PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Introduction to Big Bang Nucleosynthesis: Open and Closed Models, Anisotropies [and Discussion]

R. J. Tayler and R. Fabbri

Phil. Trans. R. Soc. Lond. A 1982 **307**, 3-17
doi: 10.1098/rsta.1982.0096

Phil. Trans. R. Soc. Lond. A 307, 3–17 (1982) [3]
Printed in Great Britain

Introduction to Big Bang nucleosynthesis: open and closed models, anisotropies

By R. J. Tayler
Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, U.K.

A variety of observations suggest that the Universe had a hot dense origin and that the pregalactic composition of the Universe was determined by nuclear reactions that occurred in the first few minutes. There is no unique hot Big Bang theory, but the simplest version produces a primeval chemical composition that is in good qualitative agreement with the abundances deduced from observation. Whether or not any Big Bang theory will provide quantitative agreement with observations depends on a variety of factors in elementary particle physics (number and masses of stable or long-lived particles, half-life of neutron, structure of grand unified theories) and from observational astronomy (present mean baryon density of the Universe, the Hubble constant and deceleration parameter). The influence of these factors on the abundances is discussed, as is the effect of departures from homogeneity and isotropy in the early Universe.

1. Introduction

The purpose of this paper is to introduce the more detailed discussions that follow. Our topic is the Big Bang and element creation, which implies some degree of belief that there was a Big Bang and that it was responsible for some of the observed element abundances. I shall therefore start with a summary of the reasons for the belief and of some possible problems that are raised.

I start by listing the main observations of cosmological significance and then I comment on all of them in turn. The relevant observations are:

(i) red shifts in the spectra of distant galaxies;
(ii) the isotropic distribution of galaxies and radio sources;
(iii) the homogeneous distribution of nearby galaxies;
(iv) evidence for evolution in the population of extragalactic radio sources;
(v) the local density of matter;
(vi) the ages of galactic objects;
(vii) the cosmic microwave background;
(viii) the chemical composition of stars and gas clouds;
(ix) charge symmetry;
(x) matter–antimatter asymmetry.

In interpreting the above observations we make a basic assumption that the laws of physics are not explicitly dependent on time and that they are the laws of physics, which have been established in our own environment, with appropriate theoretical extrapolation to conditions too extreme to be studied directly on Earth. Later authors in this symposium discuss uncertainties in some of the laws in the very early stages of the expansion of the Universe. These uncertainties will not explicitly affect my discussion but their implicit effect will subsequently be pointed out. In particular, in our interpretation of the observations, we shall assume that the large-scale structure of the Universe is governed by the General Theory of Relativity and for most of our paper we shall assume that the cosmological constant is zero.

The observation of the red shifts in spectra of distant galaxies is most readily interpreted as due to the expansion of the Universe and it is well known that General Relativity permits expanding models of the Universe. The observations can be converted into a velocity–distance relation by use of a suitable definition of distance such as luminosity distance or distance by apparent size, provided that we possess what are called *standard candles* or *standard metre rules*. These are objects whose luminosity or size is believed to be well known and unchanging with time. Hubble originally used such standard candles to obtain his velocity–distance relation

$$v = H_0 r, \tag{1.1}$$

where H_0 is called the Hubble constant and H_0^{-1} is clearly related to the age of the Universe. I shall return to standard candles and the Hubble constant after I have discussed some of the other observations.

The second and third observations have led to the view that the Universe is on the large scale homogeneous and isotropic. This means that we can discuss its evolution in terms of a cosmic time upon which all observers locally at rest agree and that, at a given value of the cosmic time, the Universe has the same properties at all points and in all directions at any point. The Universe is, of course, far from being strictly homogeneous and isotropic at present. However, although the distribution of galaxies and radio sources is clustered, there is no significant statistical deviation from isotropy in their large-scale distributions on the sky. The observation that nearby galaxies are in the large distributed uniformly in space is less clear than the observation of isotropy. It is based on counts of galaxies to different limiting magnitudes and it is discussed, for example, by Peebles (1980). The general conclusion is that large-scale homogeneity and isotropy, as seen from the Earth, is a good approximation and this is then supplemented by the assumption that the Earth is not in a privileged position in the Universe.

The fourth observation is an important indication that we live in an evolving Universe rather than in a steady-state Universe. The counts of extragalactic radio sources at different flux densities are incompatible with the assumption that the population of radio sources is unchanging with cosmic epoch, which was a vital property of the Steady State theory. In fact it now seems clear that both the space density and average luminosity of radio sources were greater in the past.

The observation that the Universe is at present homogeneous and isotropic on the large scale has led to the conjecture that it was completely homogeneous and isotropic in its earliest stages and that all of the observed structure has subsequently arisen from gravitational effects. For a strictly homogeneous and isotropic Universe the metric has the Robertson–Walker (R–W) form

$$ds^2 = c^2 dt^2 - R^2(t)\,\{dr^2/(1-kr^2) + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\}, \tag{1.2}$$

where r, θ, ϕ are dimensionless co-moving coordinates attached to fundamental observers and where $R(t)$ describes the time behaviour of the Universe, k can take the values $+1$, 0, -1; the Universe is spatially closed if $k = +1$ and open otherwise. As we shall see later, it does not in fact appear possible that the present structure can have arisen from a strictly homogeneous and isotropic initial state, although departures from the R–W metric may always have been small.

