

From Exercise 2.1, we see that the most general form for $\mathbf{G}(t, \mathbf{r})$ is given by the tensor relation

$$G_\mu(t, \mathbf{r}) = \sum_\nu A_{\mu\nu} r_\nu \quad ; \quad \lambda, \mu = 1, 2, 3 \tag{2.8}$$

where $\mathbf{r} = (r_\mu)$ is the triplet of Cartesian coordinates describing the position vector of a typical fundamental observer. The magnitude of $\mathbf{r}$ will be denoted by r. The tensor $A_{\mu\nu}$ is of second rank, and it depends on t only. Since the universe looks isotropic from any point, $A_{\mu\nu}$ cannot have any fundamental direction associated with it. It can therefore only have the isotropic form

$$A_{\mu\nu} = H(t)\delta_{\mu\nu} \tag{2.9}$$

where $H(t)$ is so far an undetermined function of t. From (2.3) we therefore get

$$\mathbf{v} = H(t)\mathbf{r}. \tag{2.10}$$

This is nothing but the velocity–distance relation obtained by Hubble! Thus Hubble's law is consistent with our postulate of homogeneity and isotropy : we do not enjoy any 'special status' by being at $\mathbf{r} = 0$, say.

We can use (2.10) to complete the integration of the differential equation (2.3) by writing

$$\mathbf{r} = S(t)\mathbf{r}_0, \tag{2.11}$$

with

$$\frac{\dot{S}}{S} = H(t). \tag{2.12}$$

The overhead dot differentiates the quantity with respect to t. We will denote the present epoch by t_0 and write $H_0 = H(t_0)$.

The factor S is often called the *scale factor* as it scales the distances with epoch. Imagine a triangle with vertex coordinates $\mathbf{a}_0$, $\mathbf{b}_0$, $\mathbf{c}_0$ at t_0. At any other epoch t, we may take these coordinates as $S(t)\mathbf{a}_0$, $S(t)\mathbf{b}_0$ and $S(t)\mathbf{c}_0$ with $S(t_0) = 1$. If $S(t)$ increases with t, our triangle is expanding. As we shall discover shortly, this happens to be the situation.

2.3 Redshift

To relate the velocity–distance relation to redshift we need to do some more work, however. Consider a galaxy at $\mathbf{r}_0 = \mathbf{a}_0$ from which we receive light. Now we will work out the propagation of light from $\mathbf{a}_0$ to ourselves by using the assumption that the velocity of light is c for every fundamental observer. Let light leave $\mathbf{r}_0 = \mathbf{a}_0$ at $t = t_a$ and reach $\mathbf{r}_0 = 0$ at $t = t_0$.

Thus the light ray propagating from $\mathbf{r}_0 = \mathbf{a}_0$ to $\mathbf{r}_0 = 0$ will pass intermediate observers at $\mathbf{r}_0 = \lambda\mathbf{a}_0$, $0 < \lambda < 1$, at time t in the range $t_a < t < t_0$. Since the velocity of such a typical en-route observer is $\mathbf{r}_0\dot{S}(t)$ away from us, the light has an effective velocity

$$\frac{dr}{dt} = -c + \lambda a_0 \dot{S}(t), \tag{2.13}$$

where $\mathbf{r} = \mathbf{r}_0 S(t) = \lambda a_0 S(t)$.

Caution : This addition of the velocity of light to the velocity of the medium is a purely Newtonian concept and it is *known to be inconsistent with facts:* one should really follow the theory of relativity. We are, however, justified in proceeding in the above manner as we are working entirely in the Newtonian framework.

Since $dr/dt = \dot{\lambda} a_0 S + \lambda a_0 \dot{S}$, we get

$$\frac{d\lambda}{dt} = -\frac{c}{a_0 S},$$

$$\text{i.e.,} \quad a_0 = \int_{t_a}^{t_0} \frac{c\,dt}{S(t)}, \tag{2.14}$$

since at $t = t_a$, $\lambda = 1$ and at $t = t_0$, $\lambda = 0$.

In deriving (2.14) we have added the velocity of light to the velocity of the intermediate observer as per the Newtonian formula for vectorial addition of velocities. Although our operation is inconsistent with special relativity, it is fully consistent within the Newtonian framework.

Consider now two light-wave crests of wavelengths λ_0 emitted by the above galaxy. The first crest leaves at t_a and arrives at t_0. The second one leaves at $t_a + \Delta t_a$, say, and arrives at $t_0 + \Delta t_0$ for which a relation similar to (2.14) holds, i.e.,

$$a_0 = \int_{t_a + \Delta t_a}^{t_0 + \Delta t_0} \frac{cdt}{S(t)}. \tag{2.15}$$

Subtracting (2.14) from (2.15) and using the approximation that Δt_0 and Δt_a are small enough intervals for treating $S(t)$ constant over them, we get

$$\frac{c\Delta t_0}{S(t_0)} = \frac{c\Delta t_a}{S(t_a)}. \tag{2.16}$$

But if λ is the wavelength received by us, then $c\Delta t_0 = \lambda$ while $c\Delta t_a = \lambda_0$. Therefore

$$1 + z \equiv \frac{\lambda}{\lambda_0} = \frac{S(t_0)}{S(t_a)}. \tag{2.17}$$

This is the relationship between z, the redshift and the scale factor $S(t)$. Since $z > 0$, $S(t_a) < S(t_0)$ for $t_a < t_0$. In other words, the scale factor increases with time, implying that the universe is expanding.

Problem 2.2 : Deduce the linear redshift–distance relation for small distances from the above derivation of redshift, and show that Hubble's constant is given by $\dot{S}/S$, evaluated at $t = t_0$.

Solution : For small distances $t_a \approx t_0$ and a Taylor expansion near $t = t_0$ gives

$$S(t_a) \cong S(t_0) - (t_0 - t_a)\dot{S}(t_0).$$

Hence,

$$1 + z = \frac{S(t_0)}{S(t_a)} = \left\{1 - (t_0 - t_a)\frac{\dot{S}(t_0)}{S(t_0)}\right\}^{-1} \approx 1 + (t_0 - t_a)\frac{\dot{S}}{S}\bigg|_{t_0}. \tag{A}$$

But, from (2.14), under the same approximation,

$$a_0 \approx \frac{c(t_0 - t_a)}{S(t_0)}. \tag{B}$$

The distance of the galaxy at $t = t_0$ is $D = aS(t_0) \cong c(t_0 - t_a)$. From (A) and (B) the result follows.

Since the coordinates $\mathbf{r}_0$ are fixed with respect to time for all galaxies, they are *comoving coordinates* and their framework describes what is often called the *cosmological rest frame*. For $S(t)$ increasing with t, the proper size of the frame increases and we describe the phenomenon as the expanding universe.

2.4 Luminosity distance

The astronomer estimates the distance of a star by the inverse square law of illumination. Thus if L is the luminosity of the star, the flux of radiation received per unit time by an observer at a distance D from the star is

$$l = \frac{L}{4\pi D^2}. \tag{2.18}$$

If L and l are known, D can be estimated and is called the luminosity distance. The cosmologist uses a similar formula but he has to modify it to include redshift. In essence, two additional factors $(1+z)$ appear in the denominator so that (2.18) changes to

$$l = \frac{L}{4\pi D^2(1+z)^2}. \tag{2.19}$$

These factors arise for the following reasons:

First, note that because of (2.16), radiation emitted in time interval Δt_a by the source is received by the observer over a time interval $\Delta t_0 = (1 + z)\Delta t_a$. Thus the 'rate' of receiving radiation is reduced by the factor $(1 + z)$. Secondly, while the number of light quanta emitted over Δt_a and received over Δt_0 remain the same, each quantum is reduced in energy by the same factor $(1+z)$ because its frequency is similarly reduced. Equation (2.19) gives us the *luminosity distance* as

$$D_L = D(1 + z) \tag{2.20}$$

where $D = a_0 S(t_0)$ is the geometrical distance. If we use (2.20), we get back (2.18) with D_L replacing D :

$$l = \frac{L}{4\pi D_L^2}. \tag{2.21}$$

The cosmologist has to exercise some caution in the use of (2.21). In practice, he does not measure the flux at *all* wavelengths; his observations are confined to a limited wavelength range, $[\lambda_1, \lambda_2]$, say. For an object of redshift z, this range corresponds to $[\lambda_1/(1+z), \lambda_2/(1+z)]$ at the source. Thus if the object has a spectral function $I(\lambda)$ with the total luminosity over this range given by

$$\int\limits_{\lambda_1/(1+z)}^{\lambda_2/(1+z)} I(\lambda)d\lambda = L\left[\frac{\lambda_1}{1+z}, \frac{\lambda_2}{1+z}\right], \tag{2.22}$$

say, then the flux received at the observer is given by

$$\begin{aligned} l[\lambda_1, \lambda_2] &= \frac{1}{4\pi D_L^2} \int\limits_{\lambda_1/(1+z)}^{\lambda_2/(1+z)} I(\lambda)d\lambda \\ &= \frac{1}{4\pi D_L^2(1+z)} \int\limits_{\lambda_1}^{\lambda_2} I(\lambda/(1+z)d\lambda. \end{aligned} \tag{2.23}$$

Thus at a redshift of $z = 2$, an observed wavelength range of 4500 Å–6000 Å requires knowledge of $I(\lambda)$ over the range 1500 Å–2000 Å, i.e., a visual range at the observer gets converted to the ultraviolet range at the source. Wrong inferences will be drawn if the observer does not include this effect, commonly called the *K-correction.*

Problem 2.3 : For $I(\lambda) \propto \lambda^2$, find the K-correction.

Solution : Let $I(\lambda) = \beta\lambda^2, \beta =$ constant. Then

$$\begin{aligned} l[\lambda_1, \lambda_2] &= \frac{1}{4\pi D_L^2(1+z)} \int\limits_{\lambda_1}^{\lambda_2} \frac{\beta\lambda^2}{(1+z)^2} d\lambda \\ &= \frac{\beta(\lambda_2^3 - \lambda_1^3)}{12\pi D_L^2(1+z)^3}. \end{aligned}$$

If the observer had not included the effect of redshift on the spectral function, he would have missed the extra $(1+z)^3$ factor in the denominator.

2.5 Cosmological models

We now introduce dynamics into our framework to calculate the form of $S(t)$. The first and simplest class of models involves 'dust' as the main component of the universe. By dust we mean pressureless fluid, an idealization implying that the large scale motion of galaxies has no random component built into it. Thus we have a typical fluid element containing density ρ of matter with a bulk velocity $\mathbf{v}$, given by the Hubble law

$$\mathbf{v} = H(t)\mathbf{r}, \qquad H(t) = \frac{\dot{S}}{S}. \tag{2.24}$$

The continuity equation of fluid mechanics then gives

$$\frac{\partial \rho}{\partial t} + \text{div}\,(\rho \mathbf{v}) = 0.$$

But, from (2.24), div $\mathbf{v} = 3H(t)$ while $\nabla \rho = 0$; which leads to

$$\frac{\partial \rho}{\partial t} + 3\frac{\dot{S}}{S}\rho = 0,$$

i.e., $\rho S^3 = \text{constant} = \rho_0 S_0^3$ (say). (2.25)

This is the density dilution during adiabatic expansion. Next we consider the Euler equations for fluid dynamics :

$$\rho\left[\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}.\nabla)\mathbf{v}\right] = -\nabla p + \rho \mathbf{F} \tag{2.26}$$

where p is the pressure and $\mathbf{F}$ is the external force per unit mass on the fluid element. In our case it is gravitational and satisfies the relation

$$\nabla.\mathbf{F} = -4\pi G\rho. \tag{2.27}$$

Substituting (2.24) in (2.26) with $p = 0$, we get

$$\{\dot{H}\mathbf{r} + H^2\mathbf{r}\} = \mathbf{F}. \tag{2.28}$$

Class I ($k = 0$): Here (2.31) has a simple power-law solution

$$S(t) = \left(\frac{t}{t_0}\right)^{2/3} S_0, \tag{2.32}$$

with the Hubble constant given by

$$H(t) = \frac{\dot{S}(t)}{S(t)} = \frac{2}{3t}, \qquad H_0 = \frac{2}{3t_0} \tag{2.33}$$

and the matter density by

$$\rho = \frac{3H^2}{8\pi G} \equiv \rho_c. \tag{2.34}$$

For reasons to be discussed later, ρ_c is called the *critical density*. This model was jointly advocated by Einstein and de Sitter in 1932, and is called the *Einstein–de Sitter model.*

Class II ($k > 0$): In this case $S(t)$ has a maximum value given by

$$S_{\max} = \frac{8\pi G \rho_0 S_0^3}{3kc^2}. \tag{2.35}$$

The universe thereafter contracts. The density at any epoch is given by

$$\rho = \frac{\rho_0 S_0^3}{S^3} = \frac{3}{8\pi G}\left(\frac{\dot{S}^2}{S^2} + \frac{kc^2}{S^2}\right) = \Omega \rho_c, \tag{2.36}$$

where

$$\Omega = 1 + \frac{kc^2}{\dot{S}^2} > 1. \tag{2.37}$$

Thus the *density parameter* Ω exceeds unity for models of this type. We may also introduce the so-called *deceleration parameter* q defined by

$$q = -\frac{1}{H^2}\frac{\ddot{S}}{S}. \tag{2.38}$$

From (2.30) we see that

$$q = \frac{1}{H^2}\cdot\frac{4\pi G \rho_0 S_0^3}{3S^3} = \frac{4\pi G \rho}{3H^2} = \frac{1}{2}\Omega \tag{2.39}$$

These definitions can be applied to cosmology of *Class I*, giving $\Omega = 1$, $q = 1/2$. For other models, Ω and q are time-dependent and we will denote them by Ω_0 and q_0, the present values of these parameters.

Class III ($k < 0$): In this case, as for *Class I*, S steadily increases from zero. We also have (2.37) and (2.39) holding here but now $\Omega < 1$ and $q < 1/2$.

If $\rho = 0$ ($\Omega = 0$), we have a linearly expanding model

$$S = S_0 . \frac{t}{t_0}. \tag{2.40}$$

E.A. Milne had arrived at this model in his *kinematic relativity*. Hence it is sometimes called the *Milne model*.

It is clear from the above discussion that the critical density ($\Omega = 1$) separates the ever expanding models ($k < 0$) from those where the expansion eventually stops and gives way to contraction ($k > 0$). Hence, the $\rho = \rho_c$ and $\Omega = 1$ case is critical in a dynamical sense. In relativistic cosmology we get exactly similar dynamical behaviour, but there k has a further significance in terms of the geometry of the space given by $t =$ constant. For $k > 0$ the space is closed (the surface of a four-dimensional hypersphere), while for $k < 0$ the space is open. Thus ρ_c is also called the *closure density*; e.g. for $\rho > \rho_c$ the space is closed, for $\rho < \rho_c$ it is open.

The dynamical feature of Ω is understood within the Newtonian framework in terms of 'escape velocity'. The relation (2.31) may be rewritten in the form

$$\frac{1}{2}\dot{S}^2 - \frac{4\pi G\rho}{3S} = -\frac{1}{2}kc^2. \tag{2.41}$$

For a unit sphere ($r_0 = 1$), the radius $R = r_0 S = S$ and $1/2\dot{R}^2 \equiv 1/2\dot{S}^2$ is the kinetic energy of outward motion of a particle of unit mass comoving with the surface of the sphere. Similarly $-4\pi G\rho/3S \equiv -4\pi G\rho/3R$ is the potential energy of that particle. Thus, $-kc^2/2$ is the total energy of the particle. The particle 'escapes' to infinity if $k < 0$, is trapped if $k > 0$ and is on the borderline for $k = 0$. The expanding universe behaves likewise!

The three types of models described here are commonly known as *Friedmann models* as they were first obtained in 1922–24 by Alexander Friedmann. Friedmann's work was, however, in relativistic cosmology. Despite the differences between the Newtonian and relativistic theories of gravity, it is something of a surprise that formally the models derived here by Newtonian methods are the same as the Friedmann models. Even the redshift formula (2.17) derived here by the Newtonian methods agrees with the relativistic formula!

2.6 The cosmological constant

In 1917 Einstein had attempted to obtain within the framework of general relativity the theoretical model of a static universe. In this he, at first, did not succeed. The reason is apparent from our dynamical equation (2.30) which does not admit a solution with $\dot{S}=0$, $\ddot{S}=0$, $S=$ constant. To get round the difficulty, Einstein added an extra term called the 'λ-term' to his equations, where λ is a constant known commonly as the *cosmological constant*.

In 1917, nebular redshift was not regarded as universally established (remember Hubble's law came in 1929); so Einstein's desire to have a static model is understandable. The additional term he introduced had negligible effect on terrestrial or even galactic gravity : it became significant only at the cosmological level. We shall shortly see why. Later, when the expanding universe concept gained currency and the 1922 models of Friedmann became relevant, Einstein realized that the λ-term was not necessary after all. He therefore retracted it as 'the greatest blunder' in his life. Nevertheless the term has survived largely because several astronomers and physicists have found it attractive for various reasons. We will therefore briefly discuss it here even though we are using the Newtonian framework.

The λ-term corresponds to a radial force of repulsion between two masses that varies *in proportion* to the distance between them. Thus, for two particles A and B separated by a vector $\mathbf{r}$, B will be repelled by a force $\frac{1}{3}\lambda\mathbf{r}$ per unit mass from A and vice versa. Therefore (2.26) gets modified to

$$\rho\left[\frac{\partial\mathbf{v}}{\partial t}+(\mathbf{v}.\nabla)\mathbf{v}\right]=-\nabla p+\rho\mathbf{F}+\frac{1}{3}\lambda\rho\mathbf{r} \tag{2.42}$$

and (2.30) is changed to

$$\ddot{S}=-\frac{4\pi G\rho_0 S_0^3}{3S^2}+\frac{1}{3}\lambda S. \tag{2.43}$$

Alternatively, we may modify equation (2.27) by adding λ to the right-hand side. Similarly, instead of (2.31) we have

$$\dot{S}^2=\frac{8\pi G\rho_0 S_0^3}{3S}-kc^2+\frac{1}{3}\lambda S^2. \tag{2.44}$$

Now we see that it is possible to solve (2.43) and (2.44) with the requirement that

$$S = S_E = \text{constant}. \tag{2.45}$$

Writing

$$\rho_E = \frac{\rho_0 S_0^3}{S_E^3}, \tag{2.46}$$

we get the relevant equations as

$$-kc^2 + \frac{1}{3}\lambda S_E^2 + \frac{8\pi G}{3}\rho_E S_E^2 = 0, \tag{2.47}$$

$$-\frac{4\pi G}{3}\rho_E S_E + \frac{1}{3}\lambda S_E = 0. \tag{2.48}$$

These are easily solved to give

$$\lambda = 4\pi G \rho_E, \quad kc^2 = \lambda S_E^2. \tag{2.49}$$

In other words, $k > 0$. The undetermined constant S_E can be fixed by setting $k = 1$. Thus the radial size of the universe is related to the density through the fundamental constants λ and G.

In relativistic cosmology also, Einstein found the corresponding answer : that the universe is closed, with a finite volume. Einstein liked the fact that in his model the radius of the universe was determined by the density of matter in it, in a clear demonstration that the geometry of space is uniquely related to the matter occupying it.

