Mon. Not. R. Astron. Soc. **286**, 173–182 (1997)

On the Hubble constant and the cosmological constant

F. Hoyle,[1] G. Burbidge[2] and J. V. Narlikar[3]

[1] *102 Admirals Walk, Bournemouth BH2 5HF*
[2] *Center for Astrophysics and Space Sciences and Department of Physics, University of California, San Diego, La Jolla, California 92093-0424, USA*
[3] *Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Ganeshkind, Pune 411 007, India*

Accepted 1996 October 28. Received 1996 June 21; in original form 1996 January 18

ABSTRACT

We review the observational determinations of the Hubble constant which have been made in recent years. We conclude that the most likely value of H_0 is 58 km s^{-1} Mpc^{-1} with uncertainties of $+10$ and -5. Thus the age of the standard big bang model is 11.2 Gyr. The discrepancy between this value and the ages of the oldest observed stars, 13–16 Gyr, appears to be real, necessitating some change in the standard model.

A currently favoured procedure for coping with this widely-admitted difficulty for the theory which has been favoured by many cosmologists in recent years is a rebirth of the cosmological constant λ. Even with this constant, the observations constrain the model very severely. There are theoretical considerations as well. The problem with this constant, as it has been seen over much of the past half-century, is that it is required to have a physical dimensionality of $(\text{length})^{-2}$ and to have a magnitude of about 10^{-56} cm^{-2}. Theoreticians have not favoured introducing such a quantity *ab initio* into cosmology, but attempts to explain the genesis of λ from particle physics have yielded results that are wide of what is required by immense factors ($\sim 10^{50}$ to $\sim 10^{100}$).

Using an approach from a scale-invariant theory of gravity, we show that λ can be derived correct to a factor of ~ 2 within the modern Universe. This derivation does not appear to be applicable to earlier phases of the Universe, which give $\lambda \simeq H^2$ rather than the relation $\lambda \simeq H_0^2$ that a true cosmological constant would require.

Key words: cosmology: miscellaneous – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

The recent determinations of Hubble's constant by various teams of observers have led to considerable flurry amongst theoreticians as to what is the correct cosmological model describing the large-scale structure of the Universe. In this context the cosmological constant has been revived once again, as it provides an enlarged parameter space for fitting the various observations.

In this paper we will review the status of H_0 (Hubble's constant) and λ (the cosmological constant) as perceived against today's backdrop of observations. Since these parameters are linked to cosmological models, we will begin with a brief summary of the latter.

(i) The standard hot big bang cosmology, with or without inflation.

(ii) The modified big bang cosmology including the cosmological constant.

(iii) The quasi-steady state cosmology proposed by the three of us in recent years (Hoyle, Burbidge & Narlikar 1993, 1994a,b, 1995; Sachs, Narlikar & Hoyle 1996).

In all the cosmologies above, the simplest models assume the Universe to be homogeneous and isotropic, with the space–time metric given (in standard notation) by the Robertson–Walker line element

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

the speed of light taken as unity, and $k=0, \pm 1$.

(i) The standard models. The classical hot big bang cosmologies follow from Einstein's equations of general relativity.

In the dust approximation, we get for matter density ρ

$$\frac{\dot{S}^2+k}{S^2}=\frac{8\pi G\rho}{3}, \tag{2}$$

$$\rho \propto S^{-3}. \tag{3}$$

For the inflationary version, $k=0$ and $S \propto t^{2/3}$. The density is given by

$$\rho=\frac{3H^2}{8\pi G}\equiv\rho_c, \tag{4}$$

where $H\equiv\dot{S}/S$ is the Hubble constant of the epoch. ρ_c is called the *closure* or *critical density*.

For $k\neq 0$, we write

$$\rho=\Omega\rho_c, \tag{5}$$

where Ω is the density parameter. For closed models $(k=1)\,\Omega>1$, while for open models $(k=-1)$, $0\leq\Omega\leq 1$. The age of the Universe, i.e. the time elapsed from $S=0$ to the present epoch, is given in the $k=0$ case by

$$t_0=\frac{2}{3H_0}, \tag{6}$$

H_0 being the present value of Hubble's constant. For $k=1$ models the age is less than this value, whereas for $k=-1$ it lies between $\frac{2}{3}H_0^{-1}$ and H_0^{-1}.

(ii) *The λ-cosmologies.* Here the Einstein equations include the λ-term and (3) is replaced by

$$\frac{\dot{S}^2+k}{S^2}-\frac{1}{3}\lambda=\frac{8\pi G\rho}{3}. \tag{7}$$

By adjusting λ one may get cosmological models with arbitrarily large ages. The limiting model in this series is the static Einstein Universe (Einstein 1917) which has $S=\text{constant}=S_0$ (say) and $k=1$ with

$$\rho=\frac{\lambda}{4\pi G},\ \lambda=\frac{1}{S_0^2}. \tag{8}$$

It is this property of the λ-models that has made them popular at the current times. Although λ can be of either sign, traditionally it is taken as positive.

(iii) *The quasi-steady state cosmology.* We will discuss the theoretical foundations of this cosmology in a later section. It requires a cosmological constant of the opposite sign, compared with (ii) above. In addition, the theory describes matter creation by means of a scalar field c. The creation activity has ups and downs, leading to a short-term oscillatory pattern superposed on a long-term exponential expansion. In place of equations (3) or (7) we have

$$\frac{\dot{S}+k}{S^2}-\frac{1}{3}\lambda=\frac{8\pi G\rho}{3}-2\pi Gf\dot{c}^2, \tag{9}$$

where f is a coupling constant of the c-field to gravity. We also have $\rho\propto S^{-3}$ and $\dot{c}\propto S^{-2}$ during an oscillation. Sachs et al. (1996) have found the solution of (9) to be

$$S(t)=\exp\left(\frac{t}{P}\right)[1+\eta\cos\theta(t)], \tag{10}$$

where P is a constant and $\theta(t)$ is a known function with a period $Q\ll P$. η is a parameter with $[\eta]<1$. The function $\theta(t)$ is almost linear $\approx 2\pi t/Q$. *Note, however, that the cosmological constant is negative in this theory.*

The quasi-steady state cosmology relates the dynamical behaviour of the Universe to creation of matter, which takes place at epochs of minima of the scalefactor S. Hoyle et al. (1994a) found that a good fit with observations is obtained for

$$P=20Q,\qquad Q=4\times10^{10}\ \text{yr},\qquad \eta=0.75,$$

present epoch $t_0=0.85Q$.