The General Theory of Relativity relates $R(t)$ to the matter in the Universe through two equations which can be written

$$8\pi G\rho R^2 = 3kc^2 + 3\dot{R}^2 - \Lambda R^2, \tag{1.3}$$

$$3\ddot{R} = [\Lambda - 4\pi G(\rho + 3P/c^2)]\,R, \tag{1.4}$$

where P, ρ are pressure and density, Λ is the cosmological constant and dot denotes differentiation with respect to time. If $\Lambda = 0$, which we shall mainly assume, the equations become

$$8\pi G\rho R^2 = 3kc^2 + 3\dot{R}^2, \tag{1.5}$$

$$3\ddot{R} = -4\pi G(\rho + 3P/c^2)\,R. \tag{1.6}$$

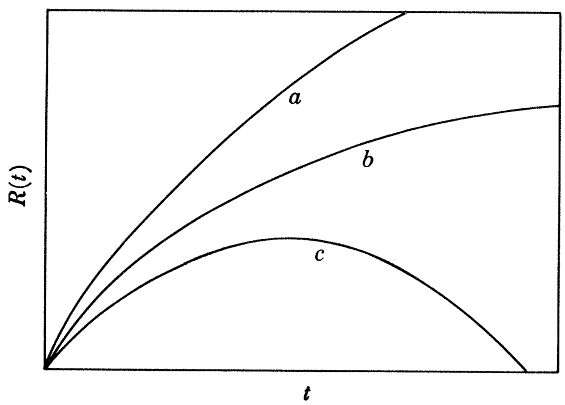

FIGURE 1. The behaviour of $R(t)$ in a $\Lambda = 0$ Universe. (*a*) Open model; (*b*) model with critical density; (*c*) closed model.

In terms of $R(t)$, the Hubble constant can be written

$$H_0 = [\dot{R}(t)/R(t)]_0, \tag{1.7}$$

where 0 denotes the present epoch. It is also useful to introduce the deceleration parameter q_0, defined by

$$q_0 = [-R\ddot{R}/\dot{R}^2]_0. \tag{1.8}$$

It can be seen from (1.6) that, if $\Lambda = 0$, q_0 is necessarily not less than 0. The possible behaviour of $R(t)$ is shown in figure 1; R increases without limit if $k \leqslant 0$, but the Universe recollapses if $k = +1$. At the present epoch we believe that $P \ll \rho c^2$. From (1.6) it then follows that the present mean density of the Universe, ρ_0, satisfies

$$q_0 = \rho_0/2\rho_c, \tag{1.9}$$

where

$$\rho_c = 3H_0^2/8\pi G. \tag{1.10}$$

It is also clear from (1.5) that, if $k = 0$, $\rho_0 = \rho_c$. Thus the values of k and q_0 are not independent; the Universe is closed if $\rho_0 > \rho_c$, $q_0 > \frac{1}{2}$ and open if $\rho_0 \leqslant \rho_c$, $q_0 \leqslant \frac{1}{2}$.

If we really possessed standard candles or metre rules, it would be possible to use the red shift – apparent luminosity relation or the red shift – apparent size relation to obtain values of both H_0 and q_0. There are, however, problems of different types affecting the determination of each of H_0 and q_0. For H_0 the problem is calibration of the local distance scale, which leads

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

to the distance to the nearest standard candle, which we wish to use at very large distances. At present different authors give

$$50 \lesssim H_0/\mathrm{km\ s^{-1}\ Mpc^{-1}} \lesssim 100. \tag{1.11}$$

It is clear from figure 1 that the present age of the Universe, t_0, must be not more than H_0^{-1}, if $\Lambda = 0$, with equality only in a completely open Universe free of matter. By using (1.11) this implies that

$$10 \lesssim t_0^{\max}/\mathrm{Ga} \lesssim 20. \tag{1.12}$$

The problem with direct determination of q_0 is that it is far from clear that we do have standard candles. It is possible to determine H_0 by using observations of galaxies that are not significantly younger than the nearby calibrating galaxies so that effects of galactic evolution are unimportant. This is not true in the determination of q_0, when we are looking so far into the past that galactic evolution must surely be important. In addition, central galaxies in clusters, which are used as standard candles, may swallow smaller galaxies. As a result of uncertainties in these effects, it is difficult to place reliability on the direct determination of q_0.

In principle, observation (v) of the local density of matter can be used to determine q_0 through (1.9); clearly that leaves a possible uncertainty of a factor of 4 in the value of q_0 because of the lack of a precise value for H_0. More significantly, observations only give a value for the density of matter in our neighbourhood whose presence is known either directly or through its gravitational influence, and the true value of ρ_0 could be significantly higher than its observed lower limit. As a result, even if ρ_c is assumed to be known, we only have a lower limit to q_0. ρ_0 is believed to have a lower limit of order a small percentage of ρ_c; Davis (this symposium) discusses problems in determining ρ_0. It is usually assumed that the known matter is baryonic matter so that we also have a lower limit to the baryonic mass density, ρ_B. This need not be so, as will be explained later.

The age of the Universe (for $\Lambda = 0$) must be less than $1/H_0$, how much less being determined by the value of q_0. The ages of galactic objects, observation (vi), should be less than the age of the Universe. The greatest ages are those of globular star clusters obtained from studies of stellar evolution and of the heavy radioactive elements obtained from a discussion of the history of nucleosynthesis in the Galaxy. These are generally believed to indicate a galactic age of about 11–12 Ga. Although there is an uncertainty in this value and indeed some authors have derived very much larger ages for globular clusters, it is difficult to see how the galactic age could be substantially less than the smaller value of 10 Ga given in inequalities (1.12). This implies that the simplest version of the Big Bang theory may be incompatible with observations if $H_0 \approx 100\ \mathrm{km\ s^{-1}\ Mpc^{-1}}$, unless the Universe is very open, which would mean that the observed lower limit to ρ_0 is very near to its true value.