The 'Einstein universe', as the model came to be known, did not long enjoy a unique status in cosmology as its creator had hoped. In 1917, a few months after Einstein's result, W. de Sitter found another solution of equations (2.43) and (2.44) :

$$S \propto e^{Ht}, \quad \rho_0 = 0, \quad k = 0, \tag{2.50}$$

where

$$\lambda = 3H^2. \tag{2.51}$$

This universe expands for ever, exponentially, but is *empty*. The de Sitter universe describes *motion without matter* in contrast to the Einstein universe which has *matter without motion*.

What about more general solutions? Looking at (2.44) we see that there are basically two parameters, k and λ. For k greater than, equal to or less than 0, we have different dynamical behaviour for different λ. Figures 2.3 (a–c) illustrate them.

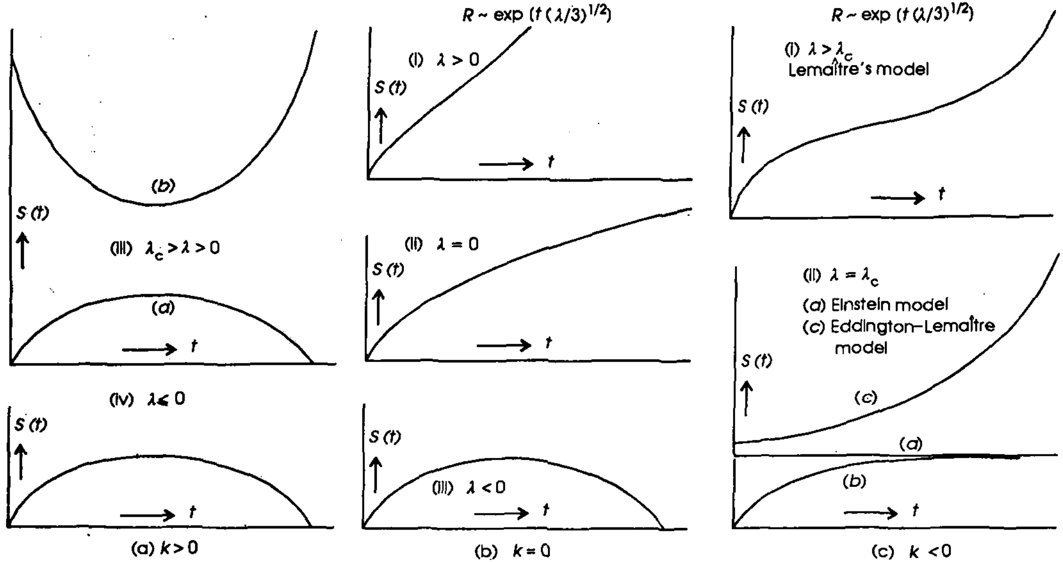


Fig.2.3 The three cases — $k > 0, k = 0, k < 0$ — are illustrated by $S(t)$ curves for different values of λ in the respective figures **(a)**, **(b)** and **(c)**.

Figure 2.3(c) shows the interesting case advocated by Eddington and Lemaitre. In Lemaitre's version, λ is very close to but slightly greater than the critical value for the Einstein universe. The model has the universe expanding from $S = 0$, coasting along close to $S = S_E$ for a considerable period and then expanding away. During the coasting period, the universe is in a pseudo-Einstein state while for the asymptotic future it is in the de Sitter state. Eddington felt that an Einstein universe would be unstable and expand, being so triggered by the process for forming galaxies. For, if galaxies form by gravitational condensation of matter, the process is helped if the universe is static or near static rather than expanding.

Problem 2.4 : How is the relation (2.39) modified in the λ-cosmologies?

Solution : Writing $\ddot{S} = -qH^2S$, $\dot{S}^2 = H^2S^2$ and $\rho_0 S_0^3/S^3 = 3H^2\Omega/8\pi G$, we get from (2.43)

$$-qH^2 = -\frac{1}{2}\Omega H^2 + \frac{1}{3}\lambda$$

i.e., $$q = \frac{1}{2}\Omega - \frac{\lambda}{3H^2}.$$

2.7 Space–time singularity

The Friedmann models of Section 2.5 and the λ-cosmologies in general have the common feature that S becomes zero at some epoch. In Newtonian cosmology this implies a state of infinite density — and possibly infinite temperature if we could extend our dust model to those wherein pressure also matters (see Chapter 3). This is an unphysical state of affairs but it gets worse in the corresponding relativistic models wherein also the $S = 0$ state appears. The reason why this is worse is that in relativity the physical contents of the universe determine its geometrical properties, and these were undefined at $S = 0$. Space–time was *singular* at this epoch.

It is usual to call this singular epoch the epoch of *big bang*, a phrase coined by Fred Hoyle, and these models are often referred to as the 'big bang models'. Some general theorems tell us that under normal physical conditions the big bang-type singular situation is unavoidable in relativity. In Newtonian cosmology, the state of infinite density does not imply space–time singularity because the close relationship between physics and geometry is not present there.

2.8 The steady state model

In 1948 Hermann Bondi, Thomas Gold and Fred Hoyle proposed an alternative to the Friedmann models, an alternative that did not have the singular phase. Bondi and Gold obtained this model by generalizing the cosmological principle to the *perfect cosmological principle* (PCP). The PCP introduces homogeneity not only in space but also in time. That is, it states that the universe on the large scale is unchanging in time. Such a universe is called the *steady state universe*.

In a steady state universe the Hubble constant

$$H(t) = \frac{\dot{S}}{S}$$

must be a constant, so that

$$S \propto \exp Ht, \quad H = \text{constant}. \tag{2.52}$$

The density ρ is also a constant in this universe. The PCP, however, is not able to determine ρ in terms of H and other physical properties of the universe since it lacks a quantitative dynamical theory. Bondi and Gold argued that the observations of the local universe together with the PCP are sufficient to determine the physical features of the universe anywhere at any epoch *without* a dynamical theory.

The constancy of ρ despite expansion means that matter must be continually created at a rate

$$Q = 3H\rho. \tag{2.53}$$

What is the physical mechanism of creation?

Hoyle tackled this question in his independent approach to the steady state theory. He proposed a scalar creation field of cosmological nature that interacts with matter at the time of creation. The creation field has *negative pressure* and *negative energy* density. We will not go into the details of the approach here. However, the currently popular inflationary universe to be discussed in the following chapter has considerable similarity with the above picture.

In 1993, Hoyle, G. Burbidge and the author had proposed a variation of the original steady state theory in which the universe has a scale factor of the kind

$$S(t) = e^{t/P}\{1 + \alpha \cos\theta(t)\}. \tag{2.54}$$

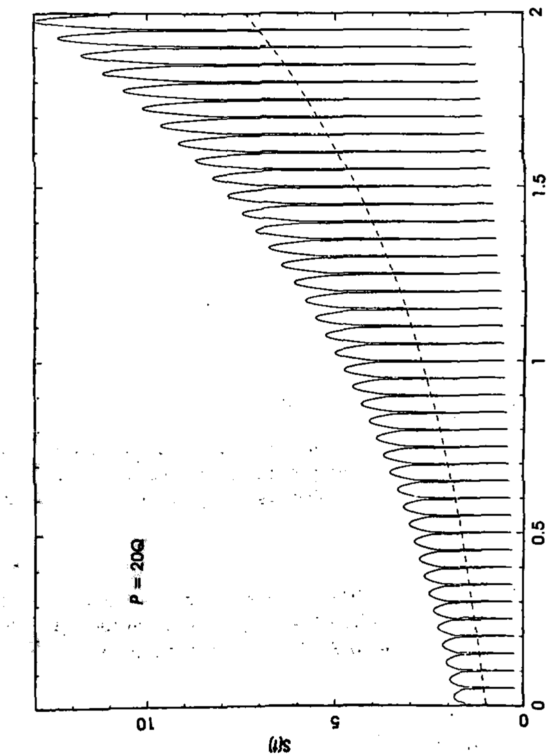


Fig.2.4 (a) The scale factor for the quasi-steady state cosmology. Here we see the long term behaviour.

Here P is a constant and the constant parameter α satisfies $0 < \alpha < 1$. The function $\theta(t)$ is determined by the dynamics of the creation process and is periodic with period Q such that $\theta(0) = 0, \quad \theta(Q) = 2\pi$. This cosmology is called the *quasi-steady state cosmology*.

The creation process in the model is periodic, with a stop–go character which causes the universe to oscillate around an average that increases exponentially with time. The characteristic period P of exponential growth is very large (say $P \approx 20Q$) compared to the period Q of oscillation. Since $|\alpha| < 1$, the universe is non-singular. Figure 2.4(a) gives the long term scale factor of this cosmology. Notice that in Fig.2.4(b) we have the possibility of some sources being seen blueshifted. We will refer back to this model in Chapter 5.

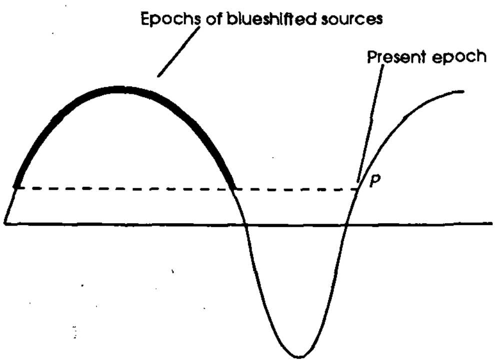


Fig.2.4(b) The scale factor for the quasi-steady state cosmology. Here we see a typical oscillation. If the present epoch is denoted by P, the thick part of the previous oscillation denotes the epochs of blueshifted sources.

2.9 Exercises

2.1 In a universe with $S(t) \propto t^{2/3}$, we observe a galaxy with redshift 1.25. How long has light taken to travel from the galaxy to us?

2.2 Show that the luminosity distance of a galaxy of redshift z in a de Sitter universe is given by

$$D_L = \frac{c}{H_0} z\,(1 + z).$$

2.3 By writing

$$S = A(1 - \cos\psi),\ t = A(\psi - \sin\psi),\ A = \text{constant}$$

solve (2.31) for the case $k > 0$.

2.4 From equation (2.44) obtain a relation between k, Ω and λ at the present epoch.

2.5 Astronomers find the density of matter in the form of galaxies to be $\sim 3 \times 10^{-31}$ g cm^{-3}. If the universe is described by the $k = 0, \lambda = 0$ Friedmann model and $H_0 = 100$ kms^{-1} Mpc^{-1}, is there a serious discrepancy between the observed density and theoretical density?

3 *History of the Universe*

3.1 Matter vs radiation-dominated universe

The simple dust-dominated models we discussed in Chapter 2 provide a reasonable description of the present universe, and possibly of its future state. However, when we extrapolate these models to the past, towards the state with $S = 0$, we need to be more cautious. As we will see next, the assumption of matter dominating over radiation becomes increasingly more suspect as we probe the past of the universe.

Let us consider gas or radiation with pressure p and density ρ, undergoing adiabatic expansion. By the laws of thermodynamics we get for a specific volume V

$$d(\rho c^2 V) + p dV = 0. \tag{3.1}$$

In an expanding universe $V \propto S^3$ and this relation becomes

$$d(\rho c^2 S^3) + 3pS^2 dS = 0. \tag{3.2}$$

If we are dealing with dust, i.e., pressure-free matter, we get

$$\rho S^3 = \text{constant}, \tag{3.3}$$

exactly as in equation (2.25). If instead we are dealing with radiation, $p = \rho c^2/3$ and (3.2) integrates to

$$\rho S^4 = \text{constant}. \tag{3.4}$$

To distinguish between dust and radiation, we will use the symbols ρ_d and ρ_r to denote their respective densities. As a result of (3.3) and (3.4) we get

$$\rho_d \propto S^{-3}, \quad \rho_r \propto S^{-4}, \quad \rho_d/\rho_r \propto S. \tag{3.5}$$

Thus in an expanding universe the dust begins to dominate over radiation as the expansion continues beyond a certain limit. What is that limit?

The evidence for 'dust' in the form of visible matter in the galaxies, intergalactic medium, etc. suggests that at present

$$\rho_d \approx 3 \times 10^{-31} \text{ g cm}^{-3}. \tag{3.6}$$

In Chapter 4 we will review the observational evidence which suggests that the universe may very likely contain dark matter far in excess of the visible matter.It is possible in view of this that the actual matter density is higher. Let us therefore enhance (3.6) by a 'dark matter factor' $\alpha > 1$ and write

$$\rho_d = 3\alpha \times 10^{-31} \text{ g cm}^{-3}. \tag{3.7}$$

As we saw in the opening chapter, the radiation background exists in all observable wavelengths but the most dominant one is in microwaves in the form of a black body radiation of temperature ~ 2.7 K. This corresponds to an energy density

$$\rho_r c^2 \approx 4 \times 10^{-13} \text{ erg cm}^{-3},$$

i.e., $$\rho_r \approx 4 \times 10^{-34} \text{ g cm}^{-3}. \tag{3.8}$$

Thus, at present

$$\frac{\rho_d}{\rho_r} \approx \frac{3\alpha}{4} \times 10^3, \quad \text{i.e.,} \quad \rho_d \gg \rho_r. \tag{3.9}$$

As the universe expands this ratio will further increase. Thus not only is the universe matter dominated today, it will continue that way so long as it expands. But what about the past?

Clearly for a small enough value of S, ρ_d was *less* than ρ_r. From equations (3.5) and (3.9) we see that this would occur at an epoch of redshift

$$z_{eq} \sim \frac{3\alpha}{4} \times 10^3. \tag{3.10}$$

At

$$S_{eq} = \frac{S_0}{1 + z_{eq}} \tag{3.11}$$

the matter and radiation densities were comparable. For $z > z_{eq}$ and $S < S_{eq}$, the universe was *radiation dominated*. Our discussion in Chapter 2 was applicable to a matter-dominated universe and will have to be modified for $S < S_{\mathrm{eq}}$. We will next consider those early epochs when the effect of dust was negligible compared to that of radiation.

3.2 The hot universe

First we should note that the Weyl postulate is not strictly followed by galaxies. They have small random motions of upto ~ 1000 km s^{-1} relative to the local standard of rest. When we look at a galaxy of redshift 0.01, say, these random motions are negligible compared to the universal expansion. How do these random motions respond to the overall expansion of the universe? Consider a small perturbation of velocity in the form

$$\mathbf{v} = \mathbf{r}\frac{\dot{S}}{S} + \mathbf{v}_1, \tag{3.12}$$

where $\mathbf{v}_1$ denotes the peculiar (i.e., random) motion. In general $\mathbf{v}_1$ will not depend systematically on the spatial coordinates but would depend on time. Thus, omitting the second and higher powers of $|\mathbf{v}|$ we get

$$\frac{\partial \mathbf{v}}{\partial t} = \mathbf{r}\dot{H} + \frac{\partial \mathbf{v}_1}{\partial t}, \qquad H = \dot{S}/S,$$

$$\mathbf{v}.\nabla\,\mathbf{v} \cong \mathbf{r}H^2 + \mathbf{v}_1.\nabla\left(\mathbf{r}\frac{\dot{S}}{S}\right) = \mathbf{r}H^2 + H\mathbf{v}_1,$$

and equation (2.26) gives us for the perturbation in velocity,

$$\frac{\partial \mathbf{v}_1}{\partial t} + H\mathbf{v}_1 = 0,$$

i.e.,

$$\mathbf{v}_1 S = \text{constant}. \tag{3.13}$$

In other words, the random motions drop off as S^{-1}. So turning the clock of the universe backwards, one finds that the random motions and hence the pressure term would be larger and larger at smaller values of S. Hence at sufficiently early epochs, the motions of fundamental observers would be random and relativistic. Of course it is unrealistic to assume that galaxies existed at such early epochs.

in Chapter 2 was applicable to a matter-dominated universe and will have to be modified for $S < S_{eq}$. We will next consider those early epochs when the effect of dust was negligible compared to that of radiation.

3.2 The hot universe

First we should note that the Weyl postulate is not strictly followed by galaxies. They have small random motions of upto ~ 1000 km s^{-1} relative to the local standard of rest. When we look at a galaxy of redshift 0.01, say, these random motions are negligible compared to the universal expansion. How do these random motions respond to the overall expansion of the universe? Consider a small perturbation of velocity in the form

$$\mathbf{v} = \mathbf{r}\frac{\dot{S}}{S} + \mathbf{v}_1, \tag{3.12}$$

where $\mathbf{v}_1$ denotes the peculiar (i.e., random) motion. In general $\mathbf{v}_1$ will not depend systematically on the spatial coordinates but would depend on time. Thus, omitting the second and higher powers of $|\mathbf{v}|$ we get

$$\frac{\partial \mathbf{v}}{\partial t} = \mathbf{r}\dot{H} + \frac{\partial \mathbf{v}_1}{\partial t}, \quad H = \dot{S}/S,$$

$$\mathbf{v} . \nabla \mathbf{v} \cong \mathbf{r}H^2 + \mathbf{v}_1 . \nabla \left(\mathbf{r}\frac{\dot{S}}{S}\right) = \mathbf{r}H^2 + H\mathbf{v}_1,$$

and equation (2.26) gives us for the perturbation in velocity,

$$\frac{\partial \mathbf{v}_1}{\partial t} + H\mathbf{v}_1 = 0,$$

i.e.,

$$\mathbf{v}_1 S = \text{constant}. \tag{3.13}$$

In other words, the random motions drop off as S^{-1}. So turning the clock of the universe backwards, one finds that the random motions and hence the pressure term would be larger and larger at smaller values of S. Hence at sufficiently early epochs, the motions of fundamental observers would be random and relativistic. Of course it is unrealistic to assume that galaxies existed at such early epochs.

Rather, we should talk of the subatomic particles moving relativistically at sufficiently early epochs. Later, as the universe continued to expand these motions slowed down; bigger units formed and galaxies subsequently appeared in the shapes and sizes we see today. How this happened is yet an unknown story but we will discuss the issue in general terms towards the end of this chapter.

Problem 3.1 Estimate the redshift at which galactic peculiar motions would be relativistic (assuming the galaxies existed then!).

Solution : From the relation (3.13), v increases from $\sim$ 1000 km s^{-1} to $c \simeq 3 \times 10^5$ kms^{-1} for a decrease of S by a factor $\sim$ 300. The redshift is therefore of this order.

The relativistic particles resemble radiation in that their equation of state satisfies the relation

$$p = \frac{1}{3}\rho c^2. \tag{3.14}$$

This causes a change in the $\rho - S$ relation to that given by (3.4) :

$$\rho S^4 = \text{constant} = A \text{ (say)}. \tag{3.15}$$

Thus equation (2.30) changes to

$$\ddot{S} = -\frac{4\pi GA}{3S^3}, \tag{3.16}$$

and (2.31) changes to

$$\dot{S}^2 = \frac{4\pi GA}{3S^2} - kc^2. \tag{3.17}$$

Henceforth we will *assume* that we are dealing with such small values of S that

$$|kc^2| \ll \frac{4\pi GA}{3S^2}. \tag{3.18}$$

Ignoring the k-term therefore we get the solution as

$$S = \beta t^{1/2} \tag{3.19}$$

where

$$\beta^4 = \frac{16\pi GA}{3}. \tag{3.20}$$

The corresponding density is

$$\rho = \frac{A}{S^4} = \frac{3}{16\pi Gt^2}. \tag{3.21}$$

Now if all of it were in the form of pure electromagnetic radiation, the black body law would tell us that its energy density is related to radiation temperature T by the formula

$$\rho c^2 = aT^4. \tag{3.22}$$

Here a is the radiation constant. Thus we have from (3.21),

$$T = \left(\frac{3c^2}{16\pi Gat^2}\right)^{1/4} \propto \frac{1}{S}. \tag{3.23}$$

In other words, the temperature drops off as $t^{-1/2}$, inversely proportional to the scale factor.