With this set of values, the time span from the last minimum to the present epoch is 14×10^9 yr. The Universe itself is, however, infinitely old and objects older than 14×10^9 yr can be easily accommodated in it.

With this theoretical introduction we now turn to the observational scenario.

2 THE VALUE OF H_0 DETERMINED FROM OBSERVATION

From the time of the early work of Hubble and Humason in the 1930s up to ~1950, the value of the Hubble constant was thought to be about 530 km s^{-1} Mpc^{-1} (Hubble 1936). Problems associated with this value of H_0 first appeared in the 1940s, and it became clear that there were two classes of pulsating stars, the classical Cepheids of Population I and the RR Lyrae stars of Population II, which have two distinct period–luminosity relations (Lundmark 1946, 1948; Baade 1952; Mineur 1952). This had the effect of doubling the distance scale. The realization that the most luminous stars in galaxies were much brighter than had originally been supposed, and also the demonstration by Humason, Mayall & Sandage (1956) that the star-like objects in external galaxies reported by Hubble were actually H II regions illuminated by more than one star, increased the distance scale further. In his Warner lecture, Sandage then showed that the value of H_0 should be reduced to about 75 km s^{-1} Mpc^{-1} (Sandage 1958). However, at IAU Symposium No. 15, Sandage (1962) summarized all of the values that were available which led to an estimate for H_0 of 98 ± 15 km s^{-1} Mpc^{-1}.

In the 1970s an extensive program to measure H_0 more accurately was started by Sandage & Tammann (1974a). They calibrated the linear sizes of H II regions as a function of spiral galaxy luminosities, and went on to determine the distances to 39 (Sc-Sd-Sm-Ir) galaxies using these values. The adopted absolute magnitudes combined with the apparent magnitudes of the Virgo spirals gave a distance to the Virgo cluster of 19.5 ± 0.8 Mpc. The Hubble diagram for first-ranked ellipticals in the Coma cluster gave a redshift for the Virgo cluster of 1111 ± 75 km s^{-1}. These values combined gave what Sandage & Tammann called a first hint at the value of $H_0=57\pm6$ km s^{-1} Mpc^{-1} (Sandage & Tammann 1974b).

Over the last twenty years, Sandage & Tammann have used a variety of methods to determine H_0 and have con-

Table 1. Estimates of the distance to the Virgo cluster.

Method	Distance		
	Sandage (1995)	van den Bergh (1992)	Jacoby et al. (1992)
Globular clusters	21.1 ± 2	19.7 ± 2.3	18.8 ± 3.8
Novae	20.6 ± 4	18.2 ± 2.5	21.1 ± 3.9
Supernovae	21.2 ± 2.2	19.1 ± 6 22.9 ± 5	19.4 ± 5
$D_n - \sigma$	23.4 ± 2	–	16.8 ± 2.4
21 cm line widths	20.9 ± 1.4	15.0 ± 1.4	15.8 ± 1.5
Size of the Galaxy	20.0 ± 1.8	–	–
Size of M31	–	17.0 ± 4	–
Size of M33	–	10.5 ± 2.5	–
Size of LMC	–	12.0 ± 2.5	–
Surface Brightness Fluctuations	–	14.9 ± 0.9	15.9 ± 0.9
Planetary Nebulae	–	14.1 ± 0.3	15.4 ± 1.1
Red Supergiants in NGC 4571	–	13.8	–
Red Supergiants in NGC 4523	–	13.2	–

cluded that the value has converged to $50 < H_0 < 60$ km s^{-1} Mpc^{-1}. One major key to the issue is the distance to the centre of mass of the Virgo cluster, which they believe that they have shown by a variety of methods to lie close to 20 Mpc. In Table 1 we show how different methods described by Tammann (Sandage 1995) have led to this result. If the true redshift of the Virgo cluster is ~ 1100 km s^{-1}, a value of $H_0 \sim 55$ km s^{-1} Mpc^{-1} is obtained. Many other methods give a similar result. We show in Table 2 a summary of the results of the determinations of H_0 also from Sandage (1995).

Is there a reasonable case to be made for saying that this value is in error? The two experts contemporary with Sandage & Tammann who have been most outspoken in their disagreement have been G. de Vaucouleurs and S. van den Bergh. Working actively in the 1970s and 1980s on this problem, de Vaucouleurs concluded that $90 \leq H_0 \leq 100$ km s^{-1} Mpc^{-1} (de Vaucouleurs 1982). In a comparatively recent review, van den Bergh (1992) concludes that $H_0 = 76 \pm 9$ km s^{-1} Mpc.

To give some idea of the different values obtained by using many of the same methods, we show in Table 1 van den Bergh's compilation of distances to the Virgo cluster using some of the same methods used by Sandage & Tammann and also different methods.

Table 2. Values of H_0 (Sandage 1995).

Method	H_0 [$km\ s^{-1} Mpc^{-1}$]
Virgo Distance	55 ± 2
ScI Hubble Diag.	49 ± 15
M101 Diameters	43 ± 11
M31 Diameters	45 ± 12
Tully-Fisher	48 ± 5
Supernovae (B)	52 ± 8
Supernovae (V)	55 ± 8
Unweighted mean	50 ± 2
Weighted mean	53 ± 2

A recent discussion of the different methods of distance determination has also been given by Jacoby et al. (1992). This paper is interesting in that it contains analyses of many methods, including comparatively new methods recently proposed by some of the authors themselves. Jacoby et al. have given a table of distance determinations of the Virgo cluster which we show also in Table 1. Jacoby et al. finally conclude that H_0 is 80 ± 11 or 73 ± 11 km s^{-1} Mpc, depending on different weighting factors.