I have not discussed four of the observations at present. These are the existence and properties of the cosmic microwave radiation; the observation that the chemical composition of stars and gas clouds is approximately two and a half or three parts hydrogen to one part helium by mass, with only a very small admixture of heavier elements; the fact that the net electrical charge of the Universe is zero or essentially zero; and that the Universe, as far as our observations extend, appears to be made of matter rather than a mixture of matter and antimatter.

I shall not comment further on the electric charge except to point out that, because the electrostatic force between particles is very much greater than the gravitational force, the large-scale structure of the Universe would be governed by electromagnetism rather than by gravitation if the charge were not essentially zero. The microwave radiation and the element abundances will be discussed in the next section. With regard to the matter–antimatter problem, it is generally but not universally agreed (Steigman 1976) that the evidence from γ-ray observations excludes regions where matter and antimatter meet, so that the Universe is made of matter. However, in a sense the matter–antimatter excess is small because the photon is its own antiparticle. It appears that the present ratio of baryon density, n_B, to photon density, n_γ, satisfies

$$n_B/n_\gamma \approx 10^{-9}. \tag{1.13}$$

The precise value of this ratio will prove to be an important point in the later discussion.

2. The standard hot Big Bang theory

The observations that I have described in some detail have given a picture of the Universe expanding from a very dense state at $t = 0$ but we have said nothing about the physics of the early Universe. It was Gamow (Alpher, Bethe & Gamow 1948) who suggested that the Universe started very hot as well as very dense so that

$$\rho = T = \infty \quad \text{at} \quad t = 0. \tag{2.1}$$

Clearly the conditions (2.1) are a mathematical idealization, but Ellis and Guth (this symposium) discuss what may have been the true conditions in the extremely early Universe. Gamow pointed out that, if the Universe were initially very hot, nuclear reactions would occur. Initially he hoped that all of the chemical elements would be synthesized by a chain of neutron capture reactions, but his detailed calculations showed that after a few minutes, when the nuclear reactions ceased, the composition was a mixture of ^1H and ^4He and only a trace of other light isotopes. Gamow also pointed out that many of the photons present in the early stages of the Universe would not subsequently have been absorbed and that they would remain today highly red-shifted, so that the Universe would be filled with microwave radiation.

The discovery of the microwave background by Penzias & Wilson (1965), which followed shortly after the realization that the hydrogen and helium contents of stars and gas clouds were probably consistent with the predictions of the hot Big Bang theory (Hoyle & Tayler 1964), led to a greatly increased interest in the subject, which has continued until now. It is the purpose of the present meeting to discuss whether quantitative agreement between theory and observation appears possible. A full account of the physical processes occurring in the early Universe can be found, for example, in Weinberg (1972). Pagel, Kinman and Sargent (this symposium) discuss the values of primeval light element abundances deduced from observations, and Schramm gives a critical discussion of the comparison between theory and observation. My purpose is to outline the key processes and the important uncertainties that remain in the theory.

In his original theory Gamow assumed that initially the Universe was composed of only neutrons and photons, but it was subsequently made clear that this was not a possible starting condition because of pair creation from photons. Instead, more recent calculations have assumed that all particles and antiparticles are in equilibrium with radiation when the mass of the particle and the temperature satisfy

$$kT \gtrsim mc^2. \tag{2.2}$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

This means that the precise composition of the Universe cannot be specified. The actual composition is then determined by ρ, T, the values of certain conserved numbers in a comoving volume and the principle of thermodynamic equilibrium.

The numbers are:

$$\left.\begin{array}{ll} C & \text{charge number,} \\ B & \text{baryon number,} \\ L_e, L_\mu, (L_\tau) & \text{electron, muon, (tauon) lepton numbers.} \end{array}\right\} \qquad (2.3)$$

It now seems probable that only C is strictly conserved, although the values of the other numbers should not change very much except in the extremely early stages of the Universe, discussed by Ellis and Guth (this symposium), or in the remote future if the Universe is open. For the moment I treat them as conserved and shall return to their possible lack of conservation later. At present, direct knowledge about the tauon and its possible neutrino is not sufficient to prove that L_τ exists, although that seems very likely. If we accept that the above numbers are independent conserved quantities, there is no single hot Big Bang theory but a fourfold (fivefold) infinity of such theories. Most attention has concentrated on the standard model with

$$C = L_e = L_\mu = (L_\tau) = 0, \quad B \neq 0, \qquad (2.4)$$

which is an updated version of Gamow's theory. There is then the single free parameter B, which measures the relative importance of matter and radiation in the Universe; a knowledge of the value of B is equivalent to a knowledge of the present value of ρ_B/ρ_γ. The results of the theory are not substantially different if any $L \neq 0$ unless $|L| \gg B$.

In the early stages of the Universe all particles and antiparticles are in equilibrium with radiation, with there being a very small matter excess because B is non-zero and positive. Each particle–antiparticle pair disappears by decay or annihilation when its mass ceases to satisfy (2.2). For a discussion of the element production the details of the Universe before $T = 10^{11}$ K (say) are unimportant except in so far as they might affect the homogeneity and isotropy of the Universe and the values of B and L. We shall assume that the Universe is strictly homogeneous and isotropic and that B and L are arbitrary. When $T \approx 10^{11}$ K the only particles that are still in equilibrium with radiation are e^+, e^-, ν_e, $\bar{\nu}_e$, ν_μ, $\bar{\nu}_\mu$, ν_τ, $\bar{\nu}_\tau$ (if ν_τ exists). In addition there are a number of neutrons and protons, which supply the net baryon number.