It is interesting to note that the right-hand side of the Newtonian relation (3.23) differs by a factor 2 from the corresponding relativistic relation

$$T = \left(\frac{3c^2}{32\pi Gat^2}\right)^{1/4}. \tag{3.24}$$

Why this discrepancy? The factor 2 could be reconciled if we argue that the gravitating density is not ρ but $\rho + 3p/c^2$. This would increase ρ by factor 2 in view of (3.14). We will not discuss this point further since it takes us farther from our present goal of describing the early universe. We will henceforth take the relativistic expression (3.24) as the correct one and likewise multiply the right-hand side of (3.16) by the factor 2.

What happens when we include relativistic particles like electrons, positrons, etc., apart from photons in our cosmic brew? In thermodynamic equilibrium each species contributes towards the overall pressure and density. Without going into details of quantum statistical mechanics we will simply state the result that emerges.

Quantum mechanics makes a distinction between two types of particles — the *fermions* and the *bosons*. The fermions are described by antisymmetric wavefunctions and are subject to Pauli's exclusion principle, whereas the bosons described by symmetric wavefunctions are not subject to that restriction. In either case, however, the rules of classical statistical mechanics have to be modified to take into account the fact that at the microscopic level, particles of the same species are indistinguishable. Thus, fermions obey the Fermi–Dirac statistics, while bosons follow the Bose–Einstein statistics, both of which are different from the classical Maxwell–Boltzmann statistics. The radiation energy density contributed by fermions F of internal degrees of freedom (spin of the particle) g_F at temperature T is

$$\rho c^2 = \frac{7}{16} g_F a T^4, \tag{3.25}$$

and the corresponding relation for bosons B is

$$\rho c^2 = \frac{1}{2} g_B a T^4. \tag{3.26}$$

The relation (3.22) for photons is a special case of (3.26) for $g_B = 2$. Now suppose the total bosonic and fermionic degrees of freedom in the cosmic mixture of relativistic particles are

$$g_b = \sum_B g_B, \quad g_f = \sum_F g_F. \tag{3.27}$$

Then (3.22) is modified to

$$\rho c^2 = \frac{1}{2} g a T^4, \tag{3.28}$$

where

$$g = g_b + \frac{7}{8} g_f. \tag{3.29}$$

The temperature–time relation (3.24) then changes to

$$T = \left(\frac{3c^2}{16\pi G g a t^2}\right)^{1/4}. \tag{3.30}$$

In the above, the assumption that a particle X of rest mass m_X is moving relativistically is justified if its thermal energy kT is large compared to its rest mass energy, i.e., if

$$T > \frac{m_X c^2}{k} \equiv T_X \text{ (say)}. \tag{3.31}$$

Here k is the Boltzmann constant. For electrons (e), $T_e = 5.93 \times 10^6$K, while for protons (p), $T_p = 10^{13}$K.

Problem 3.2 In a mixture containing relativistic electrons, positrons, electron and muon neutrinos and their antineutrinos and photons in thermodynamic equilibrium in the early universe, show that

$$T = \left(\frac{c^2}{48\pi G a}\right)^{1/4} t^{-1/2}.$$

Solution: The fermions (F) with their spin degrees of freedom are as follows :

F	e^-	e^+	ν_e	ν_μ	$\bar{\nu}_e$	$\bar{\nu}_\mu$
g_F	2	2	1	1	1	1

giving $g_f = 8$. Similarly, the only bosons are photons γ for which $g_b \equiv g_\gamma = 2$. So

$$g = \frac{7}{8} g_f + g_b = 7 + 2 = 9$$

and the result follows from (3.30).

From Problem 3.2, the relation (3.30) for $e^\pm$, ν_e, $\bar{\nu}_e$, ν_μ, $\bar{\nu}_\mu$, γ, all in thermodynamic equilibrium, becomes

$$T_{10} \cong 1.04 t_s^{-1/2}, \qquad (3.32)$$

where T_{10} is temperature in units of 10^{10} K and t_s is time measured in seconds. It is easy to check back from (3.31) that for $t_s = 1$, all the above particles are indeed moving relativistically.

The fact that the universe soon after its origin in the big bang passed rapidly through states of very high temperature was first used for a practical calculation by George Gamow in the 1940s. Along with his younger colleagues Ralph Alpher and Robert Herman, they explored the epoch 1s–200s when the temperature of the universe dropped from $\sim 10^{10}$ K to $\sim 10^8$ K. At such high temperatures, they argued, it would be possible for free neutrons and protons to come together to form atomic nuclei. They also predicted that the high temperature radiation present at that time should have cooled down to give a moderately low radiation background today. In 1965 the discovery of the microwave background by Arno Penzias and Robert Wilson lent credibility to Gamow's scheme and to the overall scenario of the hot big bang.

This time–temperature relationship therefore plays an important role in the early history of the universe. We will next discuss some of its important phases when significant physical developments took place. In this we will proceed in the chronological fashion indicated below (and not in the historical sequence of their entering the scope of cosmology) :

(i) Quantum gravity era
(ii) Inflationary phase
(iii) Quark–gluon transition
(iv) Decoupling of neutrinos
(v) Annihilation of $e^{\pm}$ pairs
(vi) Primordial nucleosynthesis
(vii) Decoupling of radiation from matter

Of these, (i)–(iii) belong to the 'very early' universe while (iv)–(vii) describe the 'early' universe. Gamow's own investigations relate to (iv)–(vi).

The temperature is an indicator of the energy of typical particles and we can write it in the so-called energy units by using the relation

$$kT = E,$$

E being the typical energy per particle. If we express E in MeV or GeV units we get the relation (3.30) in the form

$$t_s = 2.4g^{-1/2}E_{\text{MeV}}^{-2} = 2.4 \times 10^{-6}g^{-1/2}E_{\text{GeV}}^{-2}. \tag{3.33}$$

Expressed in this form we argue that t_s was the 'age' of the universe when the typical particle energy in MeV units was E_{MeV}, etc. The quantity g varies with epoch only when there is a phase transition changing the profile of the particles present.

This equation leads to the currently fashionable scenario in which the very early universe is looked upon as a particle accelerator that generated particles of very high energies at sufficiently early epochs. Particle physicists hope that their theories of grand unification at energies $\geq 10^{15}$ GeV can find applications at these very early epochs. The pioneer in this concept of applying particle physics to the early universe of course was George Gamow who discussed the era of $t_s \approx 1 - 200$ during which primordial nucleosynthesis should have taken place. It is the work on primordial nucleosynthesis that still claims to be the most definitive of all the different phases of the early/very early universe.

3.3 The very early universe

(i) The quantum gravity era : In this case, the energy typical of physical interactions is the so-called 'Planck energy'

$$E_P = M_P c^2 = \sqrt{\frac{c^5 \hbar}{G}} \cong 10^{19}\text{GeV}. \tag{3.34}$$

Here M_P is called the 'Planck mass'. The corresponding time is the 'Planck time'

$$t_P = \sqrt{\frac{G\hbar}{c^5}} \cong 5 \times 10^{-44}\text{s}. \tag{3.35}$$

The characteristic length scale of the universe is the 'Planck length'

$$L_P = \sqrt{\frac{G\hbar}{c^3}} \approx 1.5 \times 10^{-33}\text{s}. \tag{3.36}$$

What does the quantum gravity era really mean?

The era $0 < t \leq t_P$ denotes the breakdown of the classical theory of gravity. The appearance of the three fundamental constants c, $\hbar$ and G together suggests that here we are reaching a synthesis of the theory of relativity and quantum theory. The ideas of space–time measurements so crucial to general relativity break down at this level and one may have to resort to the probability language of quantum mechanics. Likewise the effects of gravity are so strong that quantum theory can no longer ignore them.

One fundamental result of quantum gravity should be to decide whether the universe as a whole has space–time singularity. Here we have to talk in the language of probability and ask for the likelihood that the universe has emerged from a singular origin. It is too early to say what the general answer to this question will be.

(ii) The inflationary phase: In 1981 Alan Guth proposed an important variation on the concept of the feedback of phase transition at grand unification on the expansion of the universe. We will only deal with the idea qualitatively.

In *grand unification*, three basic interactions — strong, weak and electromagnetic — are believed to be unified, i.e., expressed as parts of a single interaction. Just as the phenomena of electricity and magnetism get unified into the electromagnetic theory when we consider interactions of electric charges moving rapidly, so at sufficiently high energy the three interactions get unified. As shown in Fig.3.1, the strengths of the three interactions become comparable at energies of the order of 10^{15} GeV, the energy at which unification is likely.

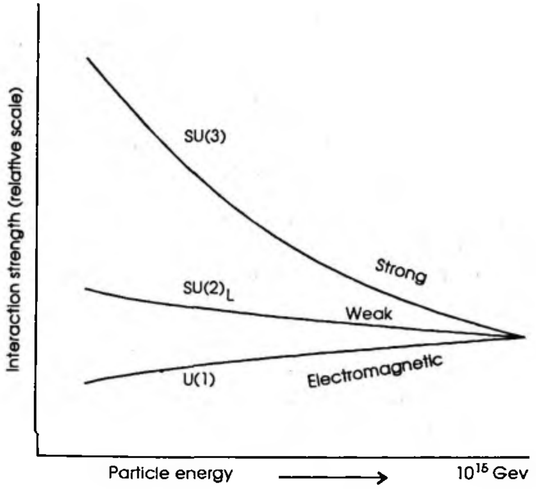


Fig.3.1 The strong, the weak and the electromagnetic interactions work with different strengths, but as the energy of participating particles is increased their relative strengths change and at particle energies of the order 10^{15} GeV they become comparable. This is where unification of interactions is expected to take place. The symbols $SU(3), SU(2)_L$ and $U(1)$ denote the group theoretic structures of these interactions.

Although a well-established grand unified theory (GUT) is not yet available, the general expectation is that at a typical energy of the above order there is a switchover from the GUT to the two component theories, i.e., the strong interaction theory and the electroweak theory (that still retains the weak and the electromagnetic interaction in one fold). Therefore a phase transition becomes due as the universe expands and cools past this energy threshold. An analogy may be drawn with the cooling of steam to form water drops.

In this analogy, it is possible to *supercool* the steam below the usual boiling point of water so that it still retains its vapour state. However, in this state the steam has higher energy (of latent heat) than the corresponding liquid state and it is consequently unstable. With slight disturbance it spontaneously condenses into water droplets, thereby releasing energy.

Likewise we have two possible energy states in the very early universe. One in which all interactions are unified is the only possible state above the GUT energy of $\sim 10^{15}$ GeV. Below this energy we expect the phase transition to separate the electroweak interaction from the strong one. In this phase transition the vacuum plays a crucial role.

In classical physics the vacuum is devoid of anything and as such it is a trivial entity. Not so in quantum physics where the vacuum has dynamic activity going on; only the activity is in the lowest key. Technically we may identify the vacuum state of a system as the state of lowest possible energy. Let us see how this notion plays a dramatic role at the GUT phase-transition epoch.

As in the case of supercooled steam, here too we may have the pre-transition phase persisting for a while even below this energy. The vacuum (i.e. ground) state of this phase will, however, have higher energy than the vacuum of the post-transition phase.

Since the vacuum is defined as the state of lowest energy, the *true vacuum* in the above scenario is that of the post-transition phase. The vacuum of the pre-transition state is called the *false vacuum*. As soon as the transition occurs, latent energy (the difference in the energies of the true vacua) is released, raising the ambient temperature of the universe.

Guth in 1981 pointed out the possible feedback this energy difference would have on the dynamics of the universe. Quantum field theory tells us that the false vacuum would have an equation of state

$$\rho_v c^2 = -p_v > 0, \tag{3.37}$$

i.e., it would resemble the λ-term of Einstein. Since $\rho_v \gg \rho_m, \rho_r$ (the vacuum energy far exceeds the particle/radiation energy) the expansion of the universe is determined by the above vacuum terms. The dynamical behaviour of vacuum feedback is described through a scalar field which has considerable similarities to the creation field of the steady state theory which also has the de Sitter expansion.

Taking the effective gravitating density as $\rho_v + 3p_v/c^2$, we see that from (3.37) this value is $-2\rho_v$. Substituting this value for ρ in equation (2.29) we get for $\rho_v =$ constant

$$H^2 = \frac{8\pi G\rho_v}{3} \equiv a^2 \quad \text{(say)}, \tag{3.38}$$

where $a =$ constant. Thus the scale factor rises exponentially

$$S \propto \exp at \tag{3.39}$$

until the phase transition is complete. Thereafter the expansion reverts to the Friedmann model, but with increased temperature of radiation.

This exponential expansion is called *inflation*. The underlying dynamics of the model depends on quantum field theory of how the false vacuum behaves vis-a-vis the true vacuum. We will not go into those details here but state briefly the advantages of such an inflationary phase. In particular, the model solves a few outstanding problems (given below) of the standard hot big bang cosmology.

a) The horizon problem Imagine an observer seeking to communicate with others at an epoch $t > 0$ in the radiation-dominated phase, when $S \propto t^{1/2}$. Using the maximum speed available — the speed of light — the distance upto which he will be able to communicate would be finite, being equal to $2ct$. This is the radius of the observer's *particle horizon*. An observer at $r = 0$, at time t, can receive signals at the most from this distance.

Problem 3.3 Show that if $S \propto t^n$, $0 < n < 1$, the radius of the particle horizon at epoch t_1 is $ct_1/(1-n)$.

Solution Consider a light ray leaving $r = r_1$ at $t = 0$, and reaching the observer at the origin at time t_1. Then for $S = At^n$ (A = constant),

$$r_1 = \int_0^{t_1} \frac{cdt}{At^n} = \frac{c}{A}\frac{t_1^{1-n}}{1-n}.$$

Hence $r_1 S(t_1) = ct_1/1-n$.

Now, for the GUT epoch, $t \sim 10^{-36}$ s, say, and the radius of the particle horizon is 6×10^{-26} cm. If the universe has expanded in the Friedmann fashion till its temperature has dropped from 10^{15} GeV to $3K \sim 3 \times 10^{-4}$ eV, the scale factor which, as seen from (3.23), varies inversely as the radiation temperature would have increased by a factor $\sim 3 \times 10^{27}$. The above particle horizon would accordingly have grown to a size of 180 cm only! Since no physical interaction travels faster than light, the particle horizon places limits on the range of causal influences. If the initial composition of the universe were determined by causality-controlled processes then we cannot expect homogeneity to be established beyond this range. But the above range of 1.8 m is absurdly small compared to the size of the observable universe today. In other words, with such small particle horizons, how did the universe manage to be homogeneous on such a large scale?

Inflation solves the problem by introducing the steep exponential rise in S. If $S \propto e^{at}$ and the inflation lasts for $t_i < t < t_f$, say, then the linear scales including the horizon size are increased by the factor

$$Z = \exp[a(t_f - t_i)]. \tag{3.40}$$

In various versions of inflation this factor can be as high as $\sim 10^{28}$ to $\sim 10^{50}$, thus explaining why homogeneity is established over large regions as observed.

b) The flatness problem Recall that in (3.18) we ignored the $|kc^2|$-term in comparison with $4\pi GA/3S^2$, thus essentially setting $k = 0$. In general relativity this means opting for the 'flat' model. The justification was that at an early enough stage this term is unimportant. However, a careful discussion of this approximation tells us something more. Consider equation (3.17) again :

$$\dot{S}^2 = -kc^2 + \frac{4\pi GA}{3S^2}. \tag{3.41}$$

If the state of the universe were determined at the GUT epoch, we expect the two terms on the right-hand side to be comparable. This means

$$|kc^2| \approx \frac{4\pi GA}{3S^2}$$

at the GUT epoch. Hence the universe would either come to a halt and recollapse in a time scale of the order of 10^{36} s (for $k > 0$) or it would disperse away to near-zero density in a similar time (for $k < 0$). How therefore did it survive as a marginally bound unit right upto the present times?

Alternatively, using (2.37), we get for any epoch t

$$\Omega - 1 = \frac{kc^2}{\dot{S}^2}, \quad \Omega_0 - 1 = \left.\frac{kc^2}{\dot{S}^2}\right|_{t_0}$$

i.e., writing $\dot{S}_0^2 \equiv \dot{S}^2$ at $t = t_0$,

$$\Omega - 1 = (\Omega_0 - 1)\frac{\dot{S}_0^2}{\dot{S}^2}. \tag{3.42}$$

At an early enough epoch $\dot{S}^2$ was very large compared to $\dot{S}_0^2$ at the present epoch. This would mean that $|(\Omega - 1)|$ was very small compared to its present value — the smallness being by a factor as small as $\sim 10^{-50} - 10^{-60}$.

Problem 3.4 Estimate the ratio $\dot{S}_0^2/\dot{S}^2$ for the GUT epoch, when $k = 0$.

Solution At present $\dot{S}_0 = H_0 S_0$, while at the GUT epoch $\dot{S} = HS$. For time t, $S \propto t^{1/2}$ and $H = 1/2t$. At time t_0, $S \propto t^{2/3}$ and $H_0 = 2/3t_0$. Therefore,

$$\frac{\dot{S}_0^2}{\dot{S}^2} = \frac{16}{9}\left(\frac{t^2}{t_0^2}\right)\cdot\frac{S_0^2}{S^2}.$$

With $S_0/S = 3 \times 10^{27}$ and $t/t_0 \approx 10^{-36}\text{s}/3 \times 10^{17}\text{s} \sim 3 \times 10^{-54}$, we get the above ratio as $\sim 10^{-52}$. Thus $|\Omega - 1|$ would have been 10^{-52} times its present value.

It follows from Problem 3.4 that unless $\Omega = 1$ strictly, any departure of the present value of Ω from unity must imply extreme fine tuning of Ω close to 1 at the GUT epoch.

Inflation avoids this situation by diminishing the importance of the kc^2-term in comparison with the $\dot{S}^2$-term *during the exponential* expansion. In other words, after the inflation is over, the kc^2-term is effectively zero.

(c) *The entropy problem* The present estimates of the ratio of photons to baryons in the universe are in the range 10^8 to 10^{10} (see equation (3.50)). Why did the number become so large? This is often expressed as the 'entropy problem', i.e., too many photons per baryons indicates a high value of entropy. Inflation solves the problem by the deduction that after phase transition and the dumping of excess energy (the difference in the energies of false and true vacua) in the form of heat, the entropy rises.