It can be seen by a comparison of the values of the distance to Virgo in Table 1, that different methods sometimes give different results, and the same methods in the hands of different investigators can also give different results. In general this is because there is intrinsic dispersion in the properties of the 'standard' objects chosen.

There is also a sociological problem which has developed between the various groups, especially between Sandage & Tammann on the one hand, and van den Bergh and the younger investigators on the other. Thus despite the fact that the differences are not large, the current view is that we are witnessing a competition between Sandage & Tammann, who have consistently claimed for twenty years that a 'low' value of H_0 is correct, and the rest who favour a 'higher value' of H_0 of 70–80 km s^{-1} Mpc^{-1}. A major issue is the structure and distance to the Virgo cluster.

Observational astronomers are aware that not only is the Virgo cluster widely extended but it appears to involve two subclusters, one predominantly involving spirals which is more extended than the other subcluster of ellipticals (cf. de Vaucouleurs 1961). De Vaucouleurs argued that there were two different clusters involved. Thus it is clear that the determination of the distance of any single galaxy will not give a good measure of the distance to the centre of mass of the cluster. This may explain the dispersion between some of the values in Table 1.

A second question is, what is the true redshift of the Virgo cluster? It is now generally believed that the Local Group galaxies involving predominantly our Galaxy, M31 and M33, are perturbed by the large mass of the local supercluster and there is a component of velocity arising from this effect. Attempts to correct for this velocity term have led to various values being used for the 'true' redshift of the Virgo cluster.

A good example of the way that these uncertainties can give rise to very different values of H_0 is given by the recent much-quoted and publicized work by Freedman et al. (1994) when compared with the results of Sandage & Tammann. Freedman et al. derive the distance of *one* galaxy in the Virgo cluster from *Hubble Space Telescope* (*HST*) observations of Cepheids, and use the distance of that galaxy M100 (17.1 ± 1.8 Mpc) combined with the redshift of the Virgo cluster which they take to be $+1404$ km s^{-1} to obtain a value of H_0 of 80 ± 17 km s^{-1} Mpc^{-1}. On the other hand, if we take the distance of the centre of mass of the Virgo Cluster according to Sandage & Tammann to be 21 Mpc (Table 1) and a redshift of 1179 km s^{-1}, which Sandage & Tammann have obtained by plotting a Hubble diagram for more distant clusters (Sandage & Tammann 1990) and extrapolating inward, $H_0 = 56$ km s^{-1} Mpc^{-1}.

In a very recent review (Kennicutt, Freedman & Mould 1995), unfortunately still only involving authors from one side of this controversy, a more moderate approach than that of Freedman et al. (1994) is taken, and it is concluded that H_0 is currently constrained to 80 ± 17 km s^{-1} Mpc^{-1}, and that when distances have been obtained in ~ 25 galaxies from observations of Cepheids using *HST* and combined with several secondary distance indicators, H_0 will be determined to ± 10 per cent.

Having studied all of the literature, much of it not referenced here, and having taken account of what we believe are genuine uncertainties, and also allowing for the prejudices of our friends, colleagues and ourselves, what is our best estimate of the final value of H_0?

In our opinion, the best standard candle appears to be the absolute magnitude at maximum of supernovae of Type Ia (SNIa). Thus a key program is to find Cepheids in galaxies in which those supernovae have been reported and derive their absolute magnitudes at maximum light. The data have shown (Branch & Tammann 1992) that plotting apparent magnitudes of SNIa against log *cz* gives a very tight relation with a slope of 5 (Sandage & Tammann 1993; Hamuy et al. 1995; Reiss et al. 1995), the intrinsic dispersion being ≤ 0.3 mag. Thus the determination of $\bar{M}$ (max) should give a good value for H_0, since supernovae can be detected in galaxies far beyond the region of the supercluster where perturbations are present.

Galaxies in which supernovae of Branch Type Ia have been identified and have had distances derived now include IC 4182 (SN 1937C), NGC 5253 (SN 1895B & SN 1972E), NGC 4536 (SN 1981B), NGC 4496 (SN 1960F), NGC 4639 (SN 1990N) and NGC 3627 (SN 1989B) (Sandage et al. 1996). This work shows that when the corrections are included, they are small and $\bar{M}_B(\max) = -19.47 \pm 0.07$ and $\bar{M}_V(\max) = -19.48 \pm 0.07$. This gives $H_0 = 56 \pm 4$ km s^{-1} Mpc^{-1} from $\bar{M}_B(\max)$ and 58 ± 4 from $\bar{M}_V(\max)$.

We have used the work of Sandage, Saha & Tammann, etc. because we believe that the *supernova method is the best currently available*.

Having put everything together we conclude that the final value for H_0 will lie in the range 58^{+10}_{-5} km s^{-1} Mpc^{-1}. We believe therefore that Sandage & Tammann were correct in their original estimates some 20 years ago. On this basis $H_0^{-1} \approx 16.9$ Gyr, and the age of the Universe in the standard model, $2/3H_0 = 11.2$ Gyr.

How old are the galaxies in the Universe? The only methods we have for determining ages are using our understanding of stellar evolution to determine the ages of star clusters, or using our knowledge of nucleosynthesis to determine the ages of the elements. As far as hard evidence is concerned, we are restricted to stellar systems which are very close to us, i.e. in our Galaxy or very nearby. We give a brief discussion of those age estimates in the following section.

3 AGES FROM STELLAR EVOLUTION AND THE AGE OF THE ELEMENTS

Age determinations for globular clusters in our Galaxy have now been made for about forty years. Over the last twenty years there has been agreement that the oldest clusters have ages in the range of 14–19 Gyr (for example see Demarque & McClure 1977; Sandage 1993; Chaboyer 1995; and many other articles). A careful analysis of the best data (van den Berg, Stetson & Bolte 1996) suggests that the oldest globular clusters, which are assumed to be those with the lowest metal abundances, are about 15 Gyr old with an uncertainty of less than ± 20 per cent. It also appears that there is little or no dispersion in age among the clusters, either as a function of the Fe/H ratio or as a function of distance to the Galactic Centre. Thus it looks as though all of those systems formed in a short time, perhaps in the way suggested in the classic paper of Eggen, Lynden-Bell & Sandage (1962).

What about the age of the elements? We know that the age of the Solar system is about 4.6 Gyr, and that the chemical elements were made in stars before this (Burbidge et al. 1957). The nuclear chronometers that can be used are ^{232}Th, ^{235}U, and ^{238}U (Rutherford 1929; Burbidge et al. 1957; Fowler & Hoyle 1960) and ^{187}Re (Clayton 1964). The age determination using U and Th was originally found to be ~ 10 Gyr at minimum (Burbidge et al. 1957). More recently Fowler (1987), assuming a constant nucleosynthesis rate up to the time the solar system formed, obtained 10.0 ± 1.6 Gyr. Using slightly different assumptions Cowan, Thielemann & Truran (1987) obtained values of 12.4–14.7 Gyr. A detailed analysis of all of the investigations by Cowan, Thielemann & Truran (1991) has led to the conclusion that the age determinations based on realistic models lie in the range 13–15 Gyr, while values in the range 10–20 Gyr cannot be excluded.

Thus it is clear that the age of the elements is compatible with the ages derived for the oldest star clusters in the sense that it is reasonable to suppose that the stars which gave rise to the heavy elements were able to synthesize them early in the history of the Galaxy.

What is the range of ages of galaxies in the Universe? There are many systems in which the dominant features are H II regions giving rise to strong emission lines, and this suggests that they are young, with ages $\ll 10^{10}$ yr. The nature of the underlying continuum radiation in these systems is not known. Information can be obtained from the form of the energy distribution measured by the colours. We can only get some measure of age from the colours if we assume that the form of the mass distribution function for stars is the same everywhere and is similar to that in our own Galaxy. However, such an argument is extremely naive.

Thus for the starburst galaxies we can either conclude that the whole system has an age $\ll 10^{10}$ yr, or that it has an underlying stellar population with stars as old, or even older, than the oldest stars in our Galaxy.

By the same token we cannot exclude the possibility that many whole galaxies are much older than $\sim 10^{10}$ yr. Such galaxies will largely be made up of evolved stars and stars with masses below about 0.7 $M_\odot$ which have not yet evolved off the main sequence.

However, using the most conservative assumption we now have good evidence that the ages of the oldest stars in clusters in our Galaxy are greater than the age derived from the standard cosmological model, so that a positive cosmological constant is indicated.

4 THE COSMOLOGICAL CONSTANT

From the early days when, in 1917, Einstein introduced the cosmological constant to accommodate a static model of the universe in the framework of general relativity, only to abandon it later as the 'greatest blunder' in his life, the big bang cosmologists have had a love/hate relationship with this parameter. It is invoked when observations cannot be satisfied by the models without it, while it is forgotten when some other remedy becomes available. We have just demonstrated that, based on the best data, the basic conflict is back and the cosmological constant is again in vogue.

Twenty years ago Gunn & Tinsley (1975) took stock of the observational situation and concluded: 'New data on the Hubble diagram, combined with constraints on the density of the universe and the ages of galaxies, suggest that the most plausible cosmological models have a positive cosmological constant, are closed, too dense to make deuterium in the big bang, and will expand for ever...'. The situation changed in a few years, however, with the advent of particle physicists into cosmology, the inflationary models, and the discovery of the so-called dark matter. In recent years, the *COBE* observations have added another dimension to the problem. How does the theory versus observation scenario look today?

An exercise along the lines of Gunn & Tinsley was carried out recently by Bagla, Padmanabhan & Narlikar (1995, hereafter BPN). We summarize below their main conclusions.

Using k for the curvature parameter, and denoting the dimensionless matter density parameter as Ω_0 and the cosmological constant parameter as Ω_λ,

$$\Omega_0 = \frac{8\pi G \rho_0}{3H_0^2}, \qquad \Omega_\lambda = \frac{\lambda c^2}{3H_0^2}, \tag{11}$$

where for this analysis we put $H_0 = 100\,h$ km s^{-1} Mpc^{-1}.

BPN assumed that the currently fashionable standard big bang model with inflation predicts

$$\Omega_0 + \Omega_\lambda = 1, \qquad k = 0. \tag{12}$$

In addition, they supposed that the model canonically predicts an initial power spectrum for the wave numbers of inhomogeneities in the form

$$\rho(\kappa) \propto \kappa^n, \qquad n = 1. \tag{13}$$

The spectrum gets modified at small scales by different physical processes and this change is described by a transfer function. BPN used the transfer function of Efstathiou, Bond & White (1992). They also took into consideration the *COBE* data for normalizing the power spectrum. In addition to the above model, BPN also considered the case $\Omega_0 < 1$, $\Omega_\lambda = 0$, $k = -1$ of the open universe without the cosmological constant.

Apart from the ages of the globular clusters and the measured value of the Hubble constant discussed earlier, BPN also considered the data on mass per unit volume in rich clusters identified from X-ray observations. One way of comparing theory with observations is to convert the number density of clusters into an amplitude of density fluctuations. This amplitude is then scaled to a typical cluster scale of $8\,h^{-1}$ Mpc assuming a power-law form for σ, the root mean square (rms) fluctuation in density perturbations. The index is chosen to match that expected in the model being considered (cf. White, Efstathiou & Frenk 1993). The result is expressed as a constraint on the parameter σ_8, which is the value of σ at $8\,h^{-1}$ Mpc, that is, it represents the rms fluctuation in mass density within spheres of radii $8\,h^{-1}$ Mpc.

Another constraint considered by BPN is the baryon mass fraction of clusters, along with the ceiling value for baryon density from primordial nucleosynthesis. This places limits on the mass density parameter Ω_0. Finally, another constraint is provided by the existence of high redshift objects with damped Lyα systems (DLAS). This tells us that the amplitude of density perturbations is of order unity at

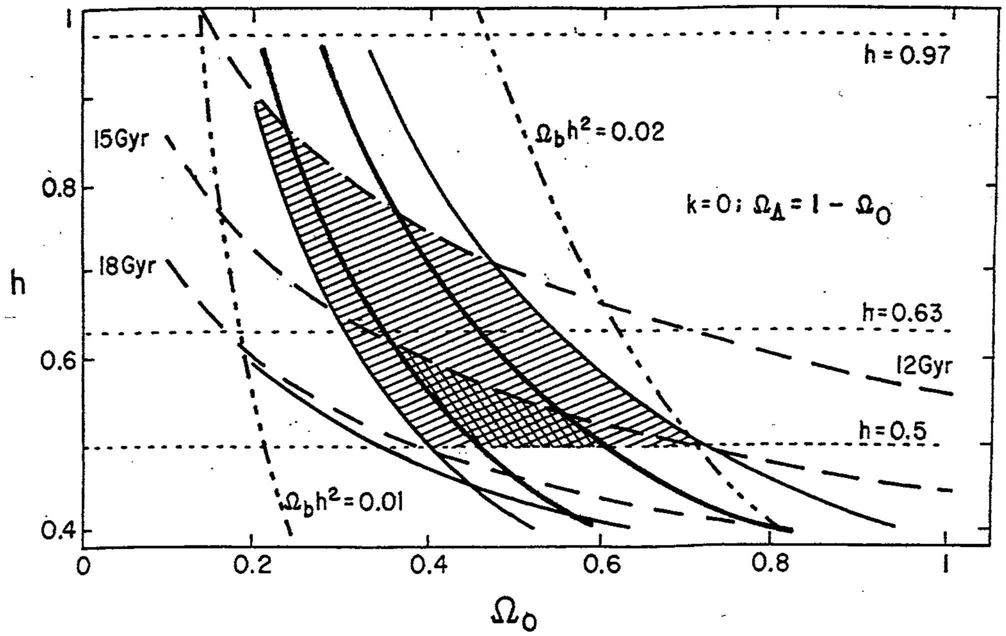

Figure 1. The figure shows the permitted ranges of parameters Ω_0 and h for inflation-supported models with curvature parameter $k=0$, $\Omega_0+\Omega_{\mathscr{A}}=1$. The density parameter Ω_0 includes baryonic and dark matter in the density measured as a fraction of the closure density, whereas h measures Hubble's constant in units of 100 km s^{-1} Mpc^{-1}. The dashed lines show models of constant specified ages marked on the curves. The dot-dashed curves indicate the extreme upper limits on baryonic densities allowed by primordial nucleosynthesis and baryonic mass fractions contributed by clusters. The thick unbroken lines enclose the region permitted by the observed abundance of clusters. The thin unbroken lines show the extent to which this region shifts if one allows for the present uncertainty in the *COBE* normalization of the power spectrum. The shaded region is permitted by these constraints for models with ages exceeding 12 Gyr and $h>0.5$. The cross-hatched region allows for age > 15 Gyr and the constraints of cluster abundance, but without taking full benefit of the *COBE* normalization constraint. However, even the cross-hatched area shrinks further if constraints from the deceleration parameter (not shown here) are taken into account. (Figure supplied by the courtesy of Bagla et al. 1995.)

$M\approx 10^{11}\,M_\odot$ at redshift $z=2$. The theoretical value of the density parameter should be greater than or equal to the observed value, as not all systems in the relevant mass range host a DLAS.

In Fig. 1 we reproduce the upper panel of fig. 3 of BPN. In the figure, plotting h against Ω_0, the cross-hatched area gives the permitted possible values of these two parameters when:

(i) the ages of the globular clusters are 15×10^9 years or more;
(ii) H_0 is considered to be possibly as low as 50 km s^{-1} Mpc^{-1}, and
(iii) the constraints on high redshift objects mentioned in the previous paragraph confine the possible values to lie between the two thick solid lines.

Weakening the observational requirements to the limit of what seems possible, and in particular lowering the ages of the oldest stars to 12×10^9 yr, permits the shaded area as possible for h and Ω_0.

The lower panel of fig. 3 of BPN (not reproduced here) shows the corresponding situation for the open Friedman model with $k=-1$, $\lambda=0$, where there is no cross-hatched region, no region with what are considered the most likely observational values and constraints. No model satisfies the most likely values without the cosmological constant, while even with the cosmological constant there is very little room for manoeuvre left.

However, far from being 'cornered', the big bang supporters tend to play down the constraints. For example, Liddle et al. (1995) conclude that 'there is a substantial parameter space still viable for these models', but their figures require $H_0 \lesssim 50$ km s^{-1} Mpc^{-1} and stellar ages $\lesssim 14\times 10^9$ yr, and in Sections 1 and 2 we have shown that this is unlikely. The reality is that there is only allowable parameter space in Fig. 1 if the uncertainties in the observed quantities are all in the right direction.

Even on theoretical grounds the deduction that, after inflation was over, a residual cosmological term remained of just the right magnitude to enable the model to fall within the tiny cross-hatched parameter space of Fig. 1 smacks of fine tuning. For, as has been pointed out by several authors (cf. Weinberg 1989), it means that from the vacuum value operating during inflation, the residual has managed to acquire the present value as a tiny fraction of a few parts in 10^{108}.

5 THE GENESIS OF THE COSMOLOGICAL CONSTANT

In this section, we derive what we consider to be a remarkable result, a deduction of the magnitude of the cosmological constant, not adrift by a factor $\sim 10^{108}$, but very close to the astronomically required value to within a factor of the order of 2. First, however, we begin a discussion of the way that λ is introduced into general relativity. As will be seen, this is an ad hoc affair with little to recommend it. Thus general relativity containing λ is obtained by applying the principle of least action, $\delta\mathscr{A}=0$ with respect to variations of the metric tensor, where $\mathscr{A}$ is given by

$$\mathscr{A}=\frac{1}{16\pi G}\int(R-2\lambda)\sqrt{-g}\,\mathrm{d}^4x-\sum_a\int m_a\,\mathrm{d}a. \qquad (14)$$

Here R is the contracted Ricci tensor, m_a is the mass of particle a, da is an element of proper length along its path, and the material content of the Universe is approximated as a set of point particles $a, b, c\ldots$.

We now discuss a gravitational theory that both contains the equations of general relativity and yields λ for the modern Universe as well as G, not as quantities inserted ab initio but as derived quantities. The circumstance that it is possible to find such a theory suggests, in our opinion, that it should supersede general relativity also in its application to cosmology.

Our mathematical object is to express gravitational theory in a scale-invariant form. A scale change implies a metric change from $\mathrm{d}s^2=g_{ik}\,\mathrm{d}x^i\,\mathrm{d}x^k$ to $\mathrm{d}s^2=\Omega^2(x)\,g_{ik}\,\mathrm{d}x^i\,\mathrm{d}x^k$, where $\Omega(x)$ is a scalar function of the space–time position x, subject to the conformal conditions that Ω is neither zero nor infinity. Since all of physics except general relativity is indeed invariant to such transformations, we think it is a reasonable objective to seek a gravitational theory that both contains general relativity and possesses this additional form of invariance. We are encouraged in this attempt by G appearing as a derived quantity, not assumed ab initio, and by there being a place also for λ in the modern Universe. First, however, we note that the action of general relativity (see equation 4) is not invariant to changes of scale of the above kind, and so cannot lead to scale-invariant gravitational equations.

It often happens that in seeking some new mathematical objective, one has to start in some apparently strange direction, and this is the case here. For each of the particles a, $b,\ldots$, obtain a field $\varphi^{(a)}$, $\phi^{(b)}(x),\ldots$, determined as the fundamental solution (Courant & Hilbert 1950) of the partial differential equations

$$\Box_x\phi^{(a)}(x)+\frac{1}{6}R(x)\,\varphi^{(a)}=\int\frac{\delta_4(x,A)}{\sqrt{-g(A)}}\,\mathrm{d}a, \qquad (15)$$

etc. for b, $c,\ldots$. Here $\delta_4(x,A)$ is the 4-dimensional delta function.

Next define a mass field $m(x)$ by

$$m(x)=\sum_{a,b,\ldots}\phi^{(a)}(x). \qquad (16)$$

Then it can be proved that a scale transformation specified by $\Omega(x)$ changes $m(x)$ to $m(x)\,\Omega^{-1}(x)$, which is the condition required for scale invariance in quantum mechanics.

Now write the action for the set of particles $a, b,\ldots$ as

$$\mathscr{A}=-\sum_{a,b,}\int m(A)\,\mathrm{d}a, \qquad (17)$$

which is evidently unchanged by a scale transformation. Thus $m(A)$ is changed to $m(A)\,\Omega^{-1}(A)$ while da at point A is changed to $\Omega(A)$ da. *The very simple form (equation 7) is to replace the far more complex action in general relativity*. It is a matter of considerable surprise that varying $\mathscr{A}$ with respect to the metric tensor and applying $\delta\mathscr{A}=0$ should lead to gravitational equations that include general relativity and also contain properties, such as scale invariance, that go outside general relativity.

The set of equations resulting from a considerable analysis is

$$K\left(R_{ik}-\frac{1}{2}g_{ik}R\right)=-T_{ik}+m_im_k-\frac{1}{2}g_{ik}g^{pq}m_pm_q$$
$$+g_{ik}\Box K-K_{;ik}, \qquad (18)$$

in which $K=\frac{1}{6}m^2$. The details of the derivation of (18), which equations are indeed scale-invariant, have been given in earlier publications (for example, Hoyle & Narlikar 1966).

Now make the scale change

$$\Omega(x)=m(x)/m_0, \qquad (19)$$

m_0 being some constant. Then the mass field $m(x)$ becomes $m(x)\,[m(x)/m_0]^{-1}$, i.e. m_0 everywhere. With the consequence that the derivative terms drop out of (18), and *defining* G by

$$G=\frac{3}{4\pi m_0^2}, \qquad (20)$$

the gravitational equations become formally the same as those of general relativity,

$$R_{ik}-\frac{1}{2}g_{ik}R=-8\pi GT_{ik}. \qquad (21)$$

Here and in (18), T_{ik} is the normal dynamical energy–momentum tensor for a set of particles,

$$T^{ik}(x)=\sum_{a,b,\ldots}\int m(A)\,\frac{\delta_4(x,A)}{\sqrt{-g(A)}}\,\frac{\mathrm{d}a^i}{\mathrm{d}a}\,\frac{\mathrm{d}a^k}{\mathrm{d}a}\,\mathrm{d}a. \qquad (22)$$

The lack of scale invariance in standard general relativity is now easy to understand. It is because the equations of standard general relativity are the form taken by the scale-invariant equation (18) when the scale is specially chosen by (19), i.e. by fixing the scale so that the mass of every particle is m_0. And provided the length unit is taken to be the Compton wavelength corresponding to m_0 then $\hbar=1$ and the relation (20) between m_0 and the gravitational constant requires m_0 to be the Planck mass. (For a more detailed discussion see Hoyle et al. 1995.)

When an older mathematical expression of a physical theory is replaced by a new form, in the present sense of the old being completely contained in the new, one hopes for the new to contain further possibilities that were not accessible from the old. It is so here, with the wave equation (5) susceptible to a generalization that does not destroy its scale-invariant property. This can be done through the non-linear addition of a cube term,

$$\Box\phi^a+\frac{1}{6}R\phi^{(a)}-\phi^{(a)3}=\int\frac{\delta_4(x,A)}{\sqrt{-g(A)}}\,\mathrm{d}a, \tag{23}$$

and similarly for the fields $\phi^{(b)},\ldots$ from particles $b,\ldots$. In the cosmological approximation permitted by the smoothed metric used in Section 1, this has the effect of introducing a cosmological constant *which turns out to have exactly the correct order of magnitude*.

Unlike the linear case, in which equation (5) for $\phi^a, \varphi^b,\ldots$ can be summed to give

$$\Box m+\frac{1}{6}Rm=\sum_a\int\frac{\delta_4(x,A)}{\sqrt{-g(A)}}\,\mathrm{d}a, \tag{24}$$

a cube term in m does not follow strictly from a summation of (13) for $a, b,\ldots$. Nevertheless, in a homogeneous, isotropic Universe we can write

$$\Box m+\frac{1}{6}Rm-\wedge m^3=\sum_a\int\frac{\delta_4(x,A)}{\sqrt{-g(A)}}\,\mathrm{d}a, \tag{25}$$

where we used the contribution of the cube terms together as a cube term in m. The cube terms, however, do not add like linear terms and their sum will have a fraction $\wedge$ multiplying m^3, where

$$\wedge=N^{-2}, \tag{26}$$

N being the effective number of particles contributing to the sum Σ_a. The latter can be considered to be determined by an Olbers-like cut-off, contributed by the portion of the Universe surrounding x to a redshift of the order of unity. In the observed Universe this total mass is of order $3\times10^{22}\,\mathrm{M}_\odot$, sufficient for about 6×10^{60} Planck masses. The actual particles are of course nucleons, of which there are about 3×10^{79}, derived in our view from physical interactions not included above. These interactions produce instability in which each Planck particle generates about 5×10^{18} nucleons. When 3×10^{79} nucleons are aggregated at 5×10^{18} per Planck particle, the required number of Planck particles is $N=6\times10^{60}$ and

$$\wedge\simeq2.8\times10^{-122}. \tag{27}$$

The reason why Planck particles are to be used here is that the basic wave equation (13) applies to Planck particles. If nucleons are used instead, factors are needed in the corresponding equation to take account of the subdivision of Planck particles into secondaries. When summations are taken leading to (25), the interpretation of A is seen to lead to the same result provided N in (26) refers to Planck particles. All of this is explained in detail in Hoyle et al. (1995), the modified form of (23) for nucleons being given there explicitly (equation 6.1, p. 206).

The next step is to notice that (16) would be obtained in usual field theory by applying stationary action with respect to variation of $m(x)$ to

$$-\frac{1}{2}\int\left(m_im^i-\frac{1}{6}Rm^2\right)\sqrt{-g}\,\mathrm{d}^4x-\frac{1}{4}\wedge\int m^4\sqrt{-g}\,\mathrm{d}^4x-\sum_a\int m(A)\,\mathrm{d}a. \tag{28}$$

Now in the special scale in which $m(x)$ is everywhere m_0, the scale of general relativity, (28) with G defined as in (20) can be written in exactly the same form as (14), but with λ given by

$$\lambda=3\wedge m_0^2. \tag{29}$$

Thus we obtain an explicit result for the magnitude of λ, a result that is beyond the range of general relativity. With 2.8×10^{-122} for $\wedge$ as in (27) and with the Planck mass m_0 about $3\times10^{32}\,\mathrm{cm}^{-1}$ ($\hbar=1$) the result (29) is given in numerical terms,

$$\lambda\simeq7\times10^{-57}\,\mathrm{cm}^{-2}, \tag{30}$$

the correct value of order H_0^{-2}.

6 DISCUSSION

This derivation of λ, agreeing well with the cosmologically required value of $\sim H_0^{-2}$, would appear at first sight to give an impetus to the present trend among cosmologists to introduce and believe in a non-zero value of λ. What has been shown is that λ is not a true constant of the theory, however, but a consequence of approximating the sum of wave equations of the form of (23) for particles $a, b, c,\ldots$ to the combined equation for $m(x)$ in (25). This is an approximation made possible by the large-scale smoothness of the spatial distribution of matter. The coefficient $\wedge=N^{-2}$ appearing in this process was critical in passing from (29) to the numerical estimate for λ in (30).

The value of N was calculated from the number of source particles of $m(x)$ lying within an Olbers-like cut-off distance, the number chosen for N being appropriate for the modern Universe. But would the same value of N apply to other epochs? The answer to this question turns out to depend on the cosmological model. The answer is *yes* for the steady state model, but it is *no* for big bang models, which lead to $\lambda\propto S^{-3}$, not λ constant.

The dynamical equation determining $S(t)$ in the standard model (with $\lambda\neq0$) is then

$$\frac{\dot S^2}{S^2}=\frac{8\pi}{3}G\rho+\frac{1}{3}\lambda=\frac{\alpha}{S^3}, \tag{31}$$

with

$$\alpha=S_0^3\left(\frac{8\pi}{3}G\rho_0+\frac{1}{3}\lambda_0\right). \tag{32}$$

The age of the Universe follows from (31) as $2/3H_0^{-1}$, and the inclusion of λ has made no difference to the usual result. This is because λ, when derived as above, varies with epoch in the same way as ρ.

The defence now for the standard model is to return to (14), with λ introduced ad hoc as a constant having a value independent of space–time position. The objections raised above then apply, namely that λ is required to have a numerical value that is explicitly related to the modern Universe. This objection is increased by the work of Section 5, in which we obtained a form of λ whose present numerical value was indeed related correctly to the modern Universe, but with the difficulty of it not being so at former times, with (31) and (32) reasserting the age difficulty set by the observed value of H_0 together with the ages of the globular clusters.

In the early 1980s, when inflationary cosmology was hailed with great fanfare, its one great merit was declared to be 'freedom from fine-tuning'. The model obtained required $\Omega = 1$, $\lambda = 0$ and argued that there must be enough matter in the Universe to make $\Omega = 1$, e.g. black holes, brown dwarfs, Jupiters, etc. However, this option was an embarrassment to the standard model for two reasons. First, it gave too large a value for $\Delta T/T$ for the microwave background – far higher than experimentally allowed. The other problem was that the primordial deuterium abundance would drop to a negligible value. So the 'non-baryonic' option was introduced.

Now, with the age problem (Section 2), the $\Omega = 1$ rule for matter has been modified and the cosmological constant has been introduced. There is no attempt to relate it to the original inflationary epoch: its present value is therefore fine-tuned, as is the value of the matter density. And the many models for structure formation invoke various kinds of non-baryonic options, cold dark matter, hot dark matter, mixed dark matter etc., which would do any culinary expert proud. Whether any of them will work appears so far to be unclear.

We understand the concern of cosmologists that unbridled speculation should not take over the field, that it is better to persist with the standard model, warts and all, than for opinions to become splintered, with the decline of professional standards which would then almost inevitably ensue.

Our response to this point of view, with which we have some sympathy, is that undesirable fragmentation has been permitted already, through the invasion of cosmology by particle physicists. If the invasion had the precision and the certainty of earlier invasions of astrophysics by atomic theory and by nuclear physics, the consequences would obviously be positive. However, one can have reservations about the advantages of becoming caught up in speculations from a different field, especially when those speculations are announced with an air of authority that will probably turn out to have been taken too seriously.

We agree with the majority of cosmologists, however, that it is probably better to stick with the standard model in the hope of something eventually turning up, than it is to rush off in some harum-scarum direction. Only if there is a clear understanding of what one is seeking to achieve is it reasonable to venture something new. Our position here is straightforward. We have explained the situation in detail in Section 5. With the rest of physics already possessing the property of scale invariance, we think it makes no sense to persist in cosmology with a gravitational theory that is not scale invariant. In discussions going outside the range of this paper, we also think it makes no sense to exclude the creation of matter from cosmology. There is ample evidence to show that observed samples of matter are not infinitely old. Excluding the creation of matter simply because current theory does not lend itself to a discussion of this question seems to us an invitation to trouble. It is much better in our view to face up to this problem than to attempt to get around current difficulties through the ad hoc introduction of λ, which would seem to achieve nothing beyond its own statement, apparently failing in its relation to the cube term in the wave equation (23).

REFERENCES

Baade W., 1952, Trans. IAU 8. Cambridge University Press, Cambridge, p. 387
Bagla J. S., Padmanabhan T., Narlikar J. V., 1995, preprint (BPN)
Branch D., Tammann G., 1992, ARA&A, 30, 359
Burbidge E. M., Burbidge G. R., Fowler W. A., Hoyle F., 1957, Rev. Mod. Phys., 29, 547
Chaboyer B., 1995, ApJ, 444, L9
Clayton D., 1964, ApJ, 139, 637
Courant R., Hilbert W., 1950, Methods of Mathematical Physics, Vol. II. Interscience, New York, London, p. 727
Cowan J., Thielemann F.-K., Truran J., 1987, ApJ, 323, 543
Cowan J., Thielemann F.-K., Truran J., 1991, ARA&A, 29, 447
DeMarque P., McClure R., 1977, in Tinsley B., Larson R. B., eds, Evolution of Galaxies and Stellar Populations. Yale Univ. Obs., New Haven, p. 199
de Vaucouleurs G., 1961, ApJS, 6, 231
de Vaucouleurs G., 1982, Lecture Notes. Australian National University, Canberra
Efstathiou G., Bond J. R., White S., 1992, MNRAS, 258, 1P
Eggen O. J., Lynden-Bell D., Sandage A. R., 1962, ApJ, 136, 748
Einstein A., 1917, Preuss Akad. Wiss. Berlin Sitzber, 142
Fowler W. A., 1987, QJRAS, 28, 87
Fowler W. A., Hoyle F., 1960, Ann. Phys., 10, 280
Freedman W. et al., 1994, Nat, 371, 757
Gunn J. E., Tinsley B. M., 1975, Nat, 257, 454
Hamuy M., Phillips M., Maza J., Suntzeff N., Schommer R., Aviles R., 1995, AJ, 109, 1
Hoyle F., Narlikar J. V., 1966, Proc. R. Soc. London, Ser. A, 294, 138
Hoyle F., Burbidge G., Narlikar J. V., 1993, ApJ, 410, 437
Hoyle F., Burbidge G., Narlikar J. V., 1994a, MNRAS, 267, 1007
Hoyle F., Burbidge G., Narlikar J. V., 1994b, A&A, 289, 729
Hoyle F., Burbidge G., Narlikar J. V., 1995, Proc. R. Soc. London, Ser. A, 448, 191
Hubble E., 1936, The Realm of the Nebulae. Oxford University Press, Oxford
Humason M. L., Mayall N. U., Sandage A. R., 1956, AJ, 61, 197
Jacoby G. et al., 1992, PASP, 104, 599
Kennicutt R., Freedman W., Mould J., 1995, AJ, 110, 1476
Liddle A. R., Lyth D. M., Roberts D., Viane P. T., 1995, preprint
Lundmark K., 1946, Obs. Medd. Lund, Vol. I, Nr 163
Lundmark K., 1948, Obs. Medd. Lund, Vol. I, Nr 187
Mineur H., 1952, C.R.Sci, 235, 1607
Reiss A. G., Press W. H., Kirshner R., 1995, ApJ, 438, L17
Rutherford E., 1929, Nat, 123, 313
Sachs R., Narlikar J. V., Hoyle F., 1996, A&A, in press
Sandage A. R., 1958, ApJ, 127, 513
Sandage A. R., 1962, in Proc. IAU Symp. 15, Problems of Extra-galactic Research. Kluwer, Dordrecht, p. 359
Sandage A. R., 1993, AJ, 106, 719

Sandage A. R., 1995, in Bingelli B., Buser R., eds, The Deep Universe (23rd Sass Fee Course). Springer, Berlin, p. 188
Sandage A. R., Tammann G., 1974a, ApJ, 190, 525
Sandage A. R., Tammann G., 1974b, ApJ, 194, 559
Sandage A. R., Tammann G., 1990, ApJ, 365, 1
Sandage A. R., Tammann G., 1993, ApJ, 415, 1
Sandage A. R., Saha A., Tammann G. A., Labhardt L., Panagia N., Macchetto F., 1996, ApJ, submitted
van den Bergh S., 1992, PASP, 104, 861
van den Berg S., Stetson H. T., Bolte M., 1996, ARA&A, in press
Weinberg S., 1989, Rev. Mod. Phys., 66, 1
White S. D. M., Efstathiou G., Frenk C. S., 1993, MNRAS, 262, 1023