In order that the present number of baryons in relation to microwave photons shall be correct, it is necessary that when T is high the rest mass energy of the baryons is infinitesimal compared with the energy of the leptons and photons and that the effective equation of state of the matter is

$$P = \tfrac{1}{3}\rho c^2. \qquad (2.5)$$

At this stage we also have

$$\rho c^2 = \tfrac{43}{8} a T^4, \qquad (2.6)$$

where aT^4 is contributed by photons, $\frac{7}{8}aT^4$ by each of e^+, e^- and $\frac{7}{16}aT^4$ by each species of neutrino and antineutrino. The solution of equations (1.3), (1.4) using the equation of state (2.5), (2.6) gives

$$\tfrac{32}{3}\pi G \rho t^2 = 1 \qquad (2.7)$$

and

$$T = 0.96 \times 10^{10} t^{-\frac{1}{2}}. \qquad (2.8)$$

The temperature of the Universe is thus about 10^{10} K when $t \sim 1$ s; values of k and Λ only affect the solution of the equations at a very much later time.

The key processes that follow are as follows.

(i) At $T \approx 10^{10}$ K the neutrinos have a mean free path that then and subsequently exceeds the radius of the observable Universe. As a result the neutrinos decouple from matter but they remain in the Universe today as 'microwave' neutrinos.

(ii) Because of the decoupling of the neutrinos, the relative number of neutrons and protons cannot be kept at the equilibrium value

$$n_n/n_p = \exp\{-(m_n - m_p)\,c^2/kT\} = \exp(-2.5 m_e c^2/kT) \tag{2.9}$$

by reactions of the type

$$\left.\begin{aligned} n + e^+ &\rightleftarrows p + \bar{\nu}_e, \\ p + e^- &\rightleftarrows n + \nu_e, \\ n &\rightleftarrows p + e^- + \bar{\nu}_e. \end{aligned}\right\} \tag{2.10}$$

As a result the ratio n_n/n_p is effectively fixed, apart from free neutron decay, which is still a slow process.

(iii) At $T \approx 10^9$ K the reverse reaction to

$$n + p \rightarrow d + \gamma \tag{2.11}$$

ceases to be effective and there is a rapid build-up of deuterium and then helium. Essentially all of the neutrons and an equal number of protons are converted to ^{4}He, with there also being a small amount of ^{2}H, ^{3}He and ^{7}Li.

(iv) Also at $T \approx 10^9$ K the electron–positron pairs annihilate through

$$e^+ + e^- \rightarrow \gamma + \gamma. \tag{2.12}$$

The energy released 'reheats' the photons relative to the neutrinos so that subsequently

$$T_\gamma = (\tfrac{11}{4})^{\frac{1}{3}}\, T_\nu = 1.4 T_\nu. \tag{2.13}$$

This enables us to calculate the present temperature of the background neutrinos from the microwave photon temperature. From this time onwards

$$n_B/T_\gamma^3 = \text{constant}. \tag{2.14}$$

This means that if we know n_{B0} and $T_{\gamma 0}$ we also in effect know B.

(v) At a temperature of a few kilokelvins the nuclei and electrons combine to form neutral matter and the photons then have a mean free path greater than the radius of the observable Universe. The photons decouple and this is the origin of the microwave radiation which we observe today and whose properties are discussed by Wilkinson, Fabbri and Richards (this symposium). With the range of values of B which are appropriate, at about the same time the rest mass energy of matter exceeds the energy of photons and neutrinos and for the first time the Universe becomes matter-dominated. Subsequently an appropriate approximation to the equation of state is

$$P = 0, \tag{2.15}$$

which has already been used in the discussion of the relation between q_0, H_0 and ρ_0.

What we are interested in are the precise abundances produced in (iii) above and how they compare with the observations. Within the standard model, which I am at present describing, there are three main sources of uncertainty:

(*a*) the value of B or, equivalently, the present value of n_B/n_γ;

(*b*) the rate of the weak interactions (2.10);

(*c*) the number of leptons with associated neutrinos, which affects the numerical constants in (2.6) and (2.8) and hence the rate at which the Universe cools.

I discuss each of these points in turn.

Provided that B is in the range indicated by observations (however uncertain) of ρ_0, H_0 and q_0, the final value of n_n/n_p and the resulting helium production are not greatly dependent on B. Thus a change of two orders of magnitude in B only leads to a change of *ca.* ± 0.02 around 0.25 in Y, the helium mass fraction. Although this change is small, it is of the order of the uncertainty in the primeval helium abundance claimed from the observations by at least some investigators, so that the precise value of B is important. The value of B determines when reaction (2.11) becomes effective. The effect on the abundance of ^{2}H and the other light isotopes is much more dramatic. How much deuterium survives depends critically on B, with its abundance being decreased by a factor of 10^3 if B increases by a factor of 10^2 in the relevant range. The observed deuterium cannot be primeval if the value of B is too high. At the same time it is not clear that there is a galactic process that might have synthesized all of it.

Points (*b*) and (*c*) both relate to uncertainties in elementary particle physics that affect the time at which the neutrinos decouple and as a result the final value of n_n/n_p and the element abundances. At energies of order 1 MeV the weak interaction cross section is proportional to (energy)^2 or in a thermal assembly to T^2. The number density of neutrinos and electrons, which are in equilibrium with radiation, is proportional to T^3. Thus the weak interaction timescale can be written

$$t_{wk} = 1/n\sigma c = AT^{-5}, \tag{2.16}$$

where the constant A is inversely proportional to the rate of the weak interactions. If we suppose that the number of types of neutrino and antineutrino is unknown, (2.6) can be replaced by

$$\rho c^2 = \alpha a T^4, \tag{2.17}$$

where

$$\alpha = \tfrac{11}{4} + \tfrac{7}{8}n \tag{2.18}$$

and n is the number of types of lepton with associated neutrinos. If we combine (2.17) with (2.7) we see that the expansion time for the Universe is

$$t_{ex} = (\tfrac{3}{32}c^2/\pi G\alpha a)^{\frac{1}{2}}\, T^{-2}. \tag{2.19}$$

The neutrinos decouple when the time given by (2.16) equals that given by (2.19) and it can be seen that this time depends on each of A and α.