(d) *The monopole problem* One consequence of the GUT phase transition is the creation of a magnetic monopole. Maxwellian electrodynamics does not permit the existence of monopoles, but they do arise in the above scenario. What is embarrassing is their excessive mass. The mass of a monopole is 10^{16} GeV$/c^2 \sim 10^{-8}$ g.

If one monopole is created in a horizon-size region, i.e., in a spherical volume of radius 6×10^{-26} cm, its density is

$$\rho_M = \frac{10^{-8}\mathrm{g}}{(4\pi/3)(6 \times 10^{-26})^3 \mathrm{cm}^3}.$$

The density is reduced at the present to

$$\rho_M = \frac{10^{-8}\mathrm{g}}{(4\pi/3)(6 \times 10^{-26})^3 \times (3 \times 10^{27})^3 \mathrm{cm}^3} = \frac{10^{-8}}{(4\pi/3)(180)^3}\mathrm{g\ cm}^{-3}$$

$$\cong 3 \times 10^{-15}\mathrm{g\ cm}^{-3}.$$

This value of ρ_M is far in excess of the cosmological density! Since the monopoles cannot be destroyed, their survival today would be unacceptable. Inflation reduces the above denisty by the factor Z^3, thereby making it insignificant.

In spite of these early successes, the idea of inflation has encountered difficulties of a different kind which require a reassessment of the concept *ab initio*. Two major difficulties are discussed here. First, the λ-term that drives inflation is extremely large compared to the λ-term discussed in Chapter 2. This means that the end of inflation leaves only a tiny remnant of the old λ behind, the magnitude of the latter being $\sim 10^{-108}$ of the former! Why this small number? If this were even slightly different it would vastly affect the dynamical behaviour of the universe including the value of the Hubble constant today. The second difficulty, which we will return to in the last chapter concerns structures that we see today. Can we relate the galaxies, clusters, superclusters, etc., to some seed-perturbations planted early on? The inflationary scenario has not yet succeeded in providing a self-consistent picture of such seeding process — despite numerous attempts to find one.

(iii) Quark–gluon transition The breakdown of GUT symmetry at energies of 10^{15} GeV leads to the separation of the strong interaction on the one hand from the electro-weak interaction on the other. The strong interaction deals with hadrons, i.e., with baryons like photons, neutrons, etc., and mesons like $\pi^0, \pi^\pm$, etc. The basic entity out of which hadrons are supposed to be built are *quarks*, which come in several types, flavours and colours. We will not go into details here except to add that three quarks make a baryon and a quark–antiquark pair makes a meson.

The quarks do not appear isolated, because a strong attractive force binds them. The force is supposed to be carried by particles called *gluons*. In the very early universe, at sufficiently high temperatures corresponding to particle energies $\gtrsim$ GeV, the quarks and gluons could have remained loosely bound like plasma. That is, we have a quark–gluon sea of plasma filling the universe. When the temperature dropped below the above threshold, the quarks combined to form hadrons.

This phase transition could, in the first approximation lead to a uniform distribution of hadrons, mainly the protons, neutrons and the pions. However, the possibility exists of a certain degree of non-uniformity leading to certain parts of the universe being rich in neutrons while others ending up rich in protons. We will return to this possiblity later in this chapter.

3.4 The early universe

We continue our chronological discussion within the sequence described in Section 3.2. However, broadly speaking, we now enter the era commonly identified with the 'early universe'. At this stage the universe has baryons and mesons as well as leptons in interactive equilibrium. How long will this picture last in view of the cooling of the universe due to expansion? Also, will some particles cease to interact with others in any significant manner as they slow down? We will consider such issues next.

(iv) Decoupling of neutrinos Neutrinos are particles with no charge, zero rest-mass and spin = 1/2. They interact very weakly with other particles and as such they are the first to decouple from the rest of the particles. This comes about in the following way.

The calculations of the electroweak theory tell us that the neutrino collision rate with other particles, β, depends on the temperature of the universe according to the formula

$$\beta = aT^5 \exp(-b/T), \qquad (3.43)$$

where a and b are constants. Thus, if the universe were static with a temperature T, then, given sufficient time, a population of neutrinos colliding with other particles would reach and subsequently maintain a state of thermodynamic equilibrium.

The universe, however, is not static. The early universe was expanding very fast, its rate of expansion being given by the then value of the Hubble constant :

$$H = \frac{\dot{S}}{S} = \frac{1}{2t} \propto T^2. \qquad (3.44)$$

Evidently, expansion by its very tendency to separate any two particles, inhibits collisions and the maintenance of thermodynamic equilibrium. We therefore have to compare β with H in order to decide whether collisions occur frequently enough in the expanding universe.

Figure 3.2 illustrates how H and β change with temperature. It is clear that at sufficiently high temperatures, β is high enough for thermodynamic equilibrium to be achieved and maintained. At lower temperatures, the collision rate cannot keep up with expansion, and the neutrinos cease to have contact with other particles. The critical temperature for this to happen is about 10^{10} K. Thus neutrinos, early on, when the universe was hotter than ten billion degrees, did collide and maintain thermodynamic equilibrium. Subsequently they became isolated and essentially decoupled from the rest of the universe.

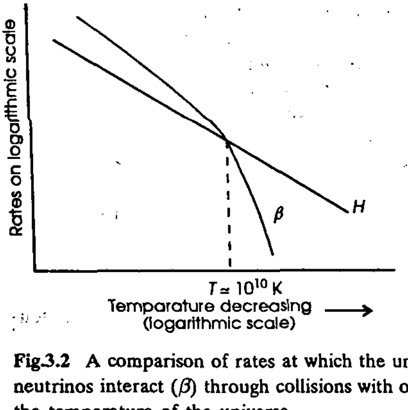


Fig.3.2 A comparison of rates at which the universe expands (H) and the neutrinos interact (β) through collisions with other leptons, plotted against the temperature of the universe.

What happens to the decoupled neutrinos? Although they no longer interact with other particles, they are still subject to gravity, and hence to the expansion of the universe. Their number density therefore drops, by a factor $1/S^3$, and their average momentum by a factor $1/S$. If they are really of zero rest mass, their distribution functions will also be reduced with expansion, such that their temperature will fall by a factor $1/S$. If, on the other hand, neutrinos do have a small rest mass, the above conclusion is altered. We will return to this point in Chapter 4, and assume for the time being that neutrinos are without rest mass.

(v) Annihilation of $e^{\pm}$ pairs By contrast to the weak interaction, the electromagnetic interaction is able to maintain a high enough collision rate even below the temperature of 10^{10} K. However, new considerations enter the picture when the temperature drops to half this value, for, (3.31) tells us that the value of T_X for electrons and positrons is comparable to this value. This is the temperature below which the electron–positron population begins to be non-relativistic. Moreover, during the cooling period from 10^{10} K to 10^9 K, we also have to take note of the tendency of electrons (e^-) and positrons (e^+) to annihilate each other and produce pairs of photons (γ):

$$e^- + e^+ \rightarrow \gamma + \gamma.$$

Of course, if a sufficiently large number of energetic photons is present, the reverse reaction also takes place:

$$\gamma + \gamma \rightarrow e^- + e^+.$$

In thermodynamic equilibrium at temperatures exceeding five billion degrees, both reactions were in fact taking place, thereby maintaining the population of electrons and positrons. As the temperature dropped below this value, however, the number of photons energetic enough to create electron–positron pairs declined sharply, and annihilations began to dominate. By the time the universe had cooled to about a billion degrees, the pairs had been effectively eliminated from the cosmic brew. The disappearance of the pairs led to an increase in the number of photons. The result was that the photon population came to have a higher temperature than the neutrinos. This increase can be seen by comparison with the temperature of the neutrinos.

Recall that at temperatures exceeding ten billion degrees the neutrinos were as much part of the thermodynamic equilibrium as the photons, and thus had the same temperature. Later, neutrinos ceased to partake in the 'equilibrium through collisions' process, although their temperature continued to decline by a factor $1/S$. Had no fresh photons been injected into the cosmic mix, the temperature of the photon population would also have continued to decline by a factor $1/S$, and would thus have remained equal to the temperature of the neutrino population. The electron–positron annihilation, however, raised the photon temperature (T_γ) above the neutrino temperature (T_ν). Exact calculation shows that

$$\frac{T_\gamma^3}{T_\nu^3} = \frac{11}{4}; \tag{3.45}$$

that is, $T_\gamma \approx 1.4T_\nu$. Assuming that after this annihilation event nothing happened to alter the numbers of photons and neutrinos in the universe, we would expect this ratio to remain valid to this day. Since the photon temperature today as exhibited by the microwave background is around 3 K, the neutrino temperature today would then be expected to be around 2 K. We will discuss the implications of this result when considering the possibility of neutrinos with non-zero rest mass.

(vi) Primordial nucleosynthesis The stage is now set for the universe to play the role of a thermonuclear reactor in which neutrons and protons at high temperature are brought together to form progressively bigger nuclei of atoms. This was the epoch George Gamow concentrated his investigations on, to evolve a theory of the origin of elements in the hot universe. First let us try to understand why the temperature range $10^9 - 10^8$ K is important for the synthesis of nuclei.

The simplest nucleus, one which needs no synthesis at all, is of course that of hydrogen. It consists of a single proton. The next one on the ladder is the heavy hydrogen (or deuterium) nucleus $^2_1$H, consisting of one proton and one neutron. We will follow the convention of denoting this nucleus, also called the 'deuteron' by the symbol d. Since it contains a neutron (n) and a proton (p) at rest, we would expect its mass, m_d, to be equal to the sum of the masses of the neutron and the proton, m_n and m_p respectively. In practice it is found to be somewhat less than this value, by an amount which we will designate Δm; thus,

$$m_d = m_n + m_p - \Delta m. \tag{3.46}$$

What is the cause of this deficit? It lies in the strong force which binds the nucleus together. It is the force of attraction which keeps the neutron and the proton together in the small nuclear region of size $\sim 10^{-15}$ m. The strong force has no effect at distances greater than this, so protons and neutrons well separated do not experience it. To strip a deuteron apart, however, we have to do work against this attractive nuclear force. The law of conservation of matter and energy requires that this work should appear as the excess energy which the free neutron and proton have over their bound-state energy. Called the 'binding energy', this excess energy, B, must of course equal Δmc^2. For the deuteron, it is 2.22 MeV.

Using our earlier discussion of the relationship between energy and temperature, we see that the binding energy of the deuterium nucleus translates into an equivalent temperature of 2.58×10^{10} K.

This figure suggests that if we have a hot gas of neutrons and protons at a temperature exceeding 2.58×10^{10} K, the binding imposed by the strong force is not adequate to hold a deuterium nucleus together. Collisions with fast-moving particles will strip it apart. Evidently we need a cooler temperature than this to form the deuteron. But how cool?

The actual value depends on the numerical distributions of neutrons and protons. So far, we have not considered these explicitly, because there were too few of them to affect the expansion of the universe at the temperatures under consideration. We now need to take note of their existence, because they are essential for the formation of nuclei. The first question we have to settle, therefore, is whether the neutrons and protons, being non-relativistic, are able to maintain thermodynamic equilibrium. Because if they are, we can say something definite about their relative abundances at any given temperature. This ratio, as we shall discover shortly, contains crucial, observable information.

It was the Japanese physicist Chushiro Hayashi who in 1950 first demonstrated that thermodynamic equilibrium between the neutron and proton populations is maintained through their collisions with electrons, positrons, and neutrinos. These 'collisions', of course, are not of the billiard-ball type, but are brought about by weak interaction. Thus processes like the following, which go both ways, were constantly taking place as the universe cooled below $T_0 = 10^{13}$ K and the neutrons and protons became non-relativistic :

$$\nu_e + \mathrm{n} \leftrightarrow \mathrm{e}^- + \mathrm{p}$$

$$\bar{\nu}_e + \mathrm{p} \leftrightarrow \bar{\nu}_e + \mathrm{p}$$

$$\nu_e + \mathrm{n} \leftrightarrow \nu_e + \mathrm{n}$$

Statistical mechanics then tells us that the ratio of the number densities of the neutrons and protons, N_n and N_p, at temperature T was

$$\frac{N_n}{N_p} = \exp\left(-\frac{1.5 \times 10^{10}}{T}\right). \qquad (3.47)$$

The slight difference between the rest-mass energies of the neutron and the proton is equivalent to an energy of 1.29 MeV, and hence to a temperature of 1.5×10^{10} K. At temperatures higher than this, N_n and N_p were almost equal. The ratio N_n/N_p dropped significantly, however, as the universe cooled below this temperature, until the actual numbers N_n and N_p themselves became so small that collisions became too infrequent to maintain thermodynamic equilibrium. Thereafter, the ratio dropped further for another reason : the one-way process of beta decay of the neutron given by

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

The lifetime of a neutron with regard to this process is about 1,013 seconds, which is longer than the age of the universe at the time this process was initiated, but still not too large to be ignored. When T drops so low that neutrons and protons are no longer in thermodynamic equilibrium, the above formula cannot be relied on to give the ratio N_n/N_p. Detailed book-keeping calculations of N_n and N_p taking into account the rates of various reactions which influenced them have to be done by computer. The interesting point is that a significant number of bound light nuclei begin to emerge when N_n/N_p reaches the value of $\sim 1/7$.

Although the deuteron was the first to form, it was not the stablest nucleus, and it subsequently grew into bigger units through reactions like the following :

$$d + d \leftrightarrow {}^3He + n$$

$${}^3He + n \leftrightarrow {}^3H + p$$

$${}^3H + d \leftrightarrow {}^4He + n$$

The process terminates to all intents and purposes when the stablest of all light nuclei, the ^{4}He nucleus is formed. This nucleus has a binding energy of 28.3 MeV, and all neutrons are taken up in forming it. Since the helium nucleus has two neutrons and two protons, it is easy to estimate the proportion of baryonic matter by mass, Y, which went to form ^{4}He. Given a neutron-to-proton ratio of 1/7 (see above), $Y \cong 1/4$. Here Y is called the mass fraction of helium; about a quarter of the total mass in the universe was thus in the form of helium (mass fractions for other nuclei are similarly defined).

The primordial nucleosynthesis cannot be continued beyond ${}^4_2$He in any significant way. Figure 3.3 shows the amounts made in typical models. A few light nuclei like d, ^{3}H, ^{3}He, ^{7}Li, and ^{11}Be are formed, but in much smaller fractions than for the ^{4}He nucleus (where $Y = 1/4$). This is because the next heavier nuclei after helium — lithium (Li), beryllium (Be), and so on — are not stable and revert to helium soon after they are formed. The stabler nuclei like carbon, oxygen, and so forth which lie beyond this gap of unstable nuclei, cannot be reached by this process of addition of neutrons and protons. It is in principle possible to produce carbon from three helium nuclei, as first pointed out by Fred Hoyle in 1954. However, after the first three minutes or so, the universe was not hot enough to bring about carbon production this way. Nonetheless, the process was shown by Hoyle to be possible inside stars and to hold the key to further nucleosynthesis of heavier nuclei in stars.

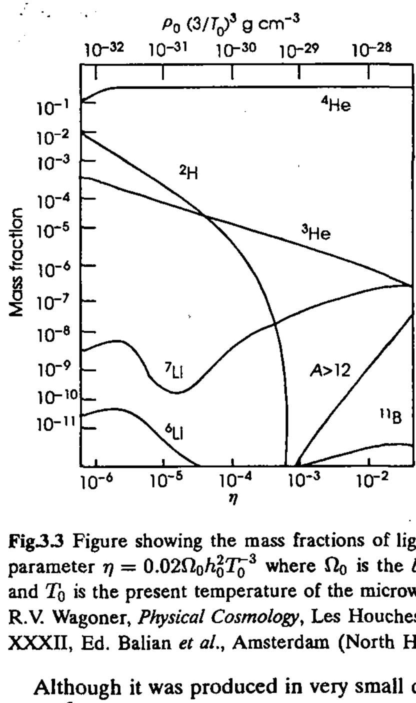


Fig.3.3 Figure showing the mass fractions of light nuclei plotted against a parameter $\eta = 0.02\Omega_0 h_0^2 T_0^{-3}$ where Ω_0 is the *baryonic* density parameter and T_0 is the present temperature of the microwave background. (Source : R.V. Wagoner, *Physical Cosmology*, Les Houches Summer School Lectures XXXII, Ed. Balian *et al.*, Amsterdam (North Holland) 1979).

Although it was produced in very small quantities (mass fraction $\leq 10^{-5}$), the deuteron holds important cosmological information. Its production, as seen from Fig.3.3, depends very sensitively on the number density of baryons in the universe. If there were far too many baryons around when nucleosynthesis was taking place, it would be easy for collisions of deuterons with neutrons and protons to take place, leading to the destruction of all the deuterons and their conversion into helium. Thus survival of deuterons is favoured in a low-density universe rather than a high-density universe. If we assume that most matter in the universe today is in the form of baryons, then the density parameter, Ω_0, can be linked with the primordial production of deuterons. For example, taking Hubble's constant as 75 kms^{-1} Mpc^{-1}, we find that in a universe with $\Omega_0 \geq 0.2$, practically no deuterons are produced.

Although Gamow's original aim of producing all observed atomic nuclei in the early universe did not materialize, the work on primordial nucleosynthesis was revived in the 1960s. Researchers like Ya. B. Zeldovich, F. Hoyle, R.J. Tayler, P.J.E. Peebles, W.A. Fowler, R.V. Wagoner, and others repeated Gamow's calculations with increasing sophistication. The reason for this revival of interest was that although stars were known to be likely sites for production of all nuclei from helium onwards, it was becoming increasingly apparent that stars cannot produce light nuclei in the quantities observed. It is precisely in this area that the early universe appears to do the trick.

Most striking in this context of course is the primordial production of helium, with a mass fraction Y of ~ 0.23. A simple calculation based on the amount of starlight produced in the galaxy from its birth to the present day provides an estimate of how much helium would have had to be synthesized in stars to generate that starlight; it comes out to a paltry value of $Y = 0.02$. This calculation of course uses the present luminosity of the galaxy and an age estimate of ten billion years. If the galaxy were considerably older or significantly brighter in the past than it is now, this estimate of Y would go up.

Actual observations of helium in several parts of the galaxy yield values of $Y \sim 0.24 - 0.28$. From these, the stellar contributions have to be subtracted. Thus a primordial contribution of $Y \sim 0.23$ seems to be about right. Notice that this value of Y, unlike the deuteron abundance, does not depend sensitively on the baryon number density.

Ultraviolet observations of absorption lines in the spectra of bright stars indicate that the deuteron mass fraction, denoted by $X(\mathrm{d})$, lies in the range $9 \times 10^{-6} - 3.5 \times 10^{-5}$. The satellite Copernicus, launched in 1973 to commemorate the quincentenary of the birth of Copernicus, gave the first reliable estimates of the abundance of deuterium in the universe. Since it is hard to come up with a stellar fusion process capable of producing even this quantity of deuterium, and since the deuteron, if created primordially, could have been destroyed subsequently in galactic processes, we need big bang models of low Ω_0 to produce at least the observed $X(\mathrm{d})$. For a Hubble constant in the range 50–100 kms^{-1} Mpc^{-1}, the value of Ω_0 cannot exceed the range 0.0375–0.15. Thus, if most matter in the universe is baryonic, the low value of $\Omega_0 (< 1)$ leads to the conclusion that the universe is open.