The rates of all the weak interaction processes (2.10) depend on that for free neutron decay or equivalently on the half-life of the neutron. The experimental value of the half-life has declined substantially from *ca.* 14 min when Gamow introduced the theory. The three latest published values are 10.6 min (Christensen *et al.* 1972), 10.1 min (Bondarenko *et al.* 1978) and 10.8 min (Byrne *et al.* 1980). A reduction in the half-life of the neutron corresponds to a later time of neutrino decoupling, a lower value of n_n/n_p and a smaller production of helium. If the uncertainty in the half-life is no greater than what is implied by these latest values, Y is uncertain by *ca.* 0.01.

A key uncertainty in the theory is the number of leptons with associated neutrinos. An increase in the value of α in (2.17) and (2.19) causes the neutrinos to decouple at a higher temperature so that the value of n_n/n_p is higher and the helium production is increased. When Gamow introduced the theory, only ν_e was known. By the time that the microwave radiation was discovered, ν_μ had also been detected. We have assumed that ν_τ also exists. If there is a small number of additional neutrinos, the helium production in the standard model may be incompatible with observations. If the number of types of neutrino and hence α becomes sufficiently large, the Universe expands and cools too rapidly for the nuclear reactions to be fully effective and the helium production is decreased again. Both the maximum value of Y and the corresponding value of α depend on B. For smaller B, the nuclear reactions become ineffective for a smaller value of Y and for small enough B even this value of Y may not be incompatible with observations. The constraints of number of neutrino types have recently been discussed by Olive *et al.* (1981) and are discussed by Schramm (this symposium).

A further uncertainty in the standard model, which is of a different character, is related to the possibility of finite neutrino mass. Provided $m_\nu \ll 1\ \mathrm{MeV}/c^2$, the discussion given above is not changed at all. The neutrinos are still relativistic when they decouple at $T \approx 10^{10}$ K and they survive in the same number to the present day. The element production for a given value of B is also unaltered. What the neutrino mass does is twofold. It affects the value of q_0 and thus places constraints on the values of H_0 that are compatible with the ages of galactic objects and it allows the possibility that the value of ρ_B and hence B is less than is generally believed. I discuss each of these points in turn.

Suppose that there are n_ν types of neutrino (neutrinos and antineutrinos counted separately) and that the average neutrino mass is m_ν. Suppose that the microwave temperature today is $T_{\gamma 0}$ so that the number density of neutrinos is the same as that of massless neutrinos with temperature $(\frac{4}{11})^{\frac{1}{3}}\,T_{\gamma 0}$. It is then easy to write down an expression for the rest mass density of neutrinos, ρ_ν, in terms of ρ_c in the form

$$\frac{\rho_\nu}{\rho_c} = 0.4\left(\frac{n_\nu}{6}\right)\left(\frac{m_\nu c^2}{kT}\right)\left(\frac{T_{\gamma 0}}{2.7}\right)^3\left(\frac{100\ \mathrm{km\ s^{-1}\ Mpc^{-1}}}{H_0}\right)^2. \tag{2.20}$$

It is clear from this formula that, with the currently accepted number of neutrino types, 6, and the estimated range of values of H_0, an average neutrino mass of a few times $10\ \mathrm{eV}/c^2$ would lead to a closure density or greater in the form of neutrinos. This density would then determine the deceleration parameter through (1.9) but, as already stated, there is not at present a good enough value of q_0 to check this point. What can, however, be said is that a high value of ρ_ν would lead to a low present age, t_0, of the Universe. Thus:

$$\left.\begin{aligned} \rho_\nu &= \rho_c, & q_0 &= \tfrac{1}{2}, & t_0 &= 0.66/H_0;\\ \rho_\nu &= 2\rho_c, & q_0 &= 1, & t_0 &= 0.57/H_0. \end{aligned}\right\} \tag{2.21}$$

Such values of t_0 would look uncomfortably small in relation to the believed ages of the heavy radioactive elements and of the globular star clusters if $H_0 \approx 100\ \mathrm{km\ s^{-1}\ Mpc^{-1}}$. The interest in finite neutrino masses was stimulated in 1980 by the suggestion of experimental evidence in favour, including a measured value (Lyubimov *et al.* 1980) of *ca.* $30\ \mathrm{eV}/c^2$ for the mass of $\bar{\nu}_e$. At present the support for the measured results is not so great but the interest in cosmological and astrophysical implications remains.

If the neutrino mass is not high enough to rule out the standard model on the age problem, it might help to reconcile observed element abundances with theory. If neutrinos are massive they may have clustered and played a role in the formation of galaxies and clusters of galaxies.

In particular some of the mass, for which we have evidence from rotation curves of single galaxies and from the motions of galaxies in binary systems and small groups but which is not directly observed, could be in the form of massive neutrinos. It has previously been thought that the mass is in the form of low luminosity or dead stars. The value of ρ_B, and hence B, which is considered to be the lowest that can be used to obtain element abundances to compare with observations, is obtained by assuming that all the mass for which we have dynamical evidence is in baryonic form. If that is not so, the value of B can be lower and the primeval value of Y is lower. This might be helpful if the observed value of Y is otherwise lower than the calculated value, although a low value of B might raise problems with the abundance of ^2H. An overproduction of ^2H in the Big Bang would require a much more efficient destruction mechanism in the Galaxy than is generally estimated.

Schramm (this symposium) discusses the detailed comparison of the standard model of the Big Bang with the observations, which are themselves discussed by Pagel, Kinman and Sargent. There is probably no clear discrepancy between theory and observation, although Rana (1982) has recently expressed the opposite view. As this remains a possibility, I turn in the next section to what can be done if theory and observation prove to be in clear contradiction.

3. Variants on the standard model

In this section I discuss two general possibilities. The first is a disagreement between the theoretical element abundances and observations, and the second is a disagreement between the age of the Universe and of galactic objects.

Consider first the possibility that for no value of B, which is compatible with other evidence, do the calculated element abundances agree with the primeval abundances deduced from observation. The obvious solution is then to ask what happens if the conditions $L_e = L_\mu = L_\tau = 0$ are relaxed. If we consider a more general version of the theory, will the discrepancy be removed? The effect of finite L_μ and L_τ is simple. If they are finite there are additional muon and tauon neutrinos (or antineutrinos) to those automatically present in the Big Bang. The effect is to increase the energy density in the early Universe and hence the value of α in (2.17) and (2.19). There is thus an increased production of helium. As any discrepancy between theory and observation at present seems to involve the standard theory's producing too much helium, it is unlikely that the introduction of large $|L_\mu|$ or $|L_\tau|$ alone will be helpful.

The effect of large $|L_e|$ is more complicated. If there is a large excess of ν_e, which are degenerate, the reaction

$$n + \nu_e \rightarrow p + e^- \tag{3.1}$$

is enhanced relative to the reverse reaction. This means that there are fewer neutrons at the time of nucleosynthesis and for large enough L_e essentially pure hydrogen results. In contrast, if there is a large excess of $\bar{\nu}_e$ the reaction

$$p + \bar{\nu}_e \rightarrow n + e^+ \tag{3.2}$$

is enhanced and there is a tendency towards pure neutrons, which later decay to protons so that there is again little helium. In fact, because of the neutron–proton mass difference, the maximum helium production occurs not with $L_e = 0$ but with some excess of $\bar{\nu}_e$ over ν_e. A large value of $|L_e|$ produces a more rapid expansion and cooling and an enhancement of helium production, but this can be more than countered by the degeneracy effects just mentioned.

Although L_μ and L_τ alone cannot have an advantageous influence, it may be necessary to have large values of them as well as L_e in order to bring all of the light-element abundances into agreement with observation. The effect of finite L_e was introduced into the subject by Wagoner *et al.* (1967) and a recent discussion of the role of non-zero lepton numbers is given by David & Reeves (1980). It must be noted that, if the effects of non-zero $|L|$ are to be significant, the number of additional neutrinos must be at least comparable with the number of thermal neutrinos, which implies that $|L| \gtrsim 10^9 B$. This means that a finite neutrino mass places even stronger constraints on such models than on the standard model.

There is an objection to models with large values of $|L|$ which arises from considerations of particle physics in the very early Universe, which will be discussed by Ellis. So far we have regarded B and L as arbitrary numbers, which cannot be determined within the cosmological theory. Recent developments in grand unified theories of particle interactions have challenged that view. They suggest that initially B and L were zero but that they are not strictly conserved and that the present value of B resulted automatically from asymmetric particle processes in the expanding Universe. It is then of course necessary for the theories to produce the observed value of n_B/n_γ. However, the important point for our present discussion is that the theories are likely to require $|B| \approx |L|$; some versions have $L_e - B$ as a conserved quantity although L_e and B are not separately conserved. This would then be a strong argument against $|L| \gg B$ being a solution to our problems. Although the non-conservation of baryon number is not important at any period in the evolution of the Universe that we have discussed, it will ultimately be very important if the Universe is open. The theories suggest that the proton is unstable with half-life *ca.* 10^{31} years so that eventually all matter would be in the form of electrons, positrons, neutrinos and photons.

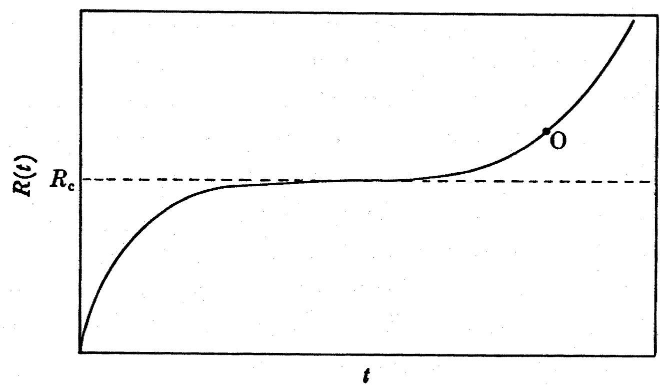

FIGURE 2. The behaviour of $R(t)$ in a $\Lambda > 0$ Universe, with a long coasting period. If the present epoch is marked by O, the age of the Universe exceeds H_0^{-1}.

We next consider what modifications might be made in the theory if the element abundances raised no problems but if there were a clear discrepancy between the ages of galactic objects and the age of the Universe as given by the standard theory. So far I have assumed a zero value for the cosmological constant Λ. Suppose we relax that assumption and consider positive values of Λ. In that case there is an additional long-range repulsive force and it is no longer necessary that $\ddot{R}/R \leqslant 0$. It can be seen from (1.4) that, if $\Lambda > 4\pi G\rho_0$, $(\ddot{R}/R)_0 > 0$ (since $P_0 \approx 0$) and that for sufficiently large values of Λ the present age of the Universe is greater than $1/H_0$ (see figure 2). Because comparison of conditions in the Universe now with those of earlier epochs only depend on R_0/R and not on how long it has taken a change in R to occur, the discussion of element production is not affected.