But this important, tight conclusion of the early 1970s has now developed a loophole! The above limit is on the *baryonic* density and not on the *total* density. Whether the universe is open or closed depends not only on the density of baryonic matter, but on the overall density of matter. Thus we could have a low baryonic component of matter as above, together with non-baryonic matter (massive neutrinos, say) resulting in $\Omega_0 \geq 1$. In this case the universe would still be closed. Whether this loophole is really operative, we will discuss later.

(vii) Decoupling of radiation from matter It is evident from the preceding discussion that light nuclei now found in the universe can be looked upon as relics of the early hot phase of the universe in much the same way that the archaeological findings at Pompeii are relics of a once flourishing city which was destroyed by the volcanic activity of nearby Mount Vesuvius. Similarly, the currently observed microwave background is a relic of a once radiation-dominated universe.

We recall that after the electrons and positrons had annihilated one another substantially and after the universe had cooled below a billion degrees, say, only radiation and neutrinos remained to control the dynamics of the universe, with the former taking on the lion's share of the job. Although the matter component of the universe in the form of baryons and left-over electrons had no effect on the cosmic expansion, they did however, continue to interact with radiation, especially the electrons (it is assumed here that there was an excess of electrons over positrons prior to annihilation. This assumption relates to a general excess of matter over antimatter, and will be discussed in the next chapter).

A typical electron scatters radiation very effectively. At low energies the scattering is by the so-called *Thomson process,* in which the energy of the incoming photon is not changed, although its direction of motion is. At high energies the scattering process is known as *Compton scattering*, which results in the electron receiving a significant part of the photon's initial energy. Since the rest energy of the electron is ~ 0.5 MeV, the Compton process is not effective for photon energies much below this value, and hence would not have been significant in the universe after it had cooled below a billion degrees.

The Thomson scattering would have been quite effective, however, so photons would have been scattered frequently. If such a situation existed in the present state of the universe, it would be impossible to study astronomy! For, photons from any astronomical source would not reach the telescopes of a remote observer intact — they would have been deflected too many times during their journey. That photons are able to make a journey of several billion light-years, bringing remote corners of the universe within the purview of our astronomical telescopes, is often expressed by the statement that the universe is *optically thin*. The early universe with its frequent scatterings of photons by electrons, by contrast, was *optically thick*.

Obviously, as the universe expanded and cooled, it changed at some stage from being optically thick to optically thin. When and how did this happen?

The answer to this question is provided, not surprisingly, by a process which removes free electrons from the cosmic brew. Remove the scatterers, and the radiation travels freely. The process occurred when the universe cooled to the temperature range of 3,000–4,000 K. At this temperature, free electrons combined with free protons to form electrically neutral hydrogen atoms.

This process is analogous to the binding of neutrons and protons into deuterium nuclei which took place earlier, at the much higher temperatures of $10^8 - 10^9$ K. In that case the *nuclear* binding force arising from the strong interaction was able to trap neutrons and protons. The chemical binding between an electron and a proton is *electrostatic* in nature, and therefore much weaker. Compared to the binding energy of the deuteron (2.22 MeV), the binding energy of the hydrogen atom is very small (only 13.59 eV). Thus, for electrons and protons to be trapped by this force, their speeds must be considerably smaller, and hence their temperature considerably lower, than in the earlier phase of the formation of the deuteron. Calculations show that, depending on the number density of electrons (which in turn can be related to the number density of protons *now present* in the universe), the bulk of the electrons were trapped into forming hydrogen atoms during the cooling of the universe from 4,000 to 3,000 K. Because this process is known as *radiative recombination*, this epoch is often called the *recombination epoch*.

In other words, by the time the universe had cooled to 3,000 K, it had become optically thin. Radiation then became decoupled from matter in much the same way that the neutrinos had earlier become decoupled from the other constituents of the universe. And like the neutrinos, the decoupled radiation preserved its equilibrium distribution of black-body type even afterwards, although with a temperature which fell by a factor $1/S$ as the universe expanded. It is therefore to be expected that the relic radiation observed today should have a black-body distribution.

As mentioned earlier, this expectation appears to be borne out by recent observations of the microwave background at different frequencies. Figure 3.4 presents the latest data on this count. The effective background temperature of 2.726 K today tells us that the decoupling of radiation from matter took place at the epoch in which the redshift was $\sim 1,000$.

We would make further progress with this explanation if we knew why the present-day temperature of the microwave background is ~ 3 K. Why is it not 7 K, as supposed by Gamow in 1953, or 5 K as proposed by Alpher and Herman in 1949? Although early universe calculations predict an almost unique value for the helium mass fraction, Y, they are not able to predict a value for the present-day temperature of the relic radiation.

This shortcoming is often expressed by observing that the photon-to-baryon ratio in the universe is not determined by the above calculation. At a temperature T_0, the formula of black-body radiation tells us that the number of photons per unit volume should be

$$N_\gamma = 19.2\pi\left(\frac{kT_0}{ch}\right)^3. \tag{3.48}$$

Thus, for $T_0 = 2.7$ K, the photon number density is ~ 400 per cm^3. If we assume that of all the matter present in the universe, a fraction f_B by mass is in the form of baryons, each with a typical mass m_P (the mass of the proton), then from the relations (2.34) and (2.36), we see that the number density of baryons in the universe at present is

$$N_B = \frac{3H_0^2\Omega_0}{8\pi G m_P} f_B. \tag{3.49}$$

After substituting numerical values for the various physical constants, we find the ratio of N_γ to N_B. Since the value of Hubble's constant is not known exactly, it is customary to express it as $H_0 = 100h_0$ kms^{-1} Mpc^{-1}, where the value of the dimensionless number, h_0, is believed to lie between 0.5 and 1. We then get from (3.48) and (3.49)

$$\frac{N_\gamma}{N_B} = 4.57 \times 10^7 (\Omega_0 f_B h_0^2)^{-1}\left(\frac{T_0}{3}\right)^3. \tag{3.50}$$

For example, for $\Omega_0 = 1$, $f_B = 0.1$, $h_0 = 0.75$ and $T_0 = 2.7$ K, this ratio is about 6×10^8. As the universe expands both N_γ and N_B fall off as $1/S^3$ and hence their ratio remains unchanged.

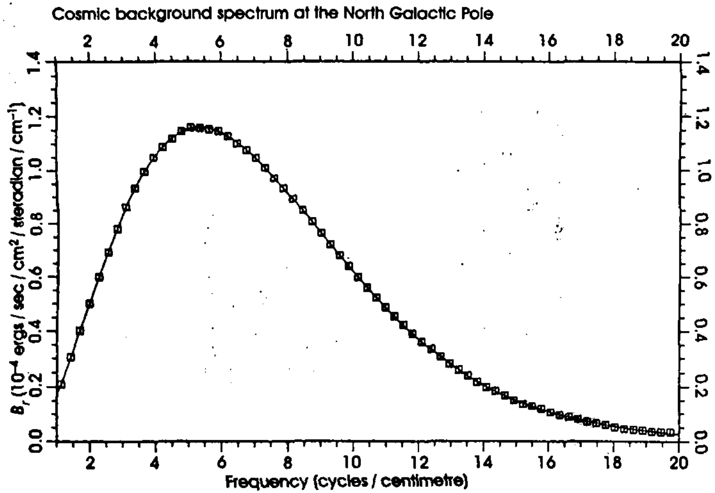


Fig.3.4 The spectrum of cosmic microwave background radiation obtained by the COBE satellite based on the work by J.C. Mather *et al.*, *The Astrophysical Journal Letters*, **354**, 37 (1990). The unit 'cycles/centimetre' is obtained by dividing the frequency by the speed of light.

Does this ratio have a significance ? Obviously the key to this ratio must lie in the history of the universe prior to the epochs we discussed earlier. In fact, questions like these prompted investigations of the very early universe that we went through in (i)–(iii).

3.5 Conclusion

We end our theoretical discussions here and pose questions any scientific theory must answer. What are the ways of testing the claims made in Chapters 2 and 3? How do we carry out measurements to test the validity of a dynamical model? How do we distinguish between the different values of the constants k and λ? How do we check on the numerous speculations of the early and the very early universe? We will try to answer these questions in the following chapter.

3.6 Exercises

3.1 Which of the following species of particles in the early universe would be moving relativistically in the early universe at a temperature of 10^{12}K?

$$e, p, n, \mu, \pi, \nu_e$$

3.2 Show that the flatness problem becomes worse as we push the epoch of initial conditions closer to $t = 0$.

3.3 Determine the time–temperature relation for the early universe if it contains only relativistic e^-, e^+ and the photons.

3.4 Using the Thomson scattering cross section for electrons, show that the optical depth of the universe at the present epoch would be $0.08\Omega_0 h_0$ if all the electrons in the universe were free and equal in number to baryons. Ω_0 is the baryonic density parameter.

4 *Observational Tests of Cosmological Models*

4.1 Introduction

We have discussed theoretical models of the universe; and in so doing we have extrapolated from the picture of the expanding universe *as at present*. To what extent are these model extrapolations correct? As in any branch of science, cosmological theories must be subject to observational proof. In this chapter, we will describe several such tests.

There are two kinds of tests we can perform to verify the cosmological ideas being discussed here.

a) Tests of the distant parts of the universe When we look at a galaxy at a distance r, we see it as it was r/c time ago. The observations of the distant parts of the universe therefore tell us about its past history. This information can be used to decide the viability of cosmological models. Of the tests in this category we will discuss the following.

(i) The redshift–magnitude relation
(ii) The counting of galaxies and radio sources
(iii) The angular diameter–redshift relation

b) Observations of the nearby universe Cosmological theories can tell us in a testable form what the state of the universe should be like in our neighbourhood. Thus we need not look very far to check the predictions of a cosmological model. The idea will become clear through our discussions of the following tests :

(i) The age of the universe
(ii) The density of matter
(iii) The radiaton background
(iv) Abundances of light nuclei

4.2 Tests of cosmological models based on observations of the distant parts of the universe

(i) The redshift–magnitude relation This relationship is also known

as the 'velocity–distance relation', and its use dates back to the very origin of modern cosmology. The present-day study of cosmology arose out of Hubble's observations of redshifts from distant galaxies. When the redshifts are plotted against the respective distances of these galaxies (from us), a straight line can be fitted to the plot. Moreover, if we relate the redshifts to velocities of recession (as discussed in Chapter 1), the observed relation may be written in the form

$$v = H_0 D, \tag{4.1}$$

where v is the velocity of recession, D is the distance away of a typical galaxy, and H_0 is Hubble's constant.

In the early 1930s, Hubble's observations extended to velocities of the order of 1000 kms^{-1}, that is, to redshifts not exceeding 0.004. Modern observations of galaxies go out to redshifts of $\gtrsim 2$, representing more than a five-hundredfold increase over the distance covered by Hubble's observations. What do these new observations tell us about the universe?

To examine this question, let us first consider the theoretical models. They all represent different dynamical behaviour. In general relativity (unlike in Newtonian gravity), they also describe different non-Euclidean geometries. Now, over a small-enough region, which in astronomical terms may mean upto a few megaparsecs (that is, up to the distances covered by Hubble), these models do not differ significantly among themselves. By looking out to greater and greater distances, however, where the differences between the various cosmological models become more and more pronounced, the astronomer hopes to pick out the one which matches the observations best.

These differences are reflected in the velocity–distance relation. Although all models predict the Hubble law as a linear (that is, a straight line) law on a graph of v versus D, for small v and D, all the models deviate from linearity for distant galaxies in different ways. This is illustrated by the graph shown in Fig.4.1. This graph does not plot velocity v against distance D but the redshift z against the apparent magnitude. The reason for this is as follows. The observer does not measure either the velocity or the distance directly. Instead he measures the redshift and the apparent magnitude. Now we have seen that redshift can be related to velocity, but the relation

$$v = cz \tag{4.2}$$

[see equation (1.7)] is only an approximate one, valid for small z. Indeed the formula is based on the concepts of Newtonian dynamics and Euclidean geometry — both of which are not applicable over the large distances covered in cosmology. Rather than talk of velocity, it is much better to relate theoretical discussions directly to the measured quantity, i.e., the redshift.

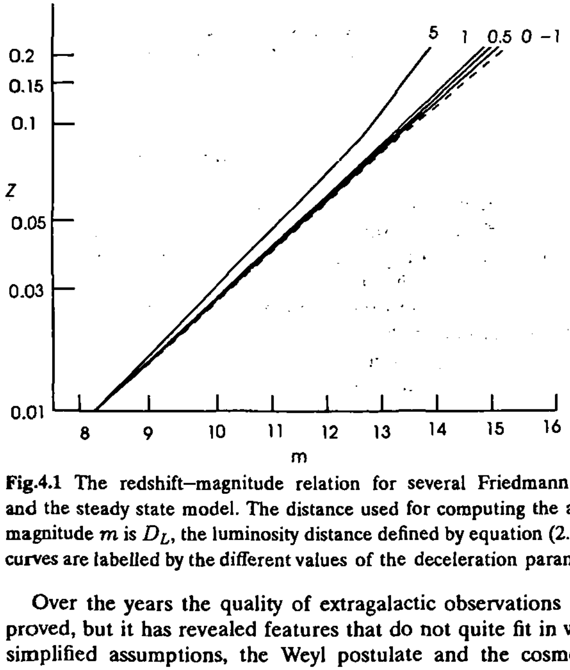


Fig.4.1 The redshift–magnitude relation for several Friedmann models and the steady state model. The distance used for computing the apparent magnitude m is D_L, the luminosity distance defined by equation (2.20). The curves are labelled by the different values of the deceleration parameter q_0.

Over the years the quality of extragalactic observations has improved, but it has revealed features that do not quite fit in with the simplified assumptions, the Weyl postulate and the cosmological principle, that form the basis of theoretical models. In particular, studies of the nearby universe extending to 10–50 Mpc have revealed large scale motions of galaxies *over and above* those of the expanding universe. Thus, the measurement of z does not necessarily mean that it is a measurement of the pure Hubble motion. For the nearby galaxies, at least a substantial part of it may be due to random, peculiar motion. It is not easy to extract the Hubble motion uniquely from the redshift measurements.

The relationship between apparent magnitude and distance is even more dubious. The apparent magnitude is a measure of distance on a logarithmic scale. *If two galaxies are equally luminous intrinsically*, the more distant of the two will have the larger apparent magnitude. However, two uncertainties are introduced at this stage.

First, the inverse-square law of illumination must be modified to take account of the expanding universe. We have seen how this modification is made in equation (2.19). So the appropriate modification must be calculated for each cosmological model separately. The different curves shown in Fig.4.2 are drawn on this basis. They are labelled by a parameter q_0, which we have defined in equation (2.38).

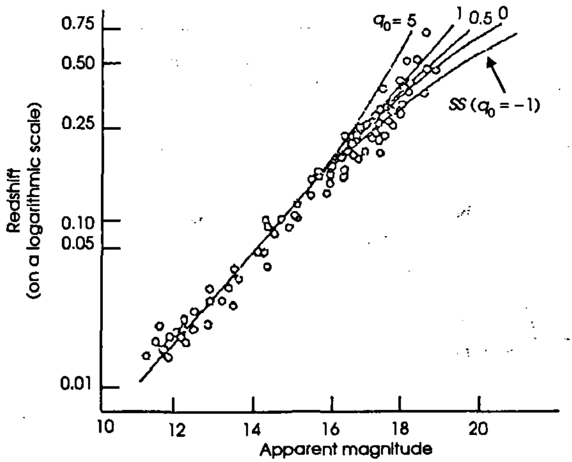


Fig.4.2 Redshift–magnitude relation for first ranked clusters number galaxies plotted along with the theoretical curves for the same cosmological models that were given in Fig.4.1. The observations are based on the work of J. Kristian *et al.*, *The Astrophysical Journal*, **221**, 383 (1978).

The second complication lies in the observations rather than in the theory. Are we sure that all galaxies are equally luminous intrinsically? Even a survey of the local region shows substantial variations — by factors of 1–100 — in the luminosities of galaxies. So in plotting their magnitudes we are liable to mistake a nearby but intrinsically faint galaxy for a distant but more luminous one. To avoid such mistakes as far as possible, Allan Sandage and his collaborators have restricted their surveys to massive elliptical galaxies which happen to be the brightest members of their clusters. It is found that the luminosities of such galaxies do not show much variation. In Fig.4.2 we see a redshift–magnitude plot using this subclass of galaxies. The relationship is fairly tight to justify confidence in Hubble's law.

In Fig.4.2 the plot is superposed on the theoretical plots for different models to see which theoretical curve fits the data best. On the basis of this, Sandage at first claimed that the curve labelled $q_0 = +1$ fits the data best. So Sandage's observations seemed to favour a model of the universe in which the expansion will slow down to a halt and will be changed to contraction. Returning to the Friedmann models, we find that $k > 0$ for a model of this type.

Problem 4.1 : Show that for the Einstein–de Sitter universe, the redshift–magnitude relation takes the form

$$m - M = 5\log\left(\frac{2c}{H_0}\right)_{pc} - 5 + 5\log\{\sqrt{z+1}(\sqrt{1+z}-1)\}.$$

Solution : Taking the scale factor from (2.32) as $S(t) = (t/t_0)^{2/3}S_0$, from (2.14) we get the time taken by light to travel from a galaxy with comoving coordinate $a > 0$ to the origin as $(t_0 - t)$, where

$$a = \int_t^{t_0} \frac{c\,dt}{S_0(t/t_0)^{2/3}} = \frac{3ct_0^{2/3}}{S_0}\{t_0^{1/3} - t^{1/3}\}.$$

The redshift of the galaxy is given by

$$1 + z = \left(\frac{t_0}{t}\right)^{2/3}.$$

Therefore, the luminosity distance is

$$D_L = S_0 a(1+z) \quad = 3ct_0\left\{1 - \frac{1}{\sqrt{1+z}}\right\}(1+z)$$

$$= \frac{2c}{H_0}\sqrt{1+z}(\sqrt{1+z}-1).$$

Since, with D_L expressed in parsecs,

$$m - M = 5\log D_L - 5,$$

the result follows provided c/H_0 is expressed in parsecs.

What are the underlying uncertainties behind this conclusion? First we note that the theoretical curves really begin to deviate at redshifts in excess of $z = 0.2$. So if several points were available on this graph beyond $z = 0.2$, we could attach a certain measure of confidence to this conclusion. Unfortunately the scatter is wide enough to admit a wide band of q and so one cannot assert that $q_0 = 1$ with any confidence. The steady state model has $q_0 = -1$, and even this case cannot be ruled out on the strength of the present data.

Another possible complication could arise from a significant change in the luminosity of a galaxy with redshift. B. Tinsley, from the University of Texas, and her colleagues made a case for a decrease in luminosity of a galaxy with age. This has the effect of lowering the earlier estimate $q_0 = 1$ of Sandage. Many astronomers, including Sandage, now believe that q_0 may be very close to zero. Such a conclusion would not be consistent with the inflationary scenario, unless, as seen in *Problem 2.4* we incorporate a λ-term.

For a clear-cut result we require a class of objects without a great deal of variation in luminosity but with redshift values far in excess of those found for galaxies. When quasi-stellar objects (QSOs) were first discovered, many with redshifts close to $z = 2$, it was hoped that they would settle the cosmological issue once for all. Today nearly 7000 QSOs are known, with a maximum redshift amongst them of $z = 4.92$. Have QSOs settled the problem as expected?

Unfortunately QSOs have made the issue more complicated than before. The Hubble diagram for QSOs, far from giving a linear redshift–magnitude relation, appears to be a perfect scatter diagram, with no hint even of a connection between redshifts and apparent magnitudes, let alone an indication of the value of q. If at all we want to reconcile this with Hubble's law we will have to argue that there is an enormous scatter in the intrinsic luminosities of QSOs, and this explanation finds favour with many astronomers. But if this interpretation is accepted, QSOs cannot be relied upon to pick out the best cosmological model. To put it another way, imagine an inversion in the chronological order in which the redshifts of QSOs and galaxies were discovered. Suppose that in the late 1920s Hubble had come across QSOs and their big redshifts. Would he, on the basis of the Hubble diagram of QSOs, have concluded that they satisfy a velocity–distance relation? Indeed no unprejudiced observer, on the basis of this data, could have come to such a conclusion. It is because such a relation was found for galaxies in the first place that the astronomer is tempted to suppose that it applies to QSOs also, but we cannot rule out the possibility of the redshifts of QSOs containing a substantial non-cosmological component.

To summarize, a somewhat frustrating situation exists with regard to this test. Theoretically, the different models predict really different results in the high-redshift range. However, in this range, where the evidence from QSOs might have been decisive, a complete anarchy exists. In the low-redshift range the galaxies do obey a systematic Hubble law, but here the observations cannot really decide which cosmological model is best suited, because most models of interest do not predict significantly different results.

(ii) The counting of galaxies and radio sources Besides the redshift–magnitude relation, Hubble also undertook another cosmological test. This involved counts of galaxies up to different distances. The basic idea of the test is simple. If we suppose that (1) we live in a static Euclidean universe, (2) there is a uniform distribution of galaxies in the universe, and (3) the galaxies are of the same luminosity, then we can perform the following simple calculation:

Let there be n galaxies per unit volume, and let each one have a luminosity L. Then the number of galaxies in a spherical region of radius R around us will be

$$N = \frac{4\pi}{3} R^3 n. \tag{4.3}$$

The faintest of these as seen from here will be those at distance R, and the rate at which light is received from any one of them over a unit area here will be

$$F = \frac{L}{4\pi R^2}. \tag{4.4}$$

From these we get the relation

$$F^3 N^2 = \frac{1}{36\pi} n^2 L^3 = \text{constant}. \tag{4.5}$$

That is, if we measure the number N of galaxies brighter than F, the product N^2F^3 should stay constant for varying levels of brightness.

If the product N^2F^3 does not appear to be constant, some or all of the assumptions (1)–(3) above must be wrong. All of these have a cosmological bearing, because they relate to the state of the universe in the past. Hubble's attempt in this direction in the 1930s did not succeed largely because the number of galaxies is fantastically large. In order to test any possible departure from the

Euclidean nature of geometry, observations must cover large regions. It is estimated that some hundred million galaxies may have to be counted to get any significant information of this type!

Modern technology with computer automation and image processing has revived interest in galaxy counts as a cosmological test. With magnitudes used instead of F, we have

$$m = -2.5 \log F + \text{ constant}, \tag{4.6}$$

and the relation (4.5) becomes

$$\log N = 0.6m + \text{ constant}. \tag{4.7}$$

In a typical expanding model, (4.7) should be the relation at the bright end, becoming progressively less steep as we probe fainter and fainter samples. While this trend is largely borne out, the result is not clearcut enough to point to a specific q_0. In fact many observers believe that evolution at the faint end is needed to reconcile the data with the model predictions. That means we have to drop all the assumptions (1)–(3) stated earlier.

When the extragalactic nature of most radio sources became established, Martin Ryle and his collaborators at Cambridge, England, set out to perform the above test for radio sources. Since a radio source is comparatively rarer than a galaxy — only a few galaxies in a million are strong radio sources — the job of counting them to a large distance is simplified.

For radio sources the quantity F is called the 'flux density', and it limits the radiation being measured to a narrow band of frequencies. The unit of flux density is Jansky, named after Karl Jansky who pioneered radio astronomy in the 1930s. One Jansky (1 Jy) is equal to 10^{-26} Watts per square metre per Hertz — Hertz (Hz) being the frequency band of 1 per second. The spectral function of a radio source is often expressed as a power law of frequency $(\nu), I(\nu) \propto \nu^{-\alpha}$, where α is called the *spectral index*.

Problem 4.2 : A radio galaxy of redshift $z = 0.1$ has a spectral index $\alpha = 1$ and luminosity of 10^{44} erg s^{-1} over the frequency range $150 \leq \nu \leq 1500$ MHz. Show that for a Hubble constant of 100 kms^{-1} Mpc^{-1} the flux density of the galaxy at 1000 MHz is $\sim$ 360 Jy (ignore cosmological effects).

Solution : Suppose the galaxy radiates $K\nu^{-1}d\nu$ erg s^{-1} over $(\nu, \nu+d\nu)$

in the given frequency band. Then the total luminosity over this band is

$$10^{44} = \int_{\nu_1}^{\nu_2} \frac{K}{\nu} d\nu = K ln \frac{\nu_2}{\nu_1}, \quad \text{with} \quad \nu_1 = 150 \text{ MHz}, \quad \nu_2 = 1500 \text{ MHz}$$

$$= 2.3\text{K},$$

i.e., $K \cong 4.3 \times 10^{43}$.

From equation (1.8) we get the distance of the galaxy as

$$D = \frac{c}{H_0} \times 0.1 \cong 300 \text{ Mpc} \approx 10^{25} m.$$

The galaxy radiates $K\nu^{-1}$ erg s^{-1} over $(\nu_1 \nu + d\nu)$. Setting $\nu = 10^3$ MHz $= 10^9$ Hz and $d\nu = 1$ Hz we get the flux density as

$$\frac{K\nu^{-1}}{4\pi D^2} \cong \frac{4.3 \times 10^{43} \text{ erg s}^{-1} \text{ Hz}^{-1}}{10^9 \times 12 \times 10^{50} \text{m}^2} \cong 3.6 \times 10^{-24} \text{Wm}^{-2} \text{ Hz}^{-1} \cong 360 \text{ Jy}.$$

The radio-astronomer plots log N against log F to test the constancy of N^2F^3. If N^2F^3 is constant, the plot should represent a straight line with a slope of -1.5, as shown in Fig.4.3. In the same figure is shown the line obtained by Ryle and his group from their survey. This line has a slope of -1.8, a significant departure from the -1.5 predicted on the basis of assumptions (1)–(3). Assuming for the moment that (1) and (3) are still valid, the implication of this result is that the number n increases as we look at farther regions, that is, as we look at regions further back in time (because radio waves take some time to travel from the source to the receiver). This would imply evolution in number density.

Alternatively, we could argue in the following way. Using one of the theoretical cosmological models we could calculate a log N versus log F curve on the basis of assumptions (2) and (3). The third curve shown in Fig.4.2 is that for the steady-state model. It begins with a slope of -1.5 and becomes progressively flatter. The big bang models give similar curves. Thus it would mean that if we abandon Euclidean geometry in favour of the non-Euclidean geometries of the expanding cosmological models, the gap between theoretical predictions and observations widens still further. The gap can be plugged in the case of the big bang models by asserting that n increased in the past. Indeed close agreement between theory and observation can be obtained in this way by requiring n to be larger in the past. This latitude is denied to the steady state theory however. This theory requires the universe to be unchanging over long periods of time. So n cannot increase either in the past or in the future. For this reason it was claimed by the Cambridge radio astronomers that the radio-source count disproves the steady state theory of the universe. To what extent is this claim justified? In this connection the following point can be made.

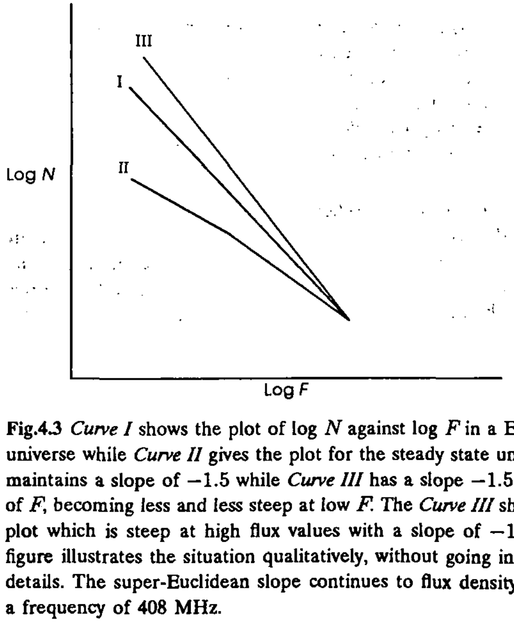


Fig.4.3 *Curve I* shows the plot of log N against log F in a Euclidean static universe while *Curve II* gives the plot for the steady state universe. *Curve I* maintains a slope of -1.5 while *Curve III* has a slope -1.5 at high values of F, becoming less and less steep at low F. The *Curve III* shows the actual plot which is steep at high flux values with a slope of -1.8. The above figure illustrates the situation qualitatively, without going into quantitative details. The super-Euclidean slope continues to flux density of ~ 2 Jy, at a frequency of 408 MHz.

The assumption (3) of constant luminosity is a very dubious one. In general, sources vary considerably in their total output of radiation. Thus a source with a low value of F may either be a strong but distant source or a weak but nearby source. How can this be decided? If the object is identified with a galaxy and its redshift measured, then that, with the help of Hubble's law, gives information about its distance. So on this basis an increase in redshift should be related to a decrease in flux level, if the latter is an indicator of distance. A plot of redshift against flux level does not reveal any correlation. In 1989 Patrick Das Gupta, Geoffrey

Burbidge and the author examined the 3C survey for which the optical identification is nearly complete. They found that when the additional information of redshifts is taken into account, there is no need to discard the hypothesis of *no evolution.*

In a more recent application of this test, Hoyle, Burbidge and the author demonstrated that the source counts fit the quasi-steady state cosmology without having to invoke evolution. In this case the steepness of the log N–log F curve is explained as due to sources from the previous cycle.

For these reasons the radio-source count has not turned out to be such a crucial cosmological test as it originally promised to be. The test is still potentially useful, but it must be applied with caution. Its application must wait until our understanding of the nature of radio sources, their optical identification, redshift measurements, etc., have considerably improved.

(iii) Angular diameter–redshift relation : In 1959 Fred Hoyle drew attention to a rather curious observable property of the non-Euclidean geometries used in cosmology. To understand this let us first consider what happens when we look at a galaxy or a radio source from successively increasing distances in a universe which obeys the laws of Euclid's geometry. To simplify matters consider a spherical object of diameter d. When we look at it from a *distant* point O, it subtends an angle $AOB = \theta$ at O, where AB is the linear diameter of the object. With $AB = d$ and the distance of O from the centre of the object as D, we have approximately,

$$\theta = \frac{d}{D}, \tag{4.8}$$

provided θ is measured in radians. This means that if we increase D, θ decreases. This is shown by the straight line of Fig.4.4(a), which is a plot of log θ against log D. The astronomer, looking at a distant object, measures θ directly, whereas his measurements of D and d are somewhat indirect, if they are at all possible. If we are looking at a lot of similar objects located at different distances, then we should observe the relation shown in Fig.4.4(a) between θ and D.

What happens in an expanding cosmological model? We have already seen that the redshift z is an indicator of distance if we accept Hubble's law. So if we examine angular diameters θ of like objects of increasing redshifts, what would we expect to find? Do we see θ decreasing to zero as we increase z to infinity? The answer, as calculated by Hoyle, was no. A typical case is described in Fig.4.4(b).

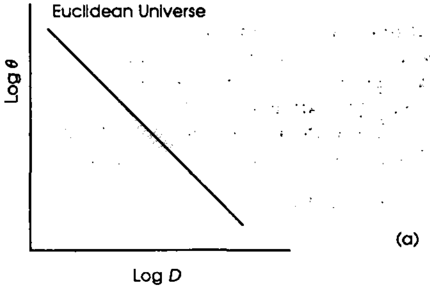


Fig.4.4(a) In **(a)** we see that θ plotted against D decreases monotonically as $1/D$ in Euclid's geometry.

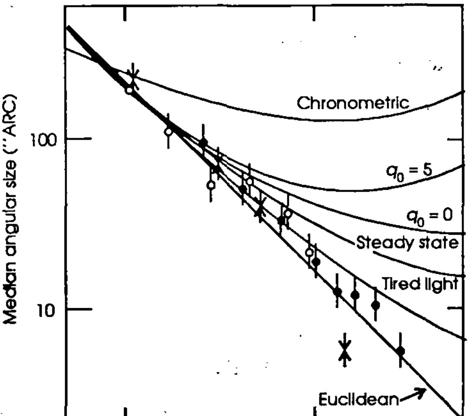


Fig.4.4(b) In **(b)** θ is plotted by a continuous curve against the redshift z in an expanding universe. Although D increases with z monotonically, θ shows a strange behaviour. It decreases to a minimum value and then starts increasing again. In the above figure several examples are shown of different cosmological predictions. In the Einstein–de Sitter universe the minimum value of θ occurs at $z = 1.25$. The observed behaviour of the median value of θ against redshift for a population of radio sources is shown by way of comparison. (Source : V.K. Kapahi, *Observational Cosmology*, IAU Symposium No. 124, Ed. A. Hewitt *et al.*, Dordrecht : Reidel, p.251 (1986).)

The argument goes like this. Imagine that we observe a spherical galaxy of diameter d at redshift z. Let the epoch when light left it be t_1, the present epoch of observation being t_0. The diameter d is measured in proper length units. However, in comoving coordinates (see equation (2.11)) the extremities of the diameter have a coordinate difference $d/S(t_1)$. If the radial coordinate of the galaxy is r_1, then the angle θ subtended by it at the observer is

$$\theta = \frac{d}{r_1 S(t_1)} = \frac{d(1+z)}{r_1 S(t_0)} = \frac{d(1+z)}{D}. \tag{4.9}$$

Thus as we look at more and more distant galaxies, their D-values increase; but so do their z-values. Hence, we may not get a monotonic decrease of θ with z.

Problem 4.3: Find the θ–z relation for the Einstein–de Sitter model. Does θ have a minimum?

Solution : From *Problem 4.2* we get

$$D = \frac{2c}{H_0}\left\{1 - \frac{1}{\sqrt{1+z}}\right\}.$$

Hence, relation (4.9) becomes

$$\theta = \frac{dH_0}{2c}\frac{(1+z)^{3/2}}{\sqrt{1+z}-1}.$$

This expression has a minimum at $z = 1.25$.

In the Einstein–de Sitter model the angular diameter first decreases to a *minimum non-zero value* as z approaches the critical value 1.25. Beyond this value θ begins to increase. Thus, the more distant the object the bigger is the angle subtended by it at the observer, provided its redshift exceeds 1.25 (see *Problem 4.3*).

This result could be a test of cosmological models provided we can apply it to a class of objects of fixed (or nearly fixed) linear size and large redshifts. The latter requirement is met by QSOs, but the former is more difficult to satisfy. QSOs exhibit large variations in their intrinsic size. If later, some subclass of QSOs turns out to be of nearly-fixed linear dimensions, this test can either be used to decide between different cosmological models or to determine whether or not QSO redshifts arise from the expansion of the universe. Present surveys show that $< \theta >$ decreases steadily as z increases, more or less as z^{-1}. Here $< \theta >$ is the median angular size at a given redshift. To explain this result one has to assume evolution — with the linear size decreasing as z increases (see Fig.4.4(b)).

4.3 Observations from the nearby universe

(i) The age of the universe If the universe began with a big bang, how long ago did this big bang take place? This period, as measured by cosmic time, may be called the age of the universe.

In the case of Friedmann models, we know the expansion factor as a function of cosmic time. If we start measuring time from the big bang, the present value of the time coordinate t_0 will be the age of the universe. This calculation is a relatively straightforward one, and the result depends on the measured values of Hubble's constant and the deceleration parameter. Figure 4.5 shows the variation of

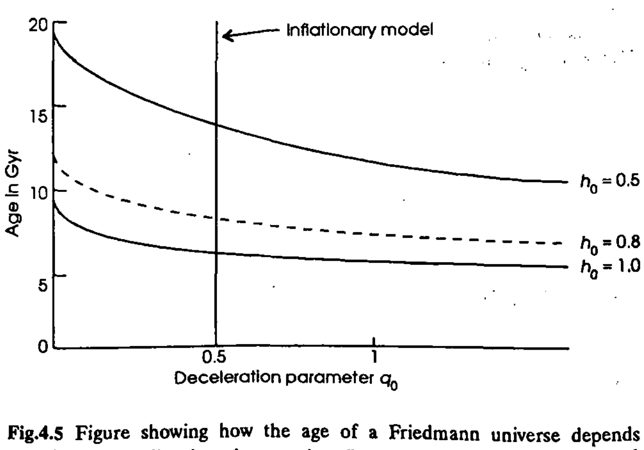


Fig.4.5 Figure showing how the age of a Friedmann universe depends on H_0 and q_0. The dotted curve describes the ages for $H_0 = 80$ kms^{-1} Mpc^{-1}, the value based on measurements of Cepheid variable stars in a Virgo cluster galaxy by the Hubble Space Telescope. The age according to the inflationary model is also shown.

the 'age' of the universe with q_0. Since we do not know q_0 with any degree of confidence, it is not possible to fix the age accurately. For a Hubble constant of 50 kms^{-1} Mpc^{-1}, the maximum value of t_0 is about 20 billion years for $q_0 = 0$. The value for $q_0 = 1$ is about 11.5 billion years. These values are reduced by half if $H_0 = 100$ kms^{-1} Mpc^{-1}. Let us see how this age band compares with the galactic ages.

The age of the galaxy may be estimated in two ways. One method makes use of the ideas of stellar evolution which tell us that a typical star spends most part of its life in the main sequence while it is deriving its energy of radiation by the fusion of hydrogen to helium. A star of mass M has a definite luminosity $L(M)$ and surface temperature $T(M)$ at this stage. The 'main sequence' is the curve on which stars of different M lie in the (T, L) diagram. Astronomers plot such stars on a $(-\log T, \log L)$ diagram, commonly known as the Hertzsprung–Russell diagram or the *H–R diagram*.

Figure 4.6 shows a schematic H–R diagram with the main sequence extending from A to B. When a typical star at C consumes all its fuseable hydrogen, it switches to a new process wherein helium is fused to carbon. In the subsequent evolution, the star core goes through a succession of fusion reactions, generating energy and synthesizing heavier nuclei. On the H–R diagram it no longer stays at one point C but becomes progressively cooler and more luminous. Thus, it branches off to the right along the curve CD on the Hertzsprung–Russell diagram of Fig.4.6. The rate at which the star

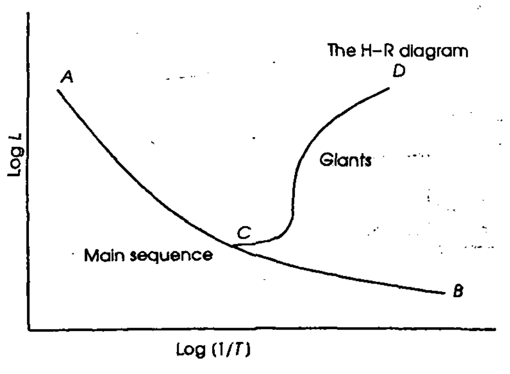


Fig.4.6 A schematic H–R diagram showing the main sequence AB and a star at C which moves to the right along the curve CD after its hydrogen burning stage is over.

covers these tracks on the diagram depends on its mass. Suppose now that we have a cluster of stars in the galaxy, all of which have branched to the right and are lying fairly close together on the H–R diagram. Then these stars must be approximately of the same age, which may be called the age of the cluster. The galaxy must be at least as old as the oldest cluster found in it.

Using this method the ages of globular clusters in the galaxy have been estimated and the figure of 12–18 billion million years has been arrived at as some kind of an average. Clusters considerably older than this are believed to exist, but detailed calculations are needed to confirm this result. So the above figure should be taken as a conservative lower limit for the age of the galaxy.

The second method makes use of the long time scales of radioactive decays of certain nuclei. For instance, the mean lives of thorium (^{232}Th) and the uranium isotopes (^{235}U and ^{238}U) are known, and they lie between 1–20 billion years. Suppose that the ratios of the abundance of these nuclei were known at the time of production and that these ratios are also known now. Then, since the nuclei decay at different rates, we can compute the time taken from production to the present epoch. The production of these nuclei is believed to take place in supernovae by the addition of neutrons. So theoretically their production ratios may be estimated and compared with present observed ratios in the neighbourhood of the Sun. In this way the decay times were estimated by Fowler and Hoyle as about $\sim$ 15 billion years.

The fact that both the estimates yield results of the same order is encouraging, since the two methods are quite unrelated. But, for the larger value of Hubble's constant these ages are clearly inconsistent with most big bang models. For example, the 1994 measurements by the Hubble Space Telescope of distances to Cepheid variable stars in another galaxy close to the Virgo cluster suggest that $H_0 = 80$ kms^{-1} Mpc^{-1}. The age curve for this value is shown in Fig.4.5 by a dotted line. Certainly the inflationary model which has age $= 2/3H_0$ is in trouble on this count, and may need the λ-term to prop it up!

(ii) The density of matter In Chapter 2 we found that the density parameter Ω defines the mean density of matter in the universe as :

$$\rho_0 = \frac{3H_0^2}{8\pi G}\Omega_0 \cong 2 \times 10^{-29}\Omega_0 h_0^2 \text{ g cm}^{-3}. \tag{4.10}$$

So for $H_0 = 100$ kms^{-1}Mpc^{-1}, $\Omega_0 = 1$; the mean density is $\sim 2 \times 10^{-29}$g cm^{-3}.

By the early 1970s, estimates were made of mass contained in the form of galaxies and the intergalactic medium, and it was found that $\rho_0 \sim 1 - 3 \times 10^{-31}$g cm^{-3}; thus clearly indicating that $\Omega_0 < 1$. Translated in terms of q_0 this result favoured the open, ever-expanding ($k < 0$) models. The deuterium abundance [see (iv) of this section on page 79] also favoured such low density models.

From the late 1970s, however, there began to appear growing indications of 'hidden mass', that is, mass that had practically no interaction with radiation but whose existence could be inferred by its gravitational effect on luminous matter. The first suggestion that dark matter might exist in substantial quantities in clusters of galaxies had, however, come much earlier from F. Zwicky in 1933. Today there are two separate lines of evidence for this dark matter.

First, the 21-cm measurements show that the rotational velocities of neutral hydrogen (H-1) clouds located beyond the visible parts of spiral (disc like) galaxies show no signs of falling off with distance from galactic centres. Keplerian orbits should show a drop of rotational velocity v with distance r as per the formula

$$v = \sqrt{\frac{GM}{r}}\ , \qquad r \geq r_0. \tag{4.11}$$

Here M is the gravitating mass within distance r. Thus we should expect $v \propto r^{-1/2}$. Instead, as shown in Fig.4.7, $v \cong$ constant.

To reconcile with this fact we need to assume that M is not constant, but $M \propto r$..That is, the galaxy has a gravitating mass even beyond its visible extent $r = r_0$. This mass has to be dark.

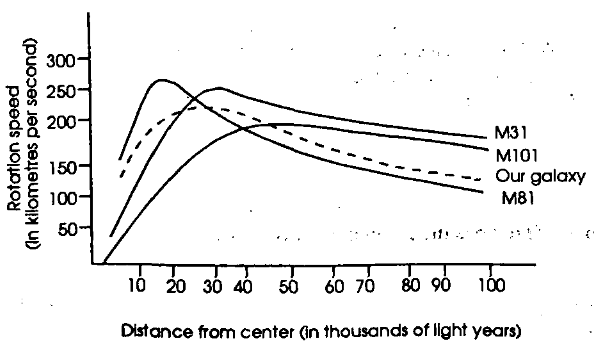


Fig.4.7 The observed flat rotation curves (continuous lines) seen in several spiral galaxies.

The other line of evidence comes from the observations of clusters of galaxies. If in a typical cluster there are N galaxies with masses $m_1 \ldots m_N$ at locations $\mathbf{r}_1 \ldots \mathbf{r}_N$, then their kinetic and potential energies are

$$T = \sum_{i=1}^{N} \frac{1}{2} m_i \mathbf{r}_i^2, \quad \Phi = -\sum_{i<j}\sum \frac{m_i m_j}{|\mathbf{r}_i - \mathbf{r}_j|}. \tag{4.12}$$

Now in a 'relaxed' cluster (i.e., wherein several close encounters and redistributions of momenta have taken place) a state of equilibrium is expected to be reached with

$$2T + \Phi = 0. \tag{4.13}$$

This is known as the *virial theorem*. In practice what is found is

$$|2T| \gg |\Phi|. \tag{4.14}$$

This could mean either that the cluster is not relaxed, or that there is a lot of dark matter around whose contribution to $|\Phi|$ has been ignored in (4.14). Most astronomers opt for the latter view and conclude that considerable dark matter is present in the intergalactic medium of a typical cluster.

How much does this increase Ω? The estimates of dark matter are many. Although the above effects raise the value of Ω well above the visible contribution, they still keep it below 1 as required by inflationary cosmology. Also, if there is so much dark matter that deuterium abundance is adversely affected, then the alternative is to make it *non-baryonic*, i.e., of matter that does not take part in primordial nucleosynthesis. It is too early to say whether $\Omega \gtrsim 1$ and whether non-baryonic matter must be present.

Several types of non-baryonic matter including massive neutrinos, axions, gravitinos, photinos, etc., have been conjectured by theoreticians. *There is no laboratory evidence of their existence* except that controversial evidence does exist for neutrinos with masses $\leq 30\text{eV}/c^2$.

Let us briefly look at the massive neutrino option for dark matter. We have seen that the relic neutrino background today will have a thermal distribution with temperature T_ν which is related to the photon background temperature T_γ by the relation (3.45). Correspondingly, the neutrino number density N_ν will be related to the photon number density N_γ by the relation

$$\frac{N_\nu}{N_\gamma} = \frac{3}{22}. \tag{4.15}$$

All this was based on the assumption of massless neutrinos. In 1972, R. Cowsik and J. McClelland examined the implications of this relation if neutrinos have a small rest mass. For massive neutrinos the difference is that if their masses exceed $\sim 2 \times 10^{-4}$ eV/c^2, they will have very small velocities relative to the expanding universe. Thus, we may assume them to be more or less at rest with density but negligible pressure. The total density contributed by neutrinos will then work out to

$$\rho_\nu = N_\nu M_\nu, \tag{4.16}$$

where M_ν is the neutrino mass. Calculations show that the density parameter Ω_ν contributed by neutrinos is

$$\Omega_\nu = \frac{M_\nu}{150}\left(\frac{T_0}{3}\right)^3 h_0^{-2}, \tag{4.17}$$

where M_ν is the mass of the neutrino expressed in electron-volts, T_0 is the present temperature of the microwave background and the Hubble constant is $100h_0$ kms^{-1} Mpc^{-1}.

If there are three species of neutrinos (ν_e, ν_μ and ν_τ) having corresponding antineutrinos, then with M_ν constant for all, we get $\Sigma\Omega_\nu \approx 1$ for $M_\nu = 25$.

Recall that the neutrinos decoupled from the rest of the matter at temperature $T_\gamma > 10^{10}$ K and so they were moving relativistically, i.e., they were 'hot' when they decoupled. Such dark matter is called 'hot dark matter' (HDM). Other forms of matter like photinos, gravitinos or axions are examples of 'cold dark matter' (CDM) — matter which decoupled from baryons when it was at low kinetic energy.

(iii) The microwave background We come now to what is regarded as the strongest evidence of all for the big bang model.

In 1965, A.A. Penzias and R.W. Wilson, two scientists from the Bell Telephone Laboratories in the United States, detected a background radiation in the microwave region (at a wavelength of about 7 cm). The discovery, like many other important discoveries in astronomy, came about accidentally. Penzias and Wilson were, in fact, making radio-astronomical measurements using an antenna originally designed to receive signals reflected from the 'Echo' satellites. The radiation they measured at 7 cm wavelength could be mostly accounted for by other sources such as atmospheric emission and ground emission. However, some residual radiation remained unaccounted for; and, more important, it was isotropic in character. That is, it did not depend on any particular direction, implying that its source or sources, if any, did not exist close by (for example, in our galaxy).

That such radiation should exist did not come as a surprise to the theoreticians, who already had a possible explanation for it. In the big bang theory, a few moments after its origin, the universe is dominated by radiation rather than matter. The radiation is that of a black body and has a temperature in excess of 10^{10} K one second after the big bang. However, as the universe expands, the temperature drops sharply, and it was estimated by Ralph Alpher and Robert Hermann in 1948 that it would be around five degrees at the present epoch. At this temperature the maximum radiation occurs in the millimetre range of wavelengths. So the detection of radiation at wavelengths of a few centimetres was not surprising. If the radiation observed by Penzias and Wilson was indeed a relic of the big bang, then its spectrum was expected to be that of a black body. Assuming it was so, the temperature could be estimated. This turned out to be about 3 K.

How can we make sure that the radiation is indeed of a black-body type? This can be done by measuring the intensity at various wavelengths. The present status of the observations in relation to the black-body curve is shown in Fig.3.4. The agreement with the black-body curve is very good. The peak itself lies near 1 mm, corresponding to a temperature of 2.73 K. Measurements close to this wavelength and at shorter wavelengths cannot be obtained from ground-based astronomy. Instead we must use balloons, rockets and satellites to make measurements above the atmosphere. Figure 3.4 shows the most comprehensive data of this kind which has come from the Cosmic Background Explorer (COBE) satellite launched in 1989.

Indirect observations based on transitions in cyanogen (CN) molecules also provide additional checks and confirm the existence of this radiation far away from the Earth. When the CN molecule is non-rotating, it has an absorption line at 3.79 cm^{-1}. In the first rotational state the wavelength is 7.58 cm^{-1}. When radiation of appropriate wavelength in the millimetre range is present, the molecule makes transitions between the two states — the relative probabilities of the states being determined by the intensity of the radiation present. These probabilities can be estimated by measuring the relative strengths of the two lines. In this way the radiation intensity can be estimated in the neighbourhood of several stars in our galaxy. This intensity is not inconsistent with the blackbody curve and demonstrates that such radiation exists elsewhere also.

Could the radiation background arise from discrete sources distributed uniformly all over the sky? This, if it works, could be an alternative to its relic-interpretation. However, source distribution in the sky will tend to introduce patchiness in the radiation — unless the number of sources is very large. There is very little patchiness in the observed microwave background, indicating that the number of discrete sources must be at least as many as (perhaps even 10–100 times more) the number of ordinary galaxies, although individually the sources could be weaker. So far no such source population has been seen. Clearly future developments of infrared astronomy are important to settle this issue. In the absence of any such alternatives the only available interpretation is that of the 'relic of hot big bang'.

The remarkable isotropy of the background radiation poses problems even for the big bang models. We discussed one aspect of this, known as the *horizon problem* in Chapter 3. In the early stages of the big bang universe the regions of communication are terribly restricted by the so-called 'particle horizons'. Thus we expect patchiness and anisotropy to be introduced at this early stage. Why then has the background remained so uniform? How, in other words, did parts of the universe out of communication with one another manage to transmit information about the intensity of background in their own region so that a uniformity could somehow be achieved? Unless this could be shown explicitly we have to conclude that the universe and the background radiation were created isotropic in the first place. While this is a possible assumption to make, it adds one more important but 'unexplained' item to the astronomer's shelf. Inflationary cosmology helps in removing this difficulty as we saw in the last chapter.

Another kind of patchiness expected in the radiation background arises from the process of formation of galaxies, clusters and superclusters in the universe. In any structure formation theory in the big bang cosmology, the assumption is that initial small-scale inhomogeneities in the matter distribution grow to become the presently observed large scale structures. These inhomogeneities of matter are also linked with inhomogeneities of radiation since the two are intimately linked until the recombination epoch. Thereafter the radiation is decoupled from matter and tends to retain those imprints to this day.

Calculations in the 1960s and 1970s suggested that these imprints should be seen as fluctuations ΔT of temperature T over regions of size $\sim 1' - 1^\circ$, depending on whether we are talking of inhomogeneities of the galactic scale or supercluster scale.

Observations, however, failed to reveal any $\Delta T/T$ of the predicted order $\sim 10^{-3}$.

This is where dark matter of non-baryonic form became useful to structure formation scenarios. Non-baryonic matter does not interact with radiation and hence the latter need not carry strong imprints of inhomogeneities of the former. If the non-baryonic matter far exceeds the baryonic matter, one can arrive at $\Delta T/T$ as low as $\sim 10^{-5}$. Various theories using HDM or CDM or mixtures of both came up in the 1980s. Till 1992, on the angular scale mentioned above no positive measurement of $\Delta T/T$ was seen even down to $\sim 2 \times 10^{-5}$. However, in April 1992 for the first time, on the scale of 7°, the COBE satellite reported fluctuations of this order. With such data it has been possible to rule out a number of candidates for dark matter. The microwave background isotropy measurements thus provide useful constraints on structure formation theories and dark matter, and it is hoped that further results will be forthcoming in the next few years.

(iv) The origin of the elements We have already looked at the current ideas on the origin of the elements in Chapter 3. Here we will be concerned with how our knowledge of the distribution of the elements can be used to test cosmological models.

Present calculations suggest that the bulk of the elements found in the universe can be synthesized in the stars. The lighter nuclei like helium and deuterium can be made both in the stars and in the big bang. The question is — Is the big bang essential for these elements? If it should turn out that the stars alone cannot manufacture enough of these nuclei, then one is forced to take the big bang origin of the universe seriously.

There has been a search for deuterium (the heavy-hydrogen nucleus) in recent years in various parts of the galaxy. In the direction of the galactic centre, the ratio of D (deuterium) or H (hydrogen) is estimated in the range of 3–50 parts in 10,000. In the Orion Nebula the ratio is thought to lie between 1 and 100 parts in 10 million (this is deduced from the radio emission lines of DCN molecules). The present calculations, which depend on a number of theoretical assumptions and parameters, indicate that the hot big bang universe is able to manufacture enough deuterium to account for the observed ratios.

A somewhat less satisfactory situation exists with regard to the ^{3}He nucleus and other light nuclei like ^{7}Li, ^{9}Be, etc. To adjust the helium and deuterium as well as the ^{7}Li and ^{9}Be abundances into the primordial framework, the hot big bang picture needs additional parameter(s). One suggestion was that a loophole might be provided by the quark–gluon plasma phase (Section 3.3, page 38) in the form of an adjustible neutron-to-proton ratio which could be different from the canonical one of equation (3.47). It is too early yet to settle on a precise framework that fits all data on the abundances of light nuclei.

Calculations for the big bang origin can be related to the microwave background because this background, if of relic cosmological origin, provides information of radiation density in the very early stages of the universe when element synthesis took place. There is at present some discrepancy, which may be summarized as follows. The abundances of the light elements (for example, the deuterium/hydrogen ratio) depend critically on the matter density and the rate of expansion of the universe in its early stages. It turns out that the universe in those critical seconds, based on the inflation values of $\Omega_0 = 1$ and $q_0 = 1/2$, does not expand fast enough to account for the observed abundances. Big bang models with Ω_0 close to zero are better placed in this respect. However, large q_0 or Ω_0 can be allowed if there is non-baryonic dark matter since its presence does not affect the deuterium abundance.

Calculations like these indicate the role of nuclear and particle physics in cosmology. If the universe did originate in a big bang, violent activity in the form of nuclear reactions must have left its mark. The abundances of the various light nuclei provide a check on the possible conditions existing in such a big bang. Needless to say, any spectacular advances in our knowledge of nuclear and elementary-particle physics could drastically alter the present picture.

4.4 Exercises

4.1 Suppose a galaxy with redshift 0.5 has a spectrum $I(\lambda) \propto \lambda^2$ in the wavelength range $2500Å < \lambda < 5000Å$. Another galaxy with redshift 0.7 but the same spectrum is being observed with the first one at a wavelength band centred on 5000Å. Show that the corrections to the magnitudes of these galaxies will differ by $\sim 0.41^m$.

4.2 A radio source survey gives $N = 10$ at $S = 12.5$ Jy and $N = 93$ at $S = 5$ Jy. Show that in a Euclidean universe this would imply either a deficit of 13 sources at the high flux end or a deficit of 53 sources at the low flux end.

4.3 Calculate the minimum angular size of a spherical source of diameter 300 kpc in an Einstein–de Sitter universe with the present Hubble constant of 100 kms^{-1} Mpc^{-1} (see *Problem 4.3* in the text).

4.4 Given that globular clusters of age 12 Gyr exist, what would be the upper limit on the present value of the Hubble constant in an Einstein–de Sitter universe?

4.5 A galaxy has visible mass of $10^{11} M_{\odot}$. It has a flat rotation curve with rotational velocity of 250 kms^{-1} extending out to 50 kpc from the centre. Estimate the ratio of dark to luminous matter in the galaxy.

4.6 Given that the number density of photons at present is 400 cm^{-3}, estimate the number density of neutrinos.

5 *Present Challenges in Cosmology*

5.1 Structure formation

We begin this chapter with the current most popular problem in theoretical cosmology — the problem of how large scale structures (briefly described in Chapter 1) came about. Without going into quantitative details, the basic approach and its problems are outlined below.

The argument begins with the assumption that the inhomogeneity seen in the form of large scale structures had its seeds in the very early epochs — in the era of quantum gravity. The universe, as it went through the GUT phase transition, already had imprinted on it primordial fluctuations of matter density ($\delta\rho/\rho$). The inflationary phase was largely responsible for modifying these seeds to make them *scale independent*. That is, if we Fourier analyse the density contrast

$$\frac{\delta\rho}{\rho} \equiv \delta(\mathbf{r},t) = \int \delta_k(t) e^{ik.\mathbf{r}} \frac{d^3k}{(2\pi)^3}, \tag{5.1}$$

we would ultimately have

$$k^3 |\delta_k(t)|^2 = \text{constant}. \tag{5.2}$$

This ensures that the root mean square fluctuation of mass as a fraction of mass contained in a region of any size R should be independent of R. A spectrum of this kind was obtained empirically by E. Harrison and Ya.B. Zeldovich independently in 1970–72.

The scale invariant spectrum is apparently seen in structures on scales of galaxies to superclusters and so this initialization seems consistent with observations, provided theories of evolution of $\delta(\mathbf{r},t)$ can demonstrate that this feature is preserved from the early epochs to the present.

The enhancement of density in a given region works towards further growth because of the attractive nature of gravity. However, the internal thermal pressure gradients inhibit the condensation process. Therefore, by balancing the two opposing forces one arrives at a criterion that growth is possible only for masses exceeding a certain critical mass, called the *Jeans mass* (named after the astrophysicist James Jeans for his pioneering work in this field).

Jeans's work was in a static background. Clearly, in an expanding universe, any gravitational tendency for contraction has to counteract the expansion of the ambient medium as well as the thermal pressures. Thus the mass of a perturbation has to exceed the Jeans mass to ensure growth of the condensation.

It is clear that for realistic structures to emerge, we must have $|\delta| \gg 1$ at the end of the structure formation process. In the beginning of the process, however, $|\delta| \ll 1$ and one may use the perturbation theory. At some stage during the growth of the structures $|\delta|$ becomes comparable to 1 and the problem becomes nonlinear. This requires numerical techniques including N-body simulations on high speed computers. The recent rapid growth in computer technology has come at the right time for such studies to be undertaken.

There are several constraints on the models of structure formation. A major one is its impact on the microwave background. As we saw in Chapter 3, the imprints on the background survive undisturbed from the recombination epoch. So the fluctuations ($\Delta T/T$ on different angular scales) that are seen today must be accounted for in terms of matter–radiation interaction *prior* to the recombination epoch.

The second constraint is due to the *present* distribution of matter, in the form of filaments with voids in between. The filaments are made up of clusters of galaxies and extend to several tens of megaparsecs. The nonlinear regime must ultimately end with this kind of structure.

A related constraint is of the large scale motion of galaxies, groups and clusters relative to the expanding universe. How did so much 'random' or 'peculiar' motion survive after expansion? Moreover, relative to the rest frame in which the microwave background looks isotropic, there are examples of large scale streaming motion with speeds $\sim 10^3$ kms^{-1}. A structure formation theory should account for this feature.

Current attempts on structure formation very critically depend on dark matter. There are two reasons why the big bang cosmology requires non-baryonic dark matter. The first is the constraint on Ω from deuterium abundance. As we saw in the earlier chapters, the constraint is expressible as an upper limit on Ω (baryonic fraction of Ω). The specific inequality is

$$h_0^2\Omega_B \leq 0.0375. \quad (5.3)$$

For inflationary cosmologies $\Omega_0 = 1$, and hence $\Omega_0 - \Omega_B$ must be non-baryonic.

The second reason is that inhomogeneities of all-baryonic type would have too large an imprint on the microwave background. Thus dark matter has to be non-baryonic; but its gravitational impact on baryonic matter would depend on whether it is hot or cold. By and large, HDM would generate initial structures on the supercluster scale which would need to break up into smaller structures like clusters and galaxies. In the CDM models, smaller nuclei of galaxy-size structures would grow first and these have to evolve into larger units. These are respectively called 'top-down' and 'bottom-up' scenarios.

At present we do not have a satisfactory theory of structure formation despite claims to the contrary by various workers in the field. It would be fair to say that theoreticians are finding it hard to match the observational detail coming from extragalactic astronomy.

5.2 Observational Cosmology

Cosmological theories can be constrained, and will be constrained due to more and more precise observations. Specific areas in which future observations will prove decisive are:

(i) Measurement of H_0 and q_0 The estimate of H_0 based on the 1994 observations of the HST has underscored the advantages of a space telescope. Measurements of extragalactic distances will ultimately narrow down the permissible range of Hubble's constant. It will greatly affect cosmological models too. For, as we saw earlier, the popular inflationary model requires the universe to be only as old as $2/3H_0$. For $H_0 = 80$ kms^{-1} Mpc^{-1} (found by the HST), this value is only ~ 8.5 Gyr. The present age estimates for globular clusters are in the range 12–18 Gyr, and they are clearly in conflict with the cosmological age. Of course, future work will make the stellar age estimates also more precise, to focus this conflict more clearly.

Can we modify the inflationary prediction of age by introducing the cosmological constant? We can; but this recourse is open to two lines of criticism. First, the inflationary model will lose the thrust of its main criticism of non-inflationary models that they involve fine tuning of parameters. Why does λ happen to have precisely the value needed for the equality

$$\Omega_0 + \frac{\lambda}{3H_0^2} = 1, \tag{5.4}$$

when this λ is $\sim 10^{-108}$ of the λ-term causing inflation?

Secondly, a large λ-term ($\lambda > H_0^2$) will lead to $q_0 < 0$, i.e., to an accelerating universe. At present, the estimates of q_0 tend to show positive values. Obviously, with improved $m - z$ relations it may be possible to determine q_0 more precisely to check this possibility.

(ii) Measurement of abundances of light nuclei These measurements have been steadily improving and will clearly help in deciding the parameter space for the big bang models. It may happen that no adjustment of available parameters will reproduce all observed abundances. In which case the model would have to be seriously modified or replaced. On the other hand, better agreement with observed abundances will enhance its credibility.

(iii) Galaxy age spectrum The big bang scenario allows the galaxies to form in a relatively narrow time span so that most of them should be roughly the same age. The discovery of very old or very young galaxies would pose problems for these cosmologies. Thus stellar and galactic astrophysics will make valuable contributions to cosmology.

(iv) Discrete source populations Detailed studies of populations of galaxies, radio sources, quasi-stellar objects, active galactic nuclei, etc., will tell us about the structure and physical environment of the universe on scales of ≥ 150 Mpc. Basically, this would cover a redshift range $z \gtrsim 0.5$. At present, $z \lesssim 5$; but we still do not know what the redshift limit is upto which discrete sources may be found.

(v) Inter-galactic medium Multi-wavelength studies of the intergalactic medium (IGM) are extremely important for physical cosmology. Thus X-ray studies of hot gas in the clusters, ultraviolet spectroscopy for cosmic deuterium and helium, infrared searches for regions of new star formation, sources of gamma ray bursts, etc., are promising areas for extragalactic astronomy. The optical and radio studies of radio sources have, in the past, enriched our knowledge of the extragalactic universe. The same will happen through such studies at other wavelengths.

(vi) Microwave background Following COBE, other studies are beginning to report inhomogeneities in the distribution of the microwave background, with $\Delta T/T \sim 10^{-5}$. Studies at different wavelengths and at different angular scales will help put constraints on structure formation theories.

5.3 Alternative Cosmologies

Although the big bang cosmology has held centre stage since the 1970s, there are several chinks in its armour. These include the failure (despite many attempts) to arrive at a good theory of structure formation, the problem of the age of the universe, the lack of direct evidence for non-baryonic dark matter, the theoretical problem of a singular origin, the absence of a physical theory of the creation of matter, etc.

It is, therefore, prudent to keep an eye open for any alternative explanations which might compete with the big bang cosmology. The competition offered by the steady state cosmology in the 1950s and the 1960s had given a boost to extragalactic astronomy.

In this book, I have largely focused on the big bang cosmology but I should not end without keeping the door open for viable alternatives. The quasi-steady state cosmology described in Chapter 2 currently claims to offer an alternative scenario. Its authors have argued that for $P \cong 8 \times 10^{11}$ yrs and $Q \cong 4 \times 10^{10}$ yrs, it is possible to explain all observations of large scale structures, to offer an alternative explanation for the microwave background and for the origin of light nuclei, and to give an entirely baryonic interpretation to dark matter. This cosmology does not have a singular beginning; it has no age-problem and it gives a physical theory for matter creation. Rather than give details of these ideas here, I refer the reader to the bibliography at the end.

There are other possible alternatives even more radical than the quasi-steady state cosmology. For example, one may offer an alternative interpretation of Hubble's law that does not demand an expanding universe, one may explore theories that require some of the fundamental constants like $G, c, \hbar$, etc., to change with epoch, or one may look upon the universe as having a fractal structure.

In the final analysis, just as the proof of the pudding lies in the eating, the proof of a cosmological theory lies in its observational tests. It is these that will ultimately decide whether a theory stands or falls.

Appendix: Mathematical, Physical and Astronomical Constants *

Mathematical Constants

π	=	3.14519
e	=	2.71828
$\zeta(3)$	=	1.20206
ln 2	=	0.69315
ln	=	2.30259
log e	=	0.43429
1 arc second	=	4.8481×10^{-6} radians
1 steradian	=	3.2828×10^{3} square degrees
Square degrees on a sphere	=	41252.96124

Physical Constants

Speed of light	$c =$	$2.99792458(1.2) \times 10^{10}$ cm s^{-1}
Planck's constant	$\hbar =$	$1.0545887(57) \times 10^{-27}$ erg s
	$=$	$6.582173(17) \times 10^{-16}$ eV s
	$h \equiv$	$2\pi\hbar = 6.62620 \times 10^{-27}$ erg s
Electron volt	1 eV $=$	$1.6021892(46) \times 10^{-12}$ erg
Gravitational constant	$G =$	$6.6720(11) \times 10^{-8}$ dyn cm^2 g^{-2}
Charge of the electron	$e =$	$4.803242(14) \times 10^{-10}$ esu
Fine structure constant	$\alpha \equiv$	$(e^2/\hbar c) = [137.03604(11)]^{-1}$
Planck length	$\sqrt{(G\hbar/c^3)} =$	1.6×10^{-33} cm
Planck time	$\sqrt{(G\hbar/c^5)} =$	5.4×10^{-44} s
Planck mass	$\sqrt{(c\hbar/G)} =$	2.2×10^{-5} g
Electron mass	$m_e =$	$9.109534(47) \times 10^{-28}$ g

contd.

* Figures given in parentheses represent 1σ uncertainty in the last digits of the main numbers.

Physical Constants *contd.*

Electron mass energy	m_ec^2	$= 0.5110034(14)$ MeV
Proton mass energy	m_pc^2	$= 938.2796(27)$ MeV
Neutron mass energy	m_nc^2	$= 939.5731(27)$ MeV
Planck energy	$\sqrt{(c^5\hbar/G)}$	$= 1.2 \times 10^{19}$ GeV
Thomson cross section	$(8\pi e^4/3m_e^2c^4)$	$= 0.06652448(33) \times 10^{-24}$ cm^2
Boltzmann constant	k	$= 1.380662(44) \times 10^{-16}$ erg K^{-1}
	k^{-1}	$= 11604.50(36)$ K(eV)$^{-1}$
Radiation constant	a	$= (8\pi^5k^4/15c^3h^3)$
		$= 7.5641 \times 10^{-15}$ erg cm^{-3} K^{-4}
Number density of photons in a blackbody radiation of temperature T	$\frac{2\zeta(3)}{\pi^2}\left(\frac{kT}{c\hbar}\right)^3$	$\cong 20.3$ T^3 cm^{-3}
Weak interaction constant	$\mathcal{G}$	$= 1.02 \times 10^{-5}(\hbar^3/m_p^2c)$
Binding energy of deuterium		$= 2.22464(4)$ MeV
Binding energy of helium		$= 28.2969(4)$ MeV

Astronomical Constants

Light year	1 ly	$= 9.4605 \times 10^{17}$ cm
Parsec	1 pc	$= 3.0856(1) \times 10^{18}$cm
		$\cong 3.26$ ly
Radius of the Sun	$R_\odot$	$= 6.959 \times 10^{10}$ cm
Mass of the Sun	$M_\odot$	$= 1.989(1) \times 10^{33}$ g
Luminosity of the Sun	$L_\odot$	$= 3.826(8) \times 10^{33}$ erg s^{-1}
Mass/light ratio for the Sun	$(M_\odot/L_\odot)$	$\cong 0.51$ g erg^{-1} s
Luminosity of a star of zero absolute magnitude ($M_{\rm bol} = 0$)	L_0	$= 2.97 \times 10^{35}$ erg s^{-1}
Flux from a star of zero apparent magnitude	l_0	$= 2.48 \times 10^{-5}$ erg cm^{-2} s^{-1}
Radio flux density (Jansky)	1 Jy	$= 10^{-26}$ Wm^{-2} Hz^{-1}
Hubble constant	H_0	$= 100h_0$km s^{-1} Mpc^{-1}, $0.5 \le h_0 \le 1$
Hubble age	T_0	$= H_0^{-1} \cong 9.8h_0^{-1} \times 10^9$ yr

Sources

1. I.M. Ryshik and I.S. Gradstein, *Tables of Series, Products and Integrals*, Veb Deutscher Verlag der Wissenschafte, Berlin (1957).
2. Particle Data Group, 1980, 'Review of particle properties', *Rev. Mod. Phys.*, **2**, 52 (1980).
3. A.H. Wapstra and N.B. Gove, 'The 1971 Atomic Mass Evaluation', *Nuclear Data Tables*, **9A**, 265 (1971).
4. C.W. Allen, *Astrophysical Quantities*, Athlone Press, London (1973).

Suggested further reading

Some introductory textbooks on cosmology and related subjects

1. H. Bondi, *Cosmology* (Cambridge University Press, Cambridge, 1961)
2. F. Hoyle and J.V. Narlikar, *The Physics–Astronomy Frontier* (W.H. Freeman, New York, 1980)
3. D. Layzer, *Constructing the Universe* (Scientific American Library, New York, 1984)
4. J.V. Narlikar, *The Primeval Universe* (Oxford University Press, Oxford, 1988)
5. J.V. Narlikar, *Introduction to Cosmology* (Cambridge University Press, Cambridge, 1993)
6. P.J.E. Peebles, *Physical Cosmology* (Princeton University Press, Princeton, 1971)
7. D.W. Sciama, *Modern Cosmology* (Cambridge University Press, Cambridge, 1971)
8. F. Shu, *Physical Universe : An Introduction to Astronomy* (University Science Books, Mill Valley, 1992)
9. S. Weinberg, *The First Three Minutes* (Basic Books Inc., New York, 1977)

Some pioneering papers in cosmology

1. A. Einstein, *Sitzungberichte der K. Preussischen Akademie der Wissenschaften zu Berlin* (for the Einstein universe), 142 (1917)
2. W. de Sitter, *Proceedings of the Royal Academy of Sciences, Amsterdam* (for the de Sitter universe), **19**, 1217 (1917)
3. A. Friedmann, *Zeitshrift für Physik* (for Friedmann models), **10**, 377 (1922)
4. A. Friedmann, *Zeitshrift für Physik* (for Friedmann models), **21**, 326 (1924)
5. E. Hubble, *Proceedings of the National Academy of Sciences* (for Hubble's law), **15**, 168 (1929)
6. F. Zwicky, *Helvetica Physica Acta* (for dark matter in clusters), **6**, 10 (1933)
7. E.A. Milne and W.H. McCrea, *Quarterly Journal of Mathematics* (for Newtonian cosmology), **5**, 73 (1934)
8. R.A. Alpher, H.A. Bethe and G. Gamow, *Physical Review* (for nucleosynthesis in the early universe), **73**, 80 (1948)
9. H. Bondi and T. Gold, *Monthly Notices of the Royal Astronomical Society* (for the steady-state theory), **108**, 252 (1948)

10. F. Hoyle, *Monthly Notices of the Royal Astronomical Society, London* (for the steady state theory with creation field), **108**, 372 (1948)
11. F. Hoyle, *Paris Symposium on Radio Astronomy* (for angular size–redshift relation), Ed. R.N. Bracewell (Stanford University Press), 529 (1959)
12. A. Sandage, *The Astrophysical Journal* (for optical tests of cosmological models), **133**, 355 (1961)
13. P.F. Scott and M. Ryle, *Monthly Notices of the Royal Astronomical Society* (for radio source counts), **122**, 389 (1961)
14. A.A. Penzias and R.W. Wilson, *The Astrophysical Journal* (for the discovery of the microwave background), **142**, 419 (1965)
15. F. Hoyle, *Proceedings of the Royal Society* (for radio source counts), **A308**, 1 (1968)
16. E.R. Harrison, *Physical Review* (for scale invariant spectrum of large scale structure), **D1**, 2726 (1970)
17. Ya.B. Zeldovich, *Monthly Notices of the Royal Astronomical Society* (for scale invariant spectrum of large scale structure), **160**, 1P (1972)
18. A.H. Guth, *Physical Review* (for inflationary cosmology), **D23**, 347 (1981)
19. R. Cowsik and J. McClelland, *The Astrophysical Journal* (for massive neutrinos), **180**, 7 (1972)
20. J.C. Mather et al, *The Astrophysical Journal Letters* (for the COBE spectrum of the microwave background), **354**, 37 (1990)
21. G.F. Smoot et al, *The Astrophysical Journal Letters* (for the COBE detection of inhomogeneities in the microwave background), **396**, 1 (1992)
22. F. Hoyle, G. Burbidge and J.V. Narlikar, *The Astrophysical Journal* (for the quasi-steady state cosmology), **410**, 437 (1993)

Index

The book

This book is based on lectures given by the author at university campuses with the aim of introducing the subject of cosmology to students and teachers at the graduate level. The subject is discussed within the framework of Newtonian gravity and mechanics because unfamiliarity with the general theory of relativity hampers the understanding of basic cosmological issues. The description is up-to-date and includes cosmological models, their physical properties and observational tests.

The author

Dr Jayant V Narlikar is the Director of the Inter-University Centre for Astronomy and Astrophysics, Pune. He is also Honorary Professor at the JNCASR, Bangalore. In a long and outstanding teaching and research career, Dr Narlikar was associated with a number of theories and studies in modern cosmology and has worked with Fred Hoyle on the Hoyle-Narlikar *theory of conformal gravity.*

Dr Narlikar has received several awards including the S S Bhatnagar Award for Physical Sciences and the M P Birla Award for Astronomy. He is a Fellow of the Third World Academy of Sciences and the President of the Cosmology Commission of the International Astronomical Union for the period 1994-97.

He is also an author of international repute and many of his books have been translated and published all over the world.

New and forthcoming books

N Kumar : *Deterministic Chaos: Complex Chance Out of Simple Necessity*

K S Valdiya : *Dynamic Himalaya*

R Gadagkar : *Cooperation and Conflict in Animal Societies*

S R Gadre : *Electrostatics of Atoms and Molecules*

Cover: The Galaxy 1346–252 in a cluster.
(Imaged by Ashish Mahabal at the Swope Telescope, Chile.)

The Educational Monographs published by Universities Press in collaboration with JNCASR address the needs of students and the teaching and research community.

Cover design by OSDATA

Universities Press

Jayant V Narlikar : *Elements of Cosmology*

ISBN 81 7371 043 0

Rs 65.00