Whether or not positive Λ cosmologies are capable of resolving a possible timescale problem depends on factors other than element abundances. If $t_0 > 1/H_0$, $q_0 < 0$. Although, as I have said earlier, the observations of q_0 are not good because of the uncertain properties of standard candles, it may prove difficult to attribute all of the difference between an apparently positive q_0 and a genuinely negative q_0 to effects of galactic evolution. In addition, models with $t_0 > 1/H_0$ tend to have $R \approx$ constant for a long period, which means that, as the red shift z of a signal emitted at t and received today satisfies

$$R_0/R(t) = 1+z, \tag{3.3}$$

a small range of red shift corresponds to a large range of time of emission. This means that we might expect to observe a bunching of objects at the particular value of the red shift. Such bunching has not been observed (see, for example, Tytler 1981) and this is an argument against at least the most extreme forms of the $\Lambda > 0$ cosmologies.

4. Deviations from homogeneity and isotropy

In our discussion so far we have assumed that the Universe in the stages relevant to the element production and to our comparison of theory with observations was strictly homogeneous and isotropic. In contrast the Universe today, although homogeneous in some average sense, is extremely inhomogeneous in detail. Any inhomogeneity must produce some small-scale anisotropy in addition to anisotropy that might be present even in a homogeneous Universe. There is no evidence for significant anisotropy of the latter type although no observations can be made further back in time than the decoupling of the microwave radiation. The present status of observations of the microwave radiation are discussed by Wilkinson and Fabbri (this symposium). Two questions that we can ask relating to homogeneity and isotropy are the following. Can the present degree of inhomogeneity and anisotropy arise from a strictly homogeneous initial period and can we assume that the Universe today is necessarily more inhomogeneous and anisotropic than it was at the earlier stages? If the answer to either of these questions is no, we must ask whether there are effects on our discussion of element production.

Consider the second question first, specifically with respect to anisotropy. We believe that the present Universe is close to Robertson–Walker form. As far as isotropy is concerned, the evidence for this comes from counts of galaxies and radio sources and observations of the microwave radiation. The latter observation tells us something about the isotropy of the Universe, and by inference about its homogeneity at the time that the microwave radiation decoupled. There are homogeneous but anisotropic metrics of the Bianchi type that depart very significantly from R–W form at small t but tend to R–W form at large t. Thus the only obvious constraint placed on such a Bianchi model is that it must not demand greater anisotropy in the microwave radiation than is consistent with observations. Hawking & Tayler (1966) and Thorne (1967) originally pointed out that allowable Bianchi models could expand so rapidly in the early stages that essentially no nuclear reactions could occur, so that helium production was suppressed. It was shown by Misner (1968) that the simple discussion given by Hawking & Tayler and Thorne was at fault because there were physical mechanisms, such as neutrino viscosity, that damped anisotropy in addition to the natural transition from a strongly anisotropic to a near R–W metric. Although this affects a comparison between microwave anisotropy and earlier anisotropy, it does not remove the general point that the element abundances might have been affected by strong early anisotropy. A more recent discussion of the relation between anisotropy and element abundances is by Barrow (1976).

Rees (this symposium) discusses the problem of origin of structure in the Universe. It seems clear that galaxies and clusters of galaxies cannot have arisen from a completely structureless early Universe of the type that we have discussed (see, for example, Peebles 1980). Whether or not there was any global anisotropy, it appears that there must always have been a spectrum of inhomogeneities, which have produced the galaxies and clusters. What implications might this have for the light-element abundances? Two types of fluctuation can be distinguished: isothermal and adiabatic. The former are inhomogeneities of the matter distribution in a smooth radiation sea. Because the matter density is negligible at early stages, the Universe is essentially R–W and the overall behaviour is unchanged from our previous description. However, B, and hence the light-element abundances, particularly those of ^{2}H, ^{3}He and ^{7}Li, vary from place to place; this was first discussed by Harrison (1968). Because the matter content of the observable Universe is very small when $T \sim 10^9$ K ($< M_\odot$), it is likely that the abundances will have been smoothed in objects observed today but some residual effects will remain. If, in contrast, the perturbations are adiabatic, the relation between B and T is the same everywhere but a given T is reached at a different time in different places. This corresponds to a variable expansion rate and there must be small modifications in the element abundances (see, for example, Gisler *et al.* 1974). These considerations must ultimately be taken into account when other arguments have identified the perturbation spectrum.

There are other departures from homogeneity that might be important. Following a discussion by Hawking (1971), it has been realized that small black holes might have been produced in the early Universe as an extreme result of departures from strict homogeneity and isotropy. Originally Hawking supposed that all of the primordial black holes would still be present but he then discovered (Hawking 1974) the mechanism of black hole evaporation, which will remove the black holes of lowest mass. Such primordial black holes might have played various roles in the development of the early Universe. The two most relevant to this present meeting are the possibility that particles emitted by black holes close to the epoch of nucleosynthesis might produce spallation reactions modifying the light-element abundances (Lindley 1980) or that the evaporation of black holes might compete with the mechanism discussed by Ellis (this symposium) for the generation of baryon number (see, for example, Zel'dovich 1976).

5. Concluding remarks

In this paper I have discussed the standard model of the hot Big Bang cosmological theory and the tests that it must pass not only if it is to explain the pregalactic abundances of the light elements but also if it is to agree with other relevant observations. The theory apparently possesses one free parameter, n_B/n_γ, although this freedom may be removed by a full understanding of elementary particle physics in the very early Universe. In addition there are other uncertainties in the laws of physics that mean that the theory is not completely defined. It seems possible that the standard theory does give a satisfactory explanation of the origin of the light elements but, if there are problems, it can apparently be generalized by the introduction of large non-zero lepton numbers or a cosmological constant. In fact both of these degrees of freedom may be apparent rather than real. If grand unified theories of particle physics determine the value of n_B/n_γ, they may also preclude large values for the lepton numbers. In addition, the observational constraints against the $\Lambda > 0$ models are already quite severe.

The standard model is a description of a Universe that is strictly homogeneous and isotropic, whereas in the real Universe at the present time there are very significant departures from homogeneity and isotropy. It was originally hoped that the present structure could have arisen from statistical irregularities in a homogeneous and isotropic Universe, but it now seems clear that there must have been a spectrum of perturbations at a very early epoch. The necessary form of such perturbations may be identified when the process of galaxy formation and clustering is fully understood. If the Universe is irregular at the time of nucleosynthesis, there must also be a variation in the element abundances from place to place and also a change in the average abundances. Although such variations are likely to be small, they must be considered in any final comparison between theory and observation.

I have not considered all possible variants of the Big Bang theory. Our critical assumption that the laws of physics are unchanging with time has ruled out consideration of the variable G cosmologies and of their modifications to the primeval element abundances (see, for example, Meisels 1982). We have also naturally not considered the possibility that the Universe had a cold dense origin and that both the light-element abundances and the microwave radiation were produced through thermonuclear reactions in pregalactic population III stars (Rees 1978; Hogan 1982). I should, however, mention that observational indications that the microwave radiation may be not exactly thermal with excess radiation near the frequency of peak emission have led to many suggestions that the excess microwave radiation, and some part of the light elements, may have been produced in pregalactic stars, which are an important ingredient in some discussions of the origin of galaxies and clusters. In effect this places an additional step in the process of deducing primeval abundances from the abundances observed today and it might therefore modify the constraints on the standard model.

References

Alpher, R. A., Bethe, H. A. & Gamow, G. 1948 *Phys. Rev.* **73**, 803–804.
Barrow, J. D. 1976 *Mon. Not. R. astr. Soc.* **175**, 359–370.
Bondarenko, L. N., Kurguzov, V. V., Prokof'ev, Yu. A., Rogov, E. V. & Spivak, P. E. 1978 *Soviet Phys. JETP Lett.* **28**, 303–307.
Byrne, J., Morse, J., Smith, K. F., Shaikh, F., Green, K. & Greene, G. L. 1980 *Phys. Lett.* B **92**, 274–278.
Christensen, C. J., Nielsen, A., Bahnsen, A., Brown, W. K. & Rustad, B. M. 1972 *Phys. Rev.* D **5**, 1628–1640.
David, Y. & Reeves, H. 1980 *Phil. Trans. R. Soc. Lond.* A **296**, 415–429.
Gisler, G. R., Harrison, E. R. & Rees, M. J. 1974 *Mon. Not. R. astr. Soc.* **166**, 663–671.
Harrison, E. R. 1968 *Astr. J.* **73**, 533–534.
Hawking, S. W. 1971 *Mon. Not. R. astr. Soc.* **152**, 75–78.
Hawking, S. W. 1974 *Nature, Lond.* **248**, 30–31.
Hawking, S. W. & Tayler, R. J. 1966 *Nature, Lond.* **209**, 1278–1279.
Hogan, C. J. 1982 *Astrophys. J.* **252**, 418–432.
Hoyle, F. & Tayler, R. J. 1964 *Nature, Lond.* **203**, 1108–1110.
Lindley, D. 1980 *Mon. Not. R. astr. Soc.* **193**, 593–601.
Lyubimov, V. A., Novikov, E. G., Nozik, V. Z., Tretyakov, E. G. & Kosik, V. S. 1980 *Phys. Lett.* B **94**, 266–268.
Meisels, A. 1982 *Astrophys. J.* **252**, 403–409.
Misner, C. W. 1968 *Astrophys. J.* **151**, 431–457.
Olive, K. A., Schramm, D. N., Steigman, G., Turner, M. S. & Yang, J. 1981 *Astrophys. J.* **246**, 557–568.
Peebles, P. J. E. 1980 *The large scale structure of the Universe*. Princeton University Press.
Penzias, A. A. & Wilson, R. W. 1965 *Astrophys. J.* **142**, 419–421.
Rana, N. C. 1982 *Phys. Rev. Lett.* **48**, 209–212.

Rees, M. J. 1978 *Nature, Lond.* **275**, 35–37.
Steigman, G. 1976 *A. Rev. Astr. Astrophys.* **14**, 339–372.
Thorne, K. S. 1967 *Astrophys. J.* **148**, 51–68.
Tytler, D. 1981 *Nature, Lond.* **291**, 289–292.
Wagoner, R. V., Fowler, W. A. & Hoyle, F. 1967 *Astrophys. J.* **148**, 3–49.
Weinberg, S. 1972 *Gravitation and cosmology*. New York: J. Wiley.
Zel'dovich, Ya. B. 1976 *Soviet Phys. JETP Lett.* **24**, 25–28.

Discussion

R. Fabbri (*Istituto di Fisica Superiore, Università di Firenze, Italy*). I heard somewhere that more refined measurements of the neutrino mass may be performed in the near future; can Professor Tayler tell us more on this point? Also, in connection with homogeneous models I would like to remark that if the intrinsic anisotropy of the 3 K background on large angular scales is $\Delta T/T \approx 10^{-4}$, then the decaying mode of anisotropy cannot account for it. However, growing modes of *homogeneous* anisotropy exist that could provide a viable explanation for the quadrupole terms.

R. J. Tayler. There are several measurements aimed at measuring the mass of the electron neutrino but these will not constrain the masses of the muon and tauon neutrinos, and it is the sum of the masses that is important.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES