Q. J. R. astr. Soc. (1993), **34**, 157–199

Modern Cosmology – A Critical Assessment*

Malcolm S.Longair
Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE

(Received 1992 December 16)

1 PRELIMINARIES

When quality newspapers such as *The Independent* devote the whole of a front-page spread to the detection of fluctuations in the Cosmic Microwave Background Radiation, we can be sure that cosmology has come of age. The story of the last 30 years has been one of quite remarkable progress in cosmology and I am one of that very lucky generation who began to carry out research in this area just as the flood was about to break. Progress has been spectacular and has opened up completely new ways of addressing problems of cosmology and fundamental physics. In their enthusiasm, however, I believe that some cosmologists have been carried away and a number of exaggerated claims have been made about how much we really understand about topics related to the origin of our Universe. My objective this evening is to attempt to redress the balance and to address four questions simultaneously:

(1) How much can we really trust?
(2) What will be the lasting contributions?
(3) What can we hope to achieve?
(4) What should everyone know?

My approach will be not to exaggerate but to take a long hard look at those bits of the story which we can really believe. In my view, the story is quite remarkable enough without speculation beyond what has already been established with a good deal of confidence. To pose the question another way, which bits of the story are so convincing that we would be prepared to die in the ditch defending them?

Before setting out on our journey, let me quote the words of two of the trustees of the Milne Society. First, I quote Sir Hermann Bondi in a recent article about the impact of Newton's achievements in deriving his laws of motion and of gravity (Bondi 1989):

> He was given a very jumbled-up heap of data, partly ordered...by Kepler's and Galileo's work. This enormous heap of data, in which the soluble and insoluble items seem inextricably mixed, Newton cut neatly into two with surgical precision. His genius separated out the question:
>
> > 'Given the positions, velocities and masses of the bodies of the Solar System today, where will they be in the future and where have they been in the past?'

---

* The text of the 16th Milne Lecture, delivered 1992 November 26.

> He solved this problem so totally that not much was left for others. The other side of the cut concerns the question:
>
> > 'Why do the bodies of the Solar System have these particular masses, positions and velocities?'
>
> This is the problem of the origin of the Solar System. After 300 years and an enormous amount of work, we are not all that wiser as regards these issues.

As someone who has devoted 10 years to helping provide the UK astronomical community with the observing facilities to tackle exactly these problems, I would have liked to disagree with this last remark but I fear that it still remains largely true. We have made enormous progress in understanding the problems of star and planet formation but, as is common in science, to make progress we change the question so that we can answer at least particular aspects of the problem, if not the ones which started the quest.

Let me quote just a little more of this passage.

> (Newton's) solution of the problem of motion in the Solar System was so complete, so total, so stunning, that it was taken as the model of what any decent theory should be like, not just in physics, but in all fields of human endeavour...
>
> I regard this as profoundly misleading. In my view, most of science is not like the Solar System but much more like weather-forecasting.

If the Solar System is such a hard nut to crack, what hope have we for the Universe as a whole? What if the origin of our Universe is more like weather-forecasting than Newton's Law of Gravity? There is no *a priori* reason why the processes involved in the origin and evolution of the Universe should be as simple as Newton's law of gravity. We can only find out by making the right observations and interpreting them in as model-independent a fashion as possible.

My second set of quotations is from Sir William McCrea who wrote what I have always considered to be a very salutary paper for cosmologists in 1970 entitled 'A Philosophy for Big-Bang Cosmology' (McCrea 1970). We will show that the Universe must have passed through a very hot dense phase in the distant past and so McCrea asks how much we can hope to learn about the very early stages of the Universe from the observations we make now. As he points out, we have a fundamental problem at the outset. We have only one Universe to study and that distinguishes the scientific study of cosmology from all other sciences. In physics, critical experiments can be carried out by independent workers with completely different apparatus and it is the agreement and repeatability of these experiments which gives us confidence in the results of the experiment and their implications for theory. In the case of the Universe, we have only one example and we cannot even do experiments with it. All we can do is observe it. We can, to some extent, carry out independent experiments by making similar observations in different regions of the Universe and, if the same results are found wherever we look, we can suspect that we have found a general rule. It is, however, applicable only for the region of the Universe we can observe and there might be some aspects which are only observable on the very largest scale for which there would be no possibility of making an independent observation.

Of the six propositions in McCrea's paper, I will quote only two which I consider to be of special significance for current cosmological research.

> *Proposition* (*B*) The less information we can get, the less we need in order to make predictions that are confirmed by observation.
> *Proposition* (*C*) From the observed properties of the present state of the Universe, we can infer less and less about earlier and earlier previous states, and almost nothing about what we might wish to call the initial state.

I believe it is worthwhile pondering the full significance of these remarks, especially in the light of some of the claims which have been made about our understanding of the early stages of our Universe. If these earliest phases are more and more inaccessible to observational study, there is considerable freedom in the choice of conceivable physical theories. I therefore wish to address the question of how valid these views are, in particular, Proposition (C), in the light of the remarkable recent discoveries of observational cosmology.

Before embarking upon this study, let me note what I consider to be the four developments which have made the advances possible.

(1) In my view, by far the most important has been the opening up of the whole of the electromagnetic spectrum for astronomical observations. Since 1945, the disciplines of radio, millimetre, infrared, ultraviolet, X-ray and γ-ray astronomy have all become major astronomical disciplines and each has had its own unique contribution to make to filling out the cosmological picture. Some of these astronomies can only be carried out from above the Earth's atmosphere – far-infrared, ultraviolet, X-ray and γ-ray astronomy – and so space and ground-based observations are complementary in the information they provide about our Universe (for more details, see Longair (1989)).

(2) Going hand-in-hand with the new astronomies has been the development of new technology. The computing and micro-electronic revolution as well as the astounding developments in instrument and detector technology mean that the telescopes for essentially all the electromagnetic wavebands are approaching their ideal efficiencies.

(3) The astrophysicists and cosmologists have rapidly absorbed all the great discoveries of modern physics into the battery of tools which they use to study the cosmos to great advantage.

(4) Finally, and by no means least, the astronomical discoveries have led to completely new astrophysical disciplines which have provided new tools for studying key astrophysical problems of cosmological importance.

To my regret, I will make little further allusion to instruments and telescopes but they are the prime source of the great discoveries and new understandings we will discuss. However ephemeral the theories, the observational and instrumental achievements are outstanding lasting contributions.

2 OBSERVATIONAL COSMOLOGY IN 1963

Why are observational and astrophysical cosmology feasible at all? When we look at the Universe, it is of quite daunting complexity. Within our own Galaxy, we observe the complexities of the birth, life and death cycle of stars – on the scale of galaxies, we observe configurations of stars and gas which range from the completely regular to the totally pathological – on the scales of whole collections of galaxies we observe regular clusters of galaxies as well as 'stringy' structures and huge 'holes' which seem to contain significantly fewer galaxies than the average. Despite the complexity, it turns out that the Universe as a whole has some very simple large-scale features and it is these which make the subject of astrophysical cosmology possible as a science.

When I began research in radio astronomy as a research student in 1963, my supervisor Dr Peter Scheuer gave me a copy of Sir Hermann Bondi's classic text *Cosmology* to absorb and warned me that

> There are only $2\frac{1}{2}$ facts in cosmology.

The point is a very important one in that, of the mass of observations which can be made of gas, stars and galaxies, most of them tell us nothing of real cosmological significance. We have to select from this plethora of data those pieces which establish real facts about the Universe as a whole. We now know many more real facts about the Universe but they are still a small finite number. My personal view is that the uncertainties are greater than many professionals would wish to believe. I will therefore start with these $2\frac{1}{2}$ facts and see how much more we have learned over the last 30 years. My aim is to distinguish between four features of these studies – Fact, Theory, Assumption and Pure Speculation.

In 1963, the $2\frac{1}{2}$ facts were as follows:

Fact 1. The sky is dark at night

This is the well-known observation which leads to what is known as *Olbers' paradox* although the paradox was well known to earlier cosmologists. Sir Hermann in his text *Cosmology* gives a thought-provoking discussion of the meaning of the paradox (Bondi 1952). The fact that the sky is not as bright as the surface of the Sun provides us with some very general information about the Universe. Probably the most general way of expressing the significance of this observation is that the Universe must, in some sense, be far from equilibrium although in what way it is in disequilibrium cannot be deduced from this very simple observation.

Fact 2. The galaxies are receding from each other as expected in a uniform expansion

This was Hubble's great discovery of 1929 and I will say much more about it in a moment. The $2\frac{1}{2}$th fact was as follows:

Fact $2\frac{1}{2}$. The contents of the Universe have probably changed as the Universe grows older

The reason for the ambiguous status of this fact was that the evidence for the evolution of extragalactic radio sources as the Universe grows older was then a matter of considerable controversy, particularly with the proponents of Steady-State cosmology. I was plunged straight into this debate as soon as I began my research programme with Martin Ryle and Peter Scheuer. As we will see, this is no longer a controversial issue – there is no question at all

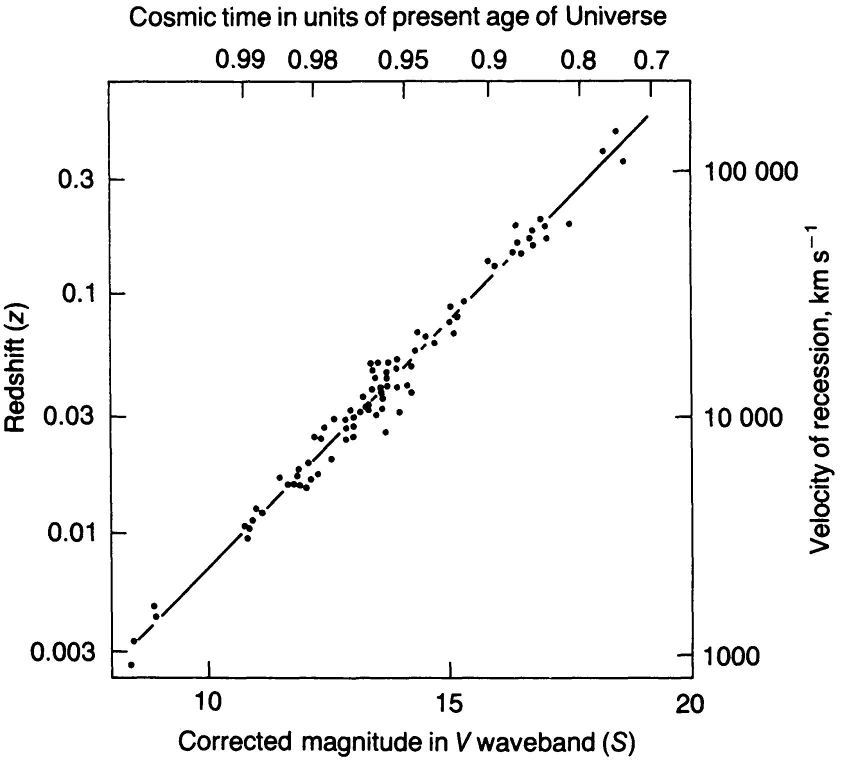


FIG. 1. A modern version of the velocity–distance relation (or Hubble diagram) for the brightest galaxies in clusters (after Sandage 1968). In this logarithmic plot, the corrected apparent magnitudes (that is $-2{\cdot}5$ times the logarithm of the flux density S) of the brightest galaxies in clusters are plotted against their redshifts z. The straight line shows what would be expected if $S \propto z^{-2}$. This correlation indicates that the brightest galaxies in clusters have remarkably standard properties and that the distances of the galaxies are proportional to their redshifts which, for small redshifts, implies that velocity is proportional to distance.

but that many classes of object exhibit evolutionary changes as the Universe grows older. Thus, in 1963, the number of real facts which characterized the Universe as a whole was very small and relatively modest progress had been made since the 1930s and the time of Milne.

To the chagrin of the observers, the standard world models which we use in our everyday work were discovered when only the first of these facts was known. Einstein completed his General Theory of Relativity in 1915 and quickly realized that he had a tool which could be used to construct meaningful models of the Universe as a whole, unlike the Newtonian Theory of Gravity which does not take account of the fact that the speed of light is finite and for which satisfactory boundary conditions at infinity cannot be found. Einstein's static model of the Universe was published in 1917 but the real breakthrough came with the work of the Russian meteorologist and theoretical physicist Aleksander Aleksandreyevich Friedman who, in the period 1922–1924, solved Einstein's field equations of General Relativity for the complete set of homogeneous, uniformly expanding models of the Universe. At that time, this was a theoretical exercise and the deep significance of his results were only appreciated when Friedman's papers were later publicized by Georges Lemaître. Friedman died of typhoid in 1925

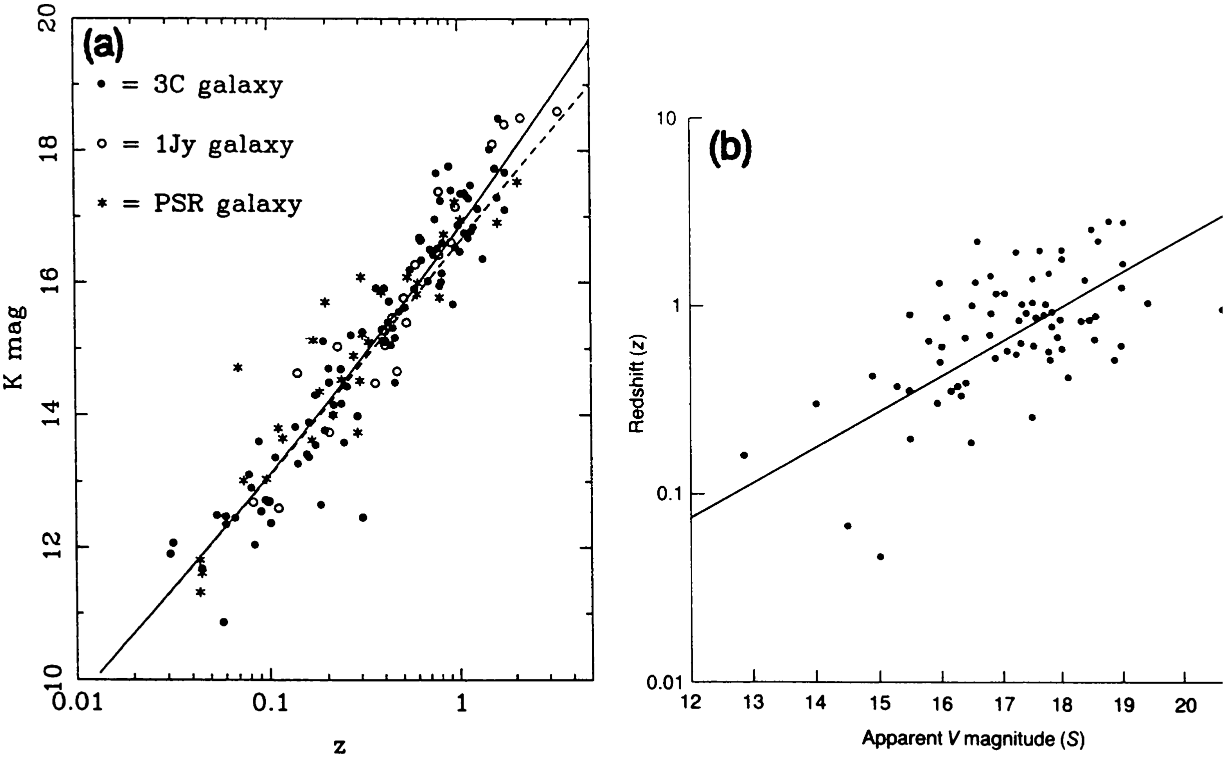


FIG. 2. (a) The $K-z$ relation for radio galaxies. This compilation includes radio galaxies from the 3CR sample (●), from the 1Jy sample (○) and the 0·1 Jy sample (asterisks) (Dunlop & Peacock 1990). (b) The redshift–magnitude relation for a complete sample of radio quasars. There is a positive correlation in the same sense as the redshift–magnitude relation for the brightest galaxies in clusters (Fig. 1) but there is much greater dispersion about the mean line. The solid line is arbitrarily drawn through the points with slope corresponding to $S \propto z^{-2}$ (Longair 1989, after Wall & Peacock 1985).

during the years of the civil war in Leningrad and so did not live to see the observational validation of his solutions of the field equations.

In 1929, Hubble announced his discovery of the relation between the velocity of recession of the galaxies and their distances. It had been known for some years that, when the velocities of galaxies are measured, they are all observed to be receding from our own Galaxy. Hubble's great discovery was that the further away a galaxy is from our own Galaxy, the greater its velocity of recession. Figure 1 shows a modern version of what is now known as the Hubble diagram and is due to Sandage (1968) who used the brightest galaxies in clusters to define the relation.

This relation is known as *Hubble's Law* and is written $v = H_0 r$ where the constant H_0 is known as Hubble's constant. All classes of extragalactic object obey this law. To illustrate this, I show in Fig. 2 the same diagrams for radio galaxies and for radio-selected quasars. In Fig. 2(a), it can be seen that the velocity–distance relation for this particular sample of radio galaxies extends to very large redshifts, $z > 2$, and that the dispersion about the mean line remains very small. In Fig. 2(b), I show the same diagram but now for radio-selected quasars. It is well-known that there is a greater dispersion in the intrinsic properties of the quasars than of luminous galaxies but nonetheless

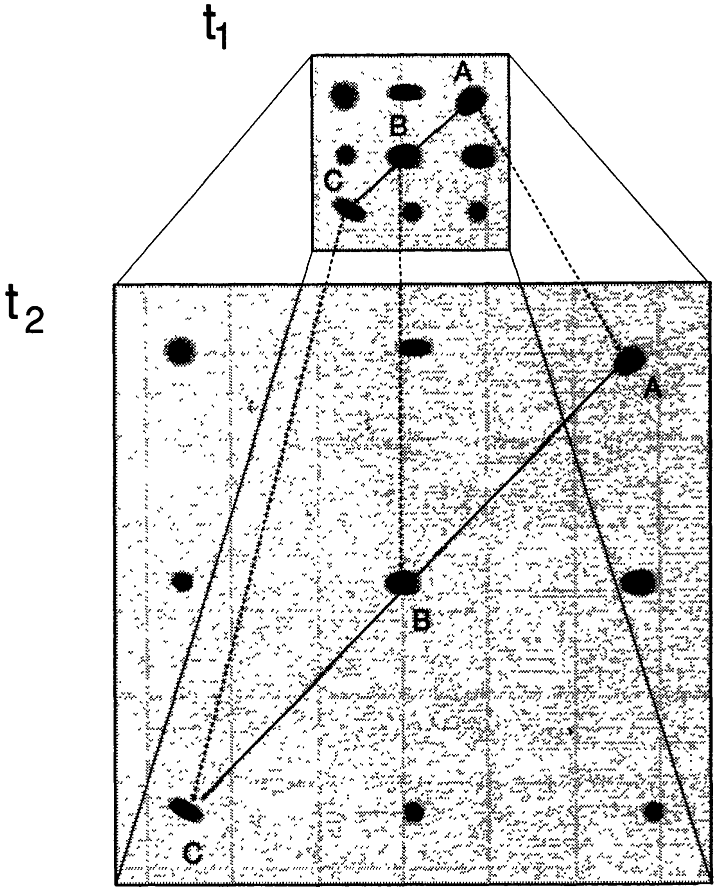


FIG. 3. Illustrating the origin of the velocity–distance relation for an isotropically expanding distribution of galaxies. The distribution of galaxies expands uniformly between the epochs t_1 and t_2. If, for example, we consider the motions of the galaxies relative to the galaxy A, it can be seen that galaxy C travels twice as far as galaxy B between the epochs t_1 and t_2 and so has twice the recession velocity of galaxy B relative to A. Since C is always twice the distance of B from A, it can be seen that the velocity–distance relation is a general property of isotropically expanding Universes.

the mean line runs through the centre of the distribution of points showing that there exists a mean redshift–magnitude relation for these objects.

Milne realized that the observation of a velocity–distance relation meant that the objects observed must be participating in a uniform expansion. This is illustrated in Fig. 3 which shows a uniform distribution of galaxies participating in a uniform expansion. A uniform expansion means that the distance between neighbouring galaxies increases by the same factor in a given time interval. The result is that, if we fix our attention upon any one galaxy in the distribution, the further it is away from the chosen point, the greater the distance it has to travel in the same time in order to preserve the uniform expansion. Thus, the observation of the velocity–distance relation for all extragalactic systems simply means that the distribution of galaxies is participating in a uniform expansion in which the distances between neighbouring galaxies continually increase with time.

In the 1930s, Milne & McCrea (1934) explained the physical content of Friedman's solutions of Einstein's field equations in terms of a simple

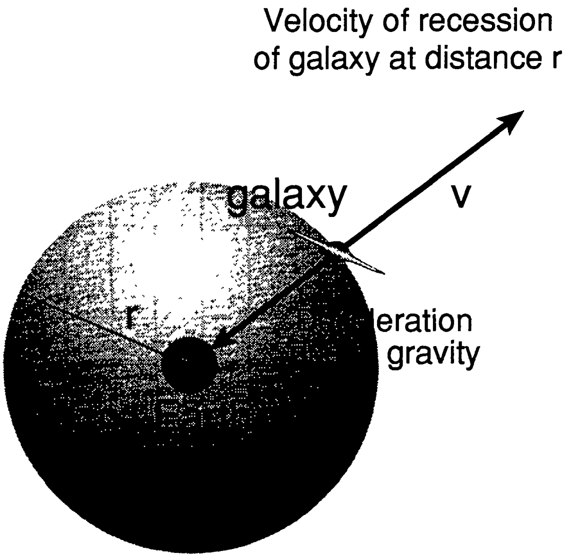


FIG. 4. Illustrating the Newtonian model for the dynamics of the classical Friedman models of General Relativity. The velocity of recession of a galaxy at distance *r* can be considered to be decelerated by the gravitational attraction of the matter within distance *r* of our own Galaxy. Because of the assumption of isotropy, an observer on any galaxy participating in the expansion of the Universe would carry out exactly the same calculation.

Newtonian picture. They realized that, in completely homogeneous, isotropic universes, global physics must be the same as local physics since every point in the Universe has to be equivalent to every other point at a given time and so the same physics applicable locally must be applicable at all points in the Universe. They showed how to derive the essential content of the Friedman models in terms of the simple Newtonian picture illustrated in Fig. 4. We observe the same velocity–distance relation in whichever direction we look and so we ask what would be the deceleration of a galaxy at distance *r* from us due to all the mass within the sphere of radius *r*. Because of the spherical symmetry of the problem, Gauss's theorem tells us that we obtain the correct expression for the deceleration of the galaxy by placing all the mass within radius *r* at the location of our Galaxy. Thus, the dynamics of the galaxy depends upon how much mass there is within radius *r* and hence upon the average density of matter within that sphere.

The classical Friedman models describe the expansion of the Universe in terms of a *scale-factor* which describes how the separation between points partaking in the universal expansion changes with time. There are three types of solution. If the Universe is of high density, the force of gravitational attraction is sufficiently great to halt the expansion and the Universe eventually collapses back to a high density, high temperature state, a state often referred to as the 'Big-Crunch'. If the Universe is of low density, the force of gravitational attraction is not sufficient to halt the universal expansion and, in the limit of infinite time, the expansion velocity remains finite. Separating these models, there is a unique model known as the *critical* or *Einstein–de Sitter* model which has zero velocity of expansion as the time

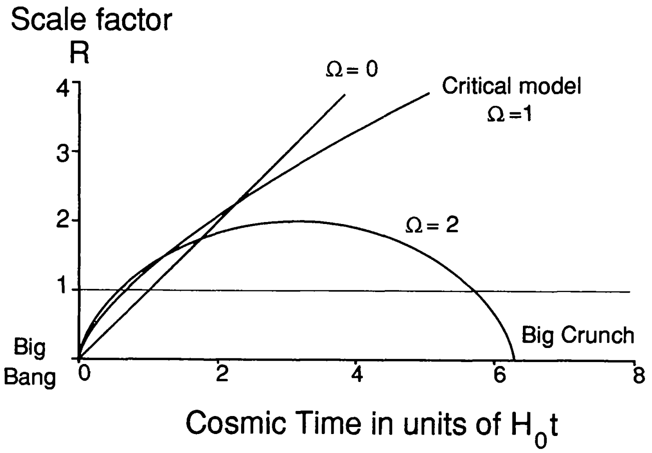


FIG. 5. The dynamics of the classical Friedman models parametrized by the density parameter $\Omega = \rho/\rho_{\text{crit}}$. If $\Omega > 1$, the Universe collapses to $R = 0$ as shown; if $\Omega < 1$, the Universe expands to infinity and has a finite velocity of expansion as R tends to infinity. In the case $\Omega = 1$, $R = (t/t_0)^{\frac{2}{3}}$ where $t_0 = (2/3)H_0^{-1}$. The time axis is given in terms of the dimensionless time $H_0 t$. At the present epoch $R = 1$ and in this presentation, the three curves have the same slope of 1 at $R = 1$, corresponding to a fixed value of Hubble's constant. If t_0 is the present age of the Universe corresponding to $R = 1$, then for $\Omega = 0$, $H_0 t_0 = 1$; for $\Omega = 1$, $H_0 t_0 = 2/3$; and for $\Omega = 2$, $H_0 t_0 = 0{\cdot}57$.

tends to infinity. In other words, the Universe just possesses its own escape velocity. There is a corresponding *critical density* ρ_{crit} associated with this model which depends only upon the value of Hubble's constant, $\rho_{\text{crit}} = 3H_0^2/8\pi G$. It is often convenient to compare the mass densities of world models (or any constituent of the models) with the critical density ρ_{crit} so that a *density parameter* $\Omega = \rho/\rho_{\text{crit}} = 8\pi G\rho/3H_0^2$ can be used to parametrize the world models. These solutions are illustrated in Fig. 5.

What cannot be incorporated into the Newtonian arguments developed by Milne & McCrea is the dependence of the global geometry of space upon the density except in some special cases. According to General Relativity, the geometry of space-time is determined by the mass-energy distribution throughout the Universe. In the homogeneous Friedman models, the density distribution is the same at all points in space at a given time and so the spatial curvature is the same everywhere at that epoch. Formally, the curvature of the geometry is given by $\kappa = 1/\mathcal{R}^2 = (\Omega - 1)/(c/H_0)^2$, where $\mathcal{R}$ is the radius of curvature of the spatial sections. In the case of the high density models, the geometry is closed and spherical while in the low density models, it is open and hyperbolic. The case of the completely empty Universe was analysed by Milne and he showed that its spatial geometry is hyperbolic (see Longair 1992 for an elementary derivation of this result). Appropriately, the empty model having $\rho = 0$ is known as the *Milne Model*. It turns out that only in the case of the critical model is the geometry flat Euclidean space. As we will see, the unique features of the critical model have a certain theoretical appeal. In principle, the geometry of space is a measurable quantity – one simply needs to measure accurately the sum of the angles of a triangle over significant cosmological distances and find out whether the sum is equal to 180° or not. In practice, this test is not feasible.

3 THE FRIEDMAN MODELS IN 1992

It will be noted that the standard Friedman models use only two pieces of observational evidence, the velocity–distance relation for galaxies and the isotropy of the Universe as a whole on the large scale. The models also assume that, on the large scale, the dynamics of the Universe can be described by General Relativity. There is also an implicit assumption made which is known as the *Cosmological Principle* according to which the large-scale features of the Universe which we observe would also be observed by any other suitably chosen observer who looks at the Universe at the same cosmic epoch. In other words, an astronomer on a distant galaxy would also observe the same Hubble's law and an isotropic Universe if the observations are made at the present epoch. As we have argued above, the interpretation of the velocity–distance relation in terms of the uniform expansion of a homogeneous distribution of galaxies is entirely consistent with this principle. It can also be subjected to direct test by comparing the properties of the distribution of galaxies at different distances and in different regions of space. So far as we can tell, there is no reason to believe that the assumption that we live at a typical point in the Universe is incorrect. Indeed, if the principle were not to be correct and we are at a very special point in the Universe, we would be forced to return to a pre-Copernican, Ptolemaic view of the Universe in which we would occupy a privileged position in the Universe. There is no observational evidence for this to be the case.

We have already discussed the modern versions of Hubble's law but we have yet to deal with the isotropy of the Universe and with the assumption that General Relativity can be used to describe the large scale dynamics of the Universe. I will elevate the first of these observations to Fact 3 because of the quite spectacular accuracy with which this has now been established.

Fact 3. The Universe is isotropic on very large scales to an accuracy of better than one part in 100000

The Universe is obviously highly inhomogeneous on a small scale with matter condensed into stars which are congregated into galaxies which are themselves clustered, the associations ranging from small groups to giant regular clusters of galaxies. If we take our averages on larger and larger scales, however, the inhomogeneity becomes less and less. Fig. 6 shows the distribution of galaxies in the northern galactic hemisphere once all the stars of our own Galaxy have been removed. The large 'bite' out of the picture in the bottom right corresponds to an area of the sky which was not observed in the Lick survey and the decrease in the numbers of galaxies towards the edges of the picture is due to extinction by interstellar dust in our own Galaxy. Therefore, only in the central region of Fig. 6 do we obtain a reasonably clean picture of the large scale distribution of galaxies in the Universe. This is a picture of the distribution of galaxies on the grandest scale, a giant cluster such as the Coma cluster corresponding to the bright dot in the centre of the picture. We now know that much of the obvious

Fig. 6. The distribution of galaxies in the northern galactic hemisphere derived from counts of galaxies undertaken by Shane, Wirtanen and their colleagues at the Lick Observatory in the 1960s. Over one million galaxies were counted in their survey. The northern galactic pole is at the centre of the picture and the galactic equator is represented by the solid circle bounding the diagram. The projection of the sky onto the plane of the picture is an equal area projection. This photographic representation of the galaxy counts was made by Peebles and his colleagues. The large sector missing from the lower right-hand corner of the picture corresponds to an area in the southern celestial hemisphere which was not surveyed by the Lick workers. The decreasing surface density of galaxies towards the circumference of the picture, that is towards the galactic equator, is due to the obscuring effect of interstellar dust in the interstellar medium of our own Galaxy. The prominent cluster of galaxies close to the centre of the picture is the Coma cluster (Seldner, Siebars, Groth & Peebles 1977).

clumping, the holes and the stringy structures are real features of the distribution of galaxies but, if we take averages over very large regions, one bit of Universe looks very much like another.

Even more impressive evidence comes from the distribution of extragalactic radio sources over the sky. It turns out that, when we make a survey of the radio sky, the objects which are easiest to observe are extragalactic radio sources associated with certain rare classes of galaxy at very great distances. Because they are rare objects, they sample the isotropy of the Universe on a large scale. Fig. 7 shows the distribution of the brightest

3000 extragalactic radio sources in the northern hemisphere. There is a hole in the centre of the distribution corresponding to a region which was not observed as part of the 4C survey but otherwise the distribution is entirely consistent with the sources being distributed uniformly at random over the sky. The radio sources are ideal for probing the large scale distribution of discrete objects since they are so readily observed at large distances. Similar maps will soon be available at X-ray wavelengths thanks to the ROSAT survey of the X-ray sky.

This is impressive enough, but it pales into insignificance compared with the recent results on the isotropy of the Cosmic Microwave Background Radiation. This background radiation was discovered in 1965 by Penzias & Wilson whilst commissioning a very sensitive telescope and receiver system for centimetre wavelengths. It was quickly established that this background radiation is remarkably uniform over the sky. The most recent results on the large scale distribution of this radiation over the sky have been obtained by the Cosmic Background Explorer (COBE) which was launched in November 1989. This satellite is dedicated to studies of the background radiation, not only in the millimetre waveband but throughout the infrared waveband as well.

The results in increasing levels of sensitivity are as follows. At sensitivity levels about one part in 1000 of the total intensity, there is a large scale anisotropy over the whole sky associated with the motion of the Earth through the frame of reference in which the radiation would be the same in all directions. This is no more than the result of Doppler effect due to the Earth's motion and as a result the radiation is about one part in a thousand more intense in one direction and exactly the same amount less intense in the opposite direction. The intensity distribution has precisely the expected dipolar distribution and it turns out that the Earth is moving at about 350 km s^{-1} with respect to the frame of reference in which the radiation would be 100% isotropic. At about the same level of intensity the plane of our Galaxy can be observed as a faint band of emission over the sky.

In the most recent analyses of the isotropy of the Cosmic Microwave Background Radiation on angular scales 10° and greater by the COBE workers, sensitivity levels of only one part in 100000 of the total intensity have been attained. At this level, the radiation from the plane of the Galaxy is intense but is confined to a relatively narrow strip centred on the galactic plane. Away from this region, the sky appears quite smooth on a large scale but careful analysis of the variation of the intensity from point to point on the sky has found convincing evidence for tiny fluctuations in intensity over and above the instrumental noise. The signal amounts to only about 1 part in 100000 of the total intensity and, when averaged over the clear region of sky, the significance of the result is at the 6σ level (Fig. 8). This is a very important result for cosmology as we will see later.

For the moment, however, our interest is in the isotropy of the Universe as a whole and we can state that there is certainly no evidence for any anisotropy in the distribution of the Cosmic Microwave Background Radiation at the level of one part in 100000 when we look on the large scale. This is quite incredible precision for any cosmological experiment since one is normally lucky in cosmology if one knows anything within a factor of

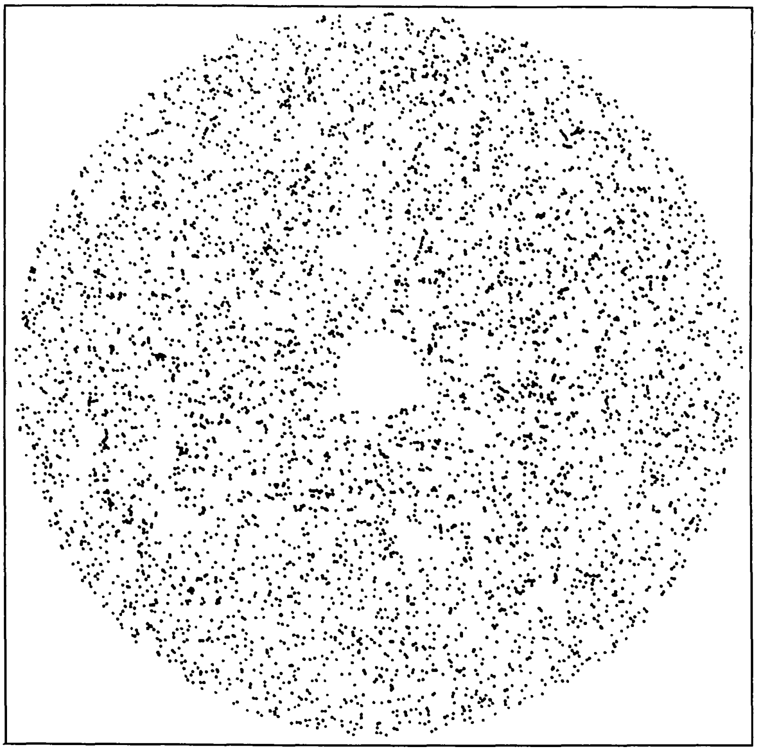


FIG. 7. The distribution of radio sources in the fourth Cambridge (4C) catalogue in the northern celestial hemisphere. This part of the catalogue contains over 3000 sources of small angular diameter. In this equal area projection, the north celestial pole is in the centre of the diagram and the celestial equator around the perimeter. The area about the north celestial pole was not surveyed. The distribution does not display any significant departure from a random distribution. (Courtesy of Dr M.Seldner.)

about 10. The obvious question is how the distribution of this radiation is related to the distribution of ordinary matter. The answer is not as straightforward as one would like and we need to understand the temperature history of the Universe to give the answer. In the standard picture of the evolution of the Hot Big Bang, when the Universe was squashed to only about one thousandth of its present size, the temperature of the Cosmic Background Radiation must have been about one thousand times greater than it is now and that was sufficiently hot for all the hydrogen in the Universe to be ionized. When this occurs, there is very strong coupling between the Cosmic Background Radiation and the ionized matter. In fact, when we look back to these epochs, it is as if we were looking at the surface of a star surrounding us in all directions but the temperature of the radiation we observe has been cooled by this factor of 1000 so that what we observe is redshifted into the millimetre waveband. This analogy also makes it clear that, because of the strong scattering of the radiation, we can only observe the very surface layers of the star and, in the same way, we can obtain no direct information about what was happening at earlier epochs as soon as we encounter the epoch at which the material of the Universe was ionized. This 'surface' at which the Universe becomes opaque to radiation is known as the *last scattering surface* and the fluctuations observed by COBE are believed to represent the very low intensity ripples present on that surface on angular

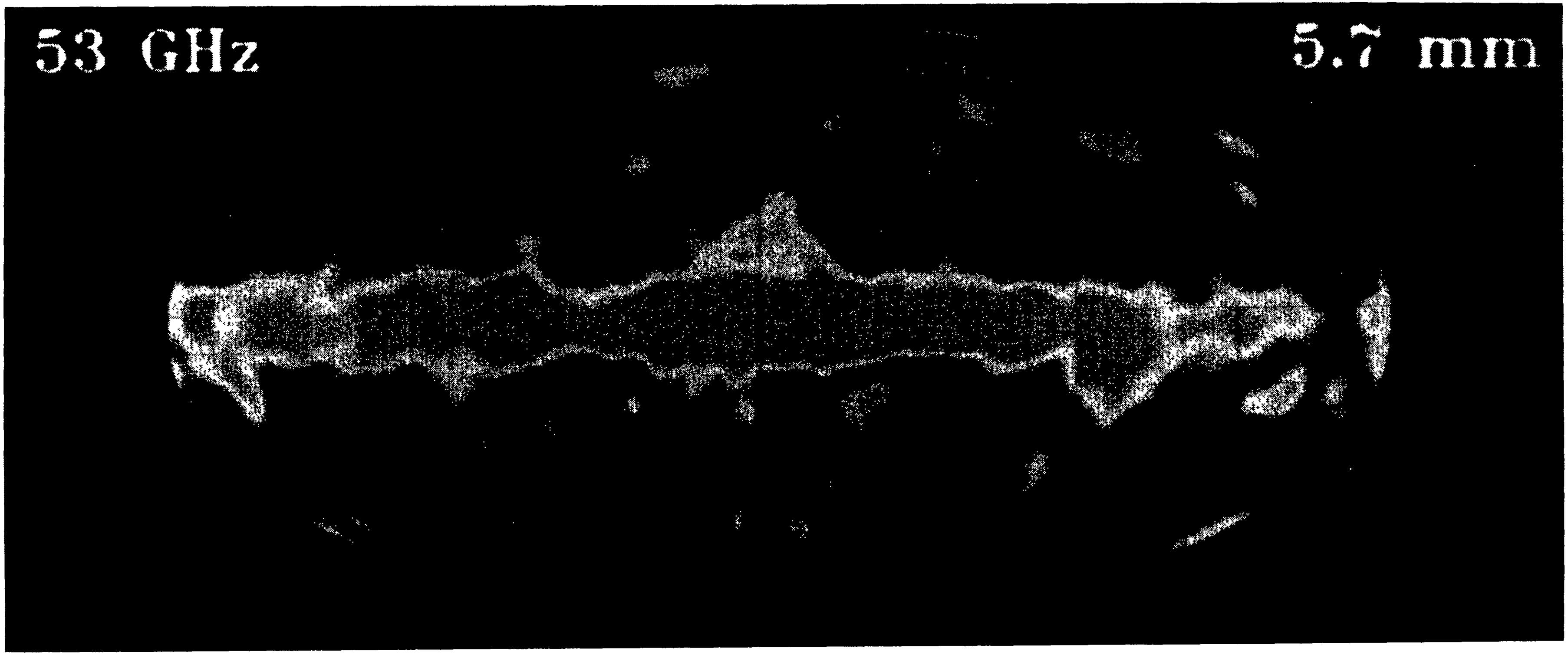


FIG. 8. The map of the whole sky in galactic coordinates as observed in the millimetre waveband at a wavelength of 5·7 mm by the COBE satellite once the dipole component associated with the motion of the Earth through the background radiation has been removed. The residual radiation from the plane of the Galaxy can be seen as a bright band across the centre of the picture. The fluctuations seen at high galactic latitudes are noise from the telescope and the instruments, the rms value at each point being 36 μK but, when statistically averaged over the whole sky at high latitudes, an excess sky noise signal of 30 ± 5 μK is observed (Smoot *et al.* 1992).

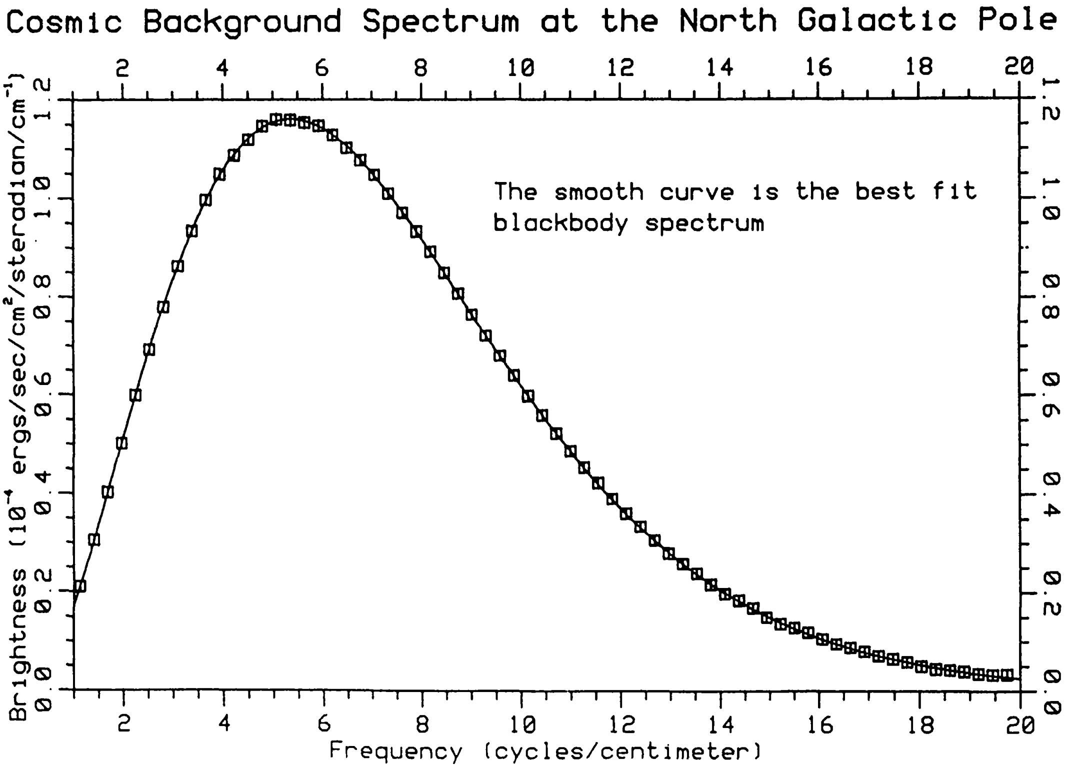


FIG. 9. The spectrum of the Cosmic Microwave Background Radiation as measured by the COBE satellite in the direction of the North Galactic Pole. Within the quoted errors, the spectrum is that of a perfect black body at radiation temperature 2·735 ± 0·06 K (Mather *et al.* 1990).

scales of 10°. Thus, strictly speaking, in the standard interpretation, the COBE results provide information about the diffuse ionized intergalactic gas when the Universe was only about one thousandth of its present size. At that stage, the galaxies had not formed and so all the ordinary matter which was eventually to become galaxies as we know them was still in the form of a remarkably smooth intergalactic ionized gas. The extragalactic radio sources provide complementary information about the large scale distribution of discrete objects such as galaxies at the present epoch once the galaxies had formed.

The upshot of this discussion is that the Cosmic Background Radiation provides us with information about the isotropy of the matter content of the Universe, not as it is now but as it was when the Universe was squashed by a factor of about 1000. Nonetheless, this is the most powerful evidence we possess that the Universe is quite remarkably isotropic on a large scale.

The COBE mission has produced another remarkable result which we can add to our list of real facts about the Universe.

Fact 4. The spectrum of the Cosmic Microwave Background Radiation has a pure black-body spectrum at a radiation temperature of 2·735 K

The evidence for this is the spectrum of the background radiation obtained by the Michelson interferometer on board the COBE satellite. The first published spectrum of the radiation is shown in Fig. 9. The boxes show the experimentally determined spectrum and the solid line is a black-body spectrum at a radiation temperature of 2·735 K. It can be seen that the line runs suspiciously precisely though the centres of all the error boxes which have been shown as 1 per cent of the peak intensity. This means that the quoted errors are somewhat conservative estimates of the uncertainties. The most recent analysis of these magnificent data which I heard discussed only 2 months ago indicated that the spectrum is now known to be a black-body with an uncertainty of only 0·25 per cent of the peak intensity at wavelengths longer than 500 μm. These are quite remarkable results. It is certainly by far the most precise black-body spectrum I know of in nature.

What is the significance of this observation? It was Planck who first showed in 1900 that the black-body spectrum is the unique radiation spectrum obtained when matter is in thermodynamic equilibrium with matter at a single temperature. The implication of the observation of such a spectrum for cosmology is that the matter and radiation must have reached a state very closely approximating thermodynamic equilibrium at some time in the past. This occurs naturally in the hot early phases of the standard Big Bang model.

The last requirement of the Friedman models is the physics of the forces which determine the large scale dynamics of the Universe. Gravity is the only large scale force we know of which acts upon all forms of matter and energy and General Relativity is the best theory of gravity we possess. When Friedman first solved the field equations of General Relativity for isotropically expanding universes, the evidence for General Relativity was good but not perhaps overwhelming. The most remarkable result was the prediction of the exact perihelion shift of Mercury which had remained an unsolved problem in the celestial mechanics of the Solar System since the time of its discovery by Leverrier in 1859.

Most of the tests of General Relativity involve the observation of astronomical objects and there has been excellent progress in testing the predictions of the theory, for example, by measuring the deflection of electromagnetic signals from distant astronomical objects as they are occulted by the Sun and the 'fourth' test of General Relativity discovered by Shapiro of the time delay as the electromagnetic radiation from a distant object passes through the gravitational potential of a massive body such as the Sun.

The most spectacular results have come, however, from radio observations of pulsars. These radio sources are rotating, magnetized neutron stars and they emit beams of radio emission from their magnetic poles as shown schematically in Fig. 10. The typical parameters of a neutron star are given in Fig. 10. Observations by Taylor and his colleagues using the Arecibo radio telescope have demonstrated that these are the most stable clocks we know of in the Universe. They have even been able to show up the variations in the time-keeping of even the most accurate laboratory clocks, as has been demonstrated by comparing the times measured by two pulsars against a standard clock.

The pulsars have enabled a wide variety of very sensitive tests to be carried out of General Relativity and the possible existence of a background flux of gravitational radiation but by far the most intriguing systems are those

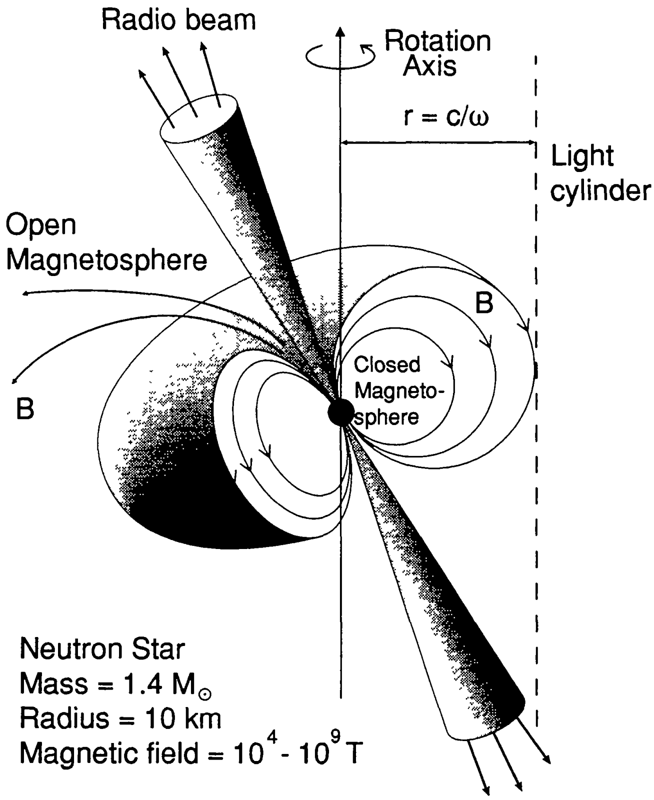
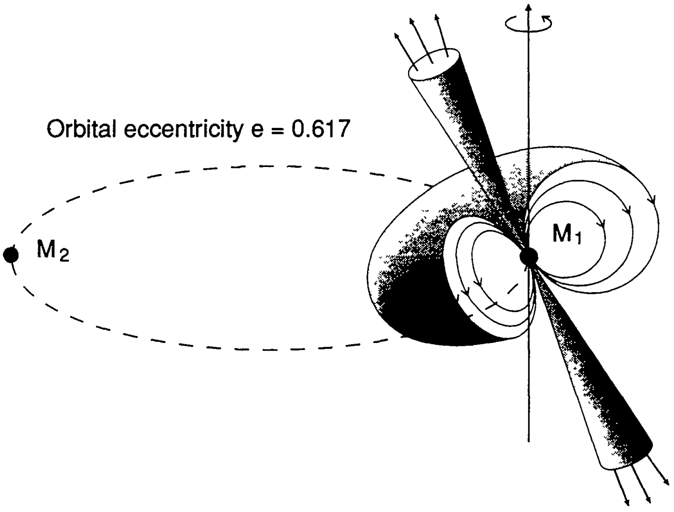


FIG. 10. A schematic diagram of a pulsar showing the displacement between the axis of the magnetic dipole and the rotation axis of the neutron star. The radio pulses are assumed to be due to beams of radio emission from the poles of the magnetic field distribution and are associated with the passage of the beam across the line of sight to the observer. Typical parameters of the neutron stars are indicated on the diagram.

pulsars which are members of binary systems. More than 20 of these are now known, the most important being those in which the other member of the binary system is also a neutron star and in which the neutron stars form a close binary system. The first of these to be discovered was the binary pulsar PSR 1913+16 which is illustrated schematically in Fig. 11. The system has a binary period of only 7·75 hours and the orbital eccentricity is large, $e = 0{\cdot}617$. This system is a pure gift for the relativist. To test General Relativity, we require a perfect clock in a rotating frame of reference and systems such as PSR 1913+16 are ideal for this purpose. The neutron stars are so inert and compact that the binary system is very 'clean' and so can be used for some of the most sensitive tests of General Relativity yet devised. To give just a few examples of the precision which can be obtained, I reproduce with the kind permission of Professor J.Taylor some of the recent tests which have been made of General Relativity.

In Fig. 12, the determination of the masses of the two neutron stars in the binary system PSR 1913+16 is shown assuming that General Relativity is the correct theory of gravity. Various parameters of the binary orbit can be measured very precisely and these provide different estimates of functions involving the masses of the two neutron stars. In Fig. 12, the various parameters of the binary orbit are shown, those which have been measured with very good accuracy being indicated by an asterisk. It can be observed



Binary period = 7.751939337 hours
Pulsar period = 59 milliseconds
Neutron star mass M_1 = 1.4411(7) $M_\odot$
Neutron star mass M_2 = 1.3874(7) $M_\odot$

FIG. 11. A schematic diagram showing the binary pulsar PSR 1913+16. As a result of the ability to measure precisely many parameters of the binary orbit from ultra-precise pulsar timing, the masses of the two neutron stars have been measured with very high precision. (Data courtesy of Professor J.Taylor.)

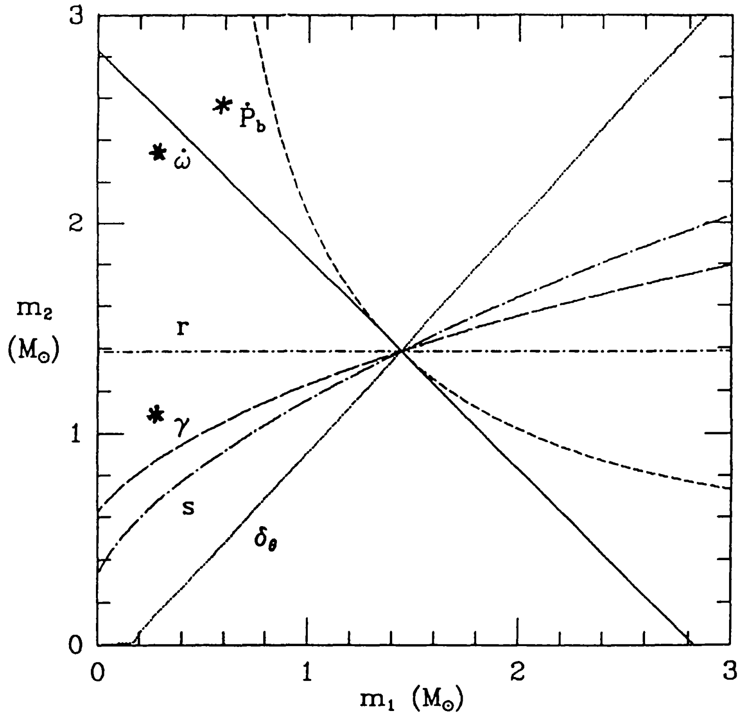


FIG. 12. The measurement of the masses of the neutron stars in the binary system PSR 1913+16 resulting from very precise timing of the arrival times of the pulses at the Earth. The different parameters of the neutron star's orbit depend upon different combinations of the masses m_1 and m_2 of the neutron stars. It can be seen that the lines intersect very precisely at a single point in the m_1–m_2 plane. *Accurately measured. (Courtesy of Professor J.Taylor.)

that the different loci intersect at a single point in the m_1/m_2 plane. Some measure of the precision with which the theory is known to be correct can be obtained from the accuracy with which the masses of the neutron stars are known as indicated in Fig. 11.

A second remarkable measurement has been the rate of loss of orbital rotational energy by the emission of gravitational waves. The binary system loses energy by the emission of gravitational radiation and the rate at which energy is lost can be precisely predicted once the masses of the neutron stars and the parameters of the binary orbit are known. The rate of change of the angular frequency Ω of the orbit due to gravitational radiation energy loss is precisely known, $-\mathrm{d}\Omega/\mathrm{d}t \propto \Omega^5$. The change in orbital phase due to the emission of gravitational waves has been observed over a period of 17 years and the observed changes over that period agree precisely with the predictions of General Relativity (Fig. 13). Thus, although the gravitational waves themselves have not been detected, exactly the correct energy loss rate from the system has been measured – it is generally assumed that this is convincing evidence for the existence of gravitational waves and this observation acts as a spur to their direct detection by future generations of gravitational wave detectors. This is a very important result for the theory of gravitation since this result alone enables a wide range of alternative theories of gravity to be eliminated. For example, since General Relativity predicts only quadrupolar emission of gravitational radiation, any theory which, say, involves the dipole emission of gravitational waves can be eliminated.

Thus, General Relativity has passed every test which has been made of the theory and we can have much greater confidence than in the past that it is an excellent description of the relativistic theory of gravity. The same techniques of accurate pulsar timing can also be used to determine whether or not there is any evidence for the gravitational constant G changing with time. These tests are slightly dependent upon the equation of state used to describe the interior of the neutron stars but for the complete range of possible equations of state, the limits of $\dot{G}/G$ are less than about 10^{-11} year^{-1}. Thus, there can have been little change in the value of the gravitational constant over typical cosmological timescales which are about $(1\text{–}2) \times 10^{10}$ years. Continued observations of certain of the binary pulsars should enable this limit to be improved by an order of magnitude. For cosmological studies, there is therefore no reason to use any theory other than General Relativity for describing the large-scale dynamics of the Universe. I feel sufficiently confident that General Relativity is by far the best theory of gravity we possess that I will elevate it to Fact Number 5.

Fact 5. Standard General Relativity has passed the most precise tests which have been devised so far and there is no astrophysical motivation for seeking any different theory

The upshot of this discussion is that we can have confidence in the basic assumptions behind the Friedman models and the next, and much more difficult, step is to determine which particular model, if any, provides the best description of the large-scale dynamics of the Universe.

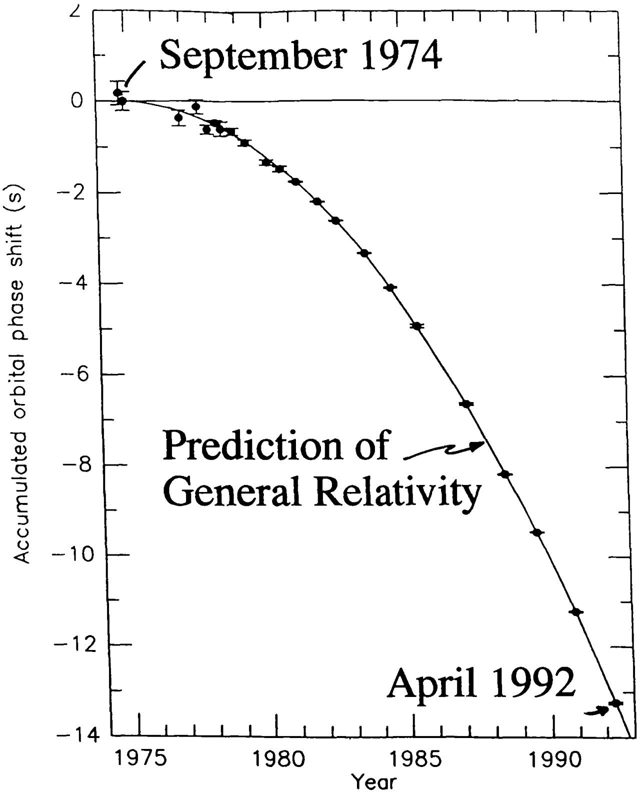


FIG. 13. The change of orbital phase as a function of time for the binary neutron star system PSR 1913+16 compared with the expected changes due to gravitational radiation energy loss by the binary system. (Courtesy of Professor J.Taylor.)

4 THE DETERMINATION OF COSMOLOGICAL PARAMETERS

Since the framework of the Friedman world models is now very well established, the next task is to pin down which of them provides the best description of the large-scale dynamics of the Universe. It turns out that the models are defined by a very small number of parameters. These are:

(1) The present rate of expansion of the Universe as defined by *Hubble's constant*, H_0.

(2) The present deceleration of the Universe as described by the *deceleration parameter* q_0.

(3) The present average density of the Universe ρ. As discussed above, it is convenient to measure the density of the Universe relative to the critical density ρ_{crit} and so define a *density parameter* $\Omega = \rho/\rho_{\mathrm{crit}}$.

In the classical world models, the deceleration of the universal expansion is entirely due to the gravitational influence of the matter content of the Universe and, according to the Friedman models, $q_0 = \Omega/2$. Now the deceleration parameter and the mean density of matter in the Universe can be measured quite independently and so this prediction provides a test of the General Theory of Relativity on the scale of the Universe itself.

There is only one wrinkle in this story and that concerns the fact that, in his paper on the static Universe of 1917, Einstein introduced a further term into the field equations, the infamous *cosmological constant* Λ. Because of the

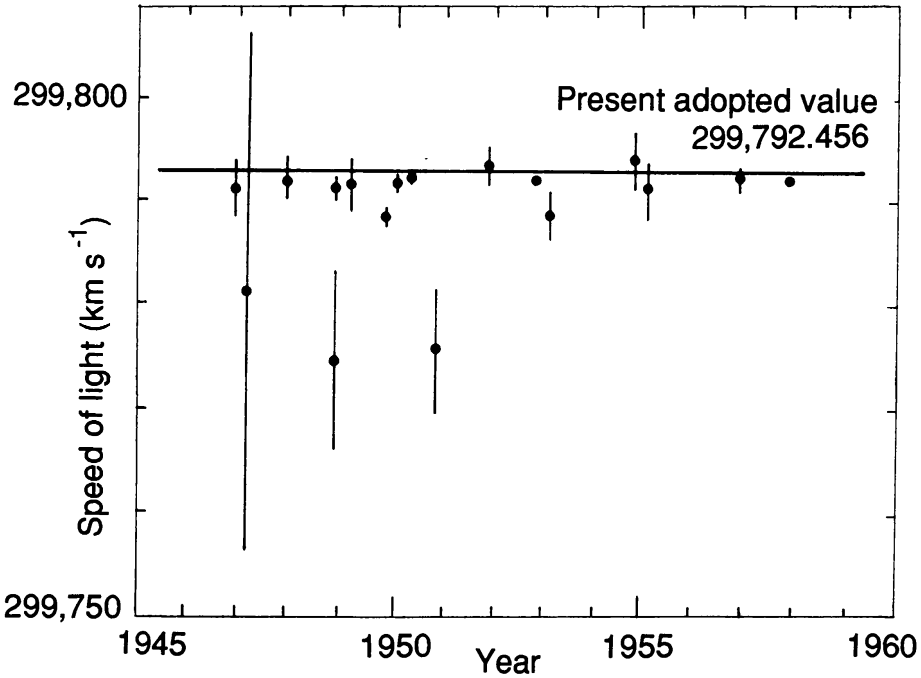


FIG. 14. Precise measurements of the speed of light published in the period 1945 to 1960. (Data from *The American Handbook of Physics*, 1969. 2nd edition. New York: McGraw Hill.)

attractive nature of gravity on the large scale, static Universes are not feasible unless a large-scale repulsive force is included in the equations to counteract the attractive influence of gravity. Once Hubble discovered that the Universe is not in fact static but expanding, there was no longer any reason to include this arbitrary term in the field equations and Einstein stated that the introduction of the Λ term 'was the greatest blunder of my life' (Einstein, quoted by Gamow 1970). The presence of the cosmological constant changes the relation between q_0 and Ω:

$$q_0 = \frac{\Omega}{2} - \frac{1}{3}\frac{\Lambda}{H_0^2}. \tag{1}$$

At the present time, there is absolutely *no* evidence from any astronomical observation that Λ is not zero. As Zeldovich has remarked, however, 'the genie is out of the bottle and, once he is out, it is very difficult to put him back again'. With great regularity, the cosmological constant has kept reappearing in the literature in response to some astronomical anomaly only to be pushed back into the bottle until its next appearance. We will find that the cosmological constant continues to haunt the subject but now in a completely different guise in the inflationary scenario for the very early Universe.

4.1 *Hubble's Constant*

There are currently two schools of thought, one finding values of H_0 in the range 80–100 km s^{-1} Mpc^{-1} and the other values in the range 45–60 km s^{-1} Mpc^{-1}. The big trouble is the difficulty of measuring accurate distances for the galaxies which sample the overall Hubble flow. The extragalactic distance scale has traditionally been determined by calibrating one set of distance indicators against another and so proceeding from the distance scale established in our Galaxy to extragalactic distances. This is an unresolved problem and my personal opinion is that the discrepancy simply reflects the difficulty of measuring extragalactic distances precisely. I have in mind a diagram I produced for pedagogical purposes of precise measurements of the speed of light between 1945 and 1960 (Fig. 14). It is noteworthy how often successive measurements lie outside the formal uncertainties of previous measurements and how many of the measurements are formally inconsistent with the present adopted value for the speed of light. Note also that these are laboratory measurements for which one might have thought that the errors could be precisely estimated. Almost certainly the problem lies in some unrecognized systematic error in the distance indicators used to measure the distances of the galaxies.

I would make two comments about this problem. The first is that there is an urgent need to develop better *physical* methods of measuring extragalactic distances. Most of the steps in the chain of arguments which lead to values of Hubble's constant by the traditional arguments involve assuming that the same types of astronomical object can be selected in different galaxies at different distances – this is why they are called distance indicators. In contrast, in the direct methods, the need to use distance indicators is eliminated by evaluating some physical dimension d at the distant galaxy and then, by measuring its angular size θ, the distance D to the galaxy can be measured from $D = d/\theta$. The trick is to be able to find a method of measuring physical sizes of objects at extragalactic distances. I will mention only three of the more promising techniques. One method uses supernova explosions of Type II. In these explosions, the speed of expansion of the supernova as well as its spectrum and luminosity are measured as it decreases from maximum light. The physical rate of expansion is measured from the width of the emission lines and the change in angular size can be estimated from the change in surface brightness as the supernova expands. In the most recent observations, accuracies as good as those obtained by the traditional techniques are being found (Branch 1988, Kirshner & Schmidt (personal communication 1992)). A second method is to use the phenomenon of gravitational lenses to work out the geometry of the lensing galaxies and thus measure physical distances at the distance of the lensing galaxy. A third good example is the use of the Sunyaev–Zeldovich effect in clusters of galaxies which contain large amounts of hot X-ray emitting gas. In this effect, a decrement is observed in the intensity of the Cosmic Microwave Background Radiation in the direction of the hot gas cloud because of Compton scattering of the background photons by the electrons of the hot gas. Combining all the data on the properties of the hot gas cloud enables its physical size to be determined and hence, by observing its angular size, its distance can be measured.

Another beautiful example of a physical measurement of an extragalactic distance has resulted from Hubble Space Telescope observations of the Type II supernova SN 1987A which exploded in the Large Magellanic Cloud in 1987. A thin ring of ionized gas was observed by the Hubble Space Telescope about the supernova which was excited by the initial outburst of ultraviolet

1993QJRAS..34..157L

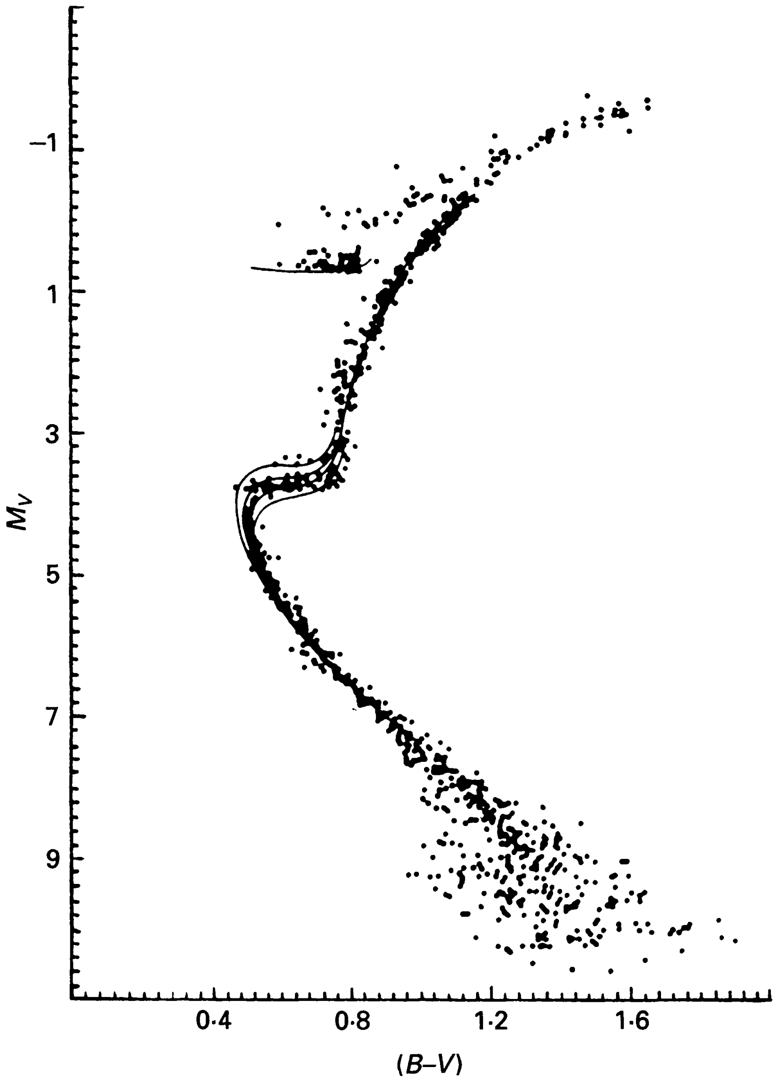


FIG. 15. The H-R diagram for the globular cluster 47 Tucanae. The scatter in the points increases towards faint magnitudes because of the increase in observational error associated with the photometry of faint stars. The solid lines show best fits to the data using theoretical models for the evolution of stars from the main sequence onto the giant branch due to VandenBerg. For this cluster, the best-fit isochrones have ages between about 12 and 14×10^9 years and the cluster is metal rich relative to other globular clusters, the metal abundance corresponding to about 20 per cent of the solar value. (From Hesser, Harris, VandenBerg, Allwright, Schott and Stetson [1989].)

ionizing radiation. The ultraviolet emission lines from the same ring were monitored by the International Ultraviolet Observatory and they increased very abruptly from zero to maximum intensity and then decayed. The convincing picture which has been adopted to explain these observations is that the time evolution of the strengths of the ultraviolet emission lines gives information about the time delay of the arrival of light at the observer from different parts of the ring. From this, the physical size of the ring can be found and hence, knowing its angular size, the distance to the supernova can be found (Panagia *et al.* 1991). This technique has resulted in as accurate a measurement of the distance to the Magellanic Cloud as any of the more

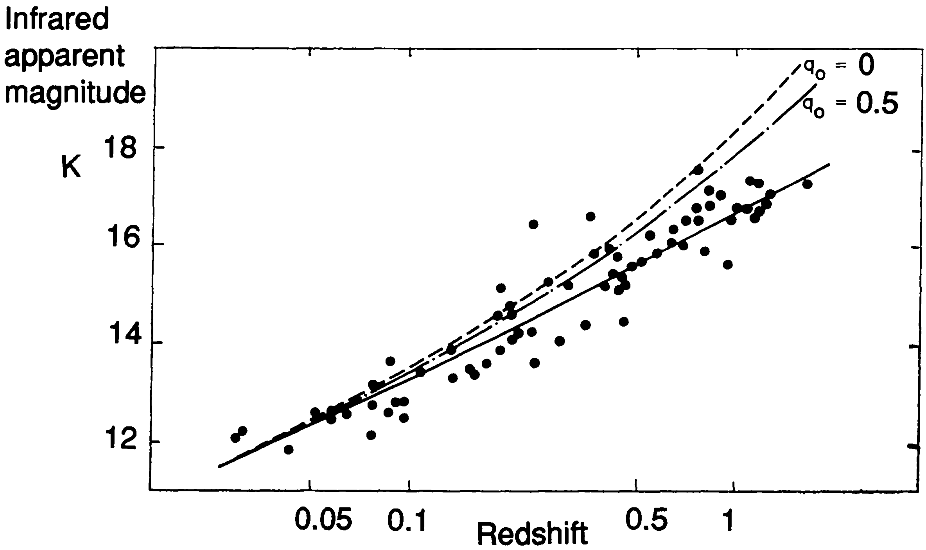


FIG. 16. The infrared redshift–magnitude relation for a complete sample of radio galaxies selected from radio sources in the revised 3C catalogue. The dashed lines show the expectations of world models having $q_0 = 0$ and 0·5. The solid line shows the expected relation when account is taken of the evolution of the stellar populations of the radio galaxies. (Lilly & Longair 1984.)

traditional methods and agrees well with the estimate of the distance to the supernova by the technique described above.

The second point is the use of more astrophysical methods of setting limits to Hubble's constant from the measurement of the ages of the oldest stellar systems in galaxies. In the standard world models, the age of the Universe is $t_0 = f(\Omega)/H_0$ where $f(\Omega)$ is less than or equal to 1. For the critical model, $\Omega = 1, f(\Omega) = 2/3$ and only for $\Omega = 0$, the Milne model, is $f(\Omega) = 1$. Thus, by measuring the ages of the oldest objects we can find in the Universe, we can find limits to both H_0 and Ω. The best example of this approach is the beautiful work of Hesser and his colleagues in estimating the age of the globular cluster 47 Tucanae. The method involves measuring very precise colours and magnitudes for large numbers of stars in the cluster and then fitting the distribution of points in the resulting Hertzsprung–Russell diagram by models of the evolution of stars in old clusters. The type of analysis which can now be undertaken is illustrated in Fig. 15. According to Hesser *et al.* (1989), the age of the globular cluster is probably between about $(12–14) \times 10^9$ years old. If Ω were equal to 1, then H_0 would have to be less than or equal to 50 km s^{-1} Mpc^{-1}. In my opinion, this type of constraint upon H_0 is just as important as the traditional route through the calibration of the Hubble diagram.

There is clearly still about a factor of 2 uncertainty in the value of Hubble's constant. My instincts are to adopt a low value for Hubble's constant but that is a working hypothesis rather than an established cosmological fact.

4.2 *The deceleration parameter* q_0

Effectively, in this test, we attempt to measure the rate of change of Hubble's constant as the Universe grows older. We therefore have to make observations of the Universe when it was significantly younger than it is now, that is, when the scale factor $R \sim 0{\cdot}5$. It is a general property of the homogeneous, isotropic world models that the scale factor is directly related to the redshift of the galaxy z, through the relation $R = (1+z)^{-1}$ if we normalize the scale factor so that it takes the value 1 at the present epoch. Thus, we have to measure galaxies with redshifts $z \sim 0{\cdot}5 - 1$ in order to measure the deceleration of the Universe. The effects of the deceleration are reflected in variations in the properties of identical objects as they are observed at different redshifts. The method which has proved to be the most promising is to identify a class of galaxy with more or less standard properties and to find out how the observed intensity changes with increasing redshift. The problem is to find the classes of galaxy which can be observed at large redshifts and which have a narrow dispersion in their intrinsic properties.

Unfortunately, the options are quite limited. The quasars extend to very large redshifts, almost up to a redshift of 5, but they have a very wide dispersion in their intrinsic luminosities (Fig. 2(b)). Even worse, it appears that their mean luminosities have changed with cosmic epoch and so they are not very useful as standard candles. The brightest galaxies in clusters are probably the best standard extragalactic systems we have but at present the samples of clusters extend only out to redshifts of about 0·5 which is scarcely far enough to measure the deceleration parameter accurately without very large samples of clusters at that redshift which are not yet available (Fig. 1). Probably the best samples at the moment are the radio galaxies which are almost as good standard candles as the brightest galaxies in clusters but which can be observed out to redshifts of about 3 (Fig. 2(a)). There is, however, a fundamental problem and that is that we have to take account of the evolution of the properties of the galaxies with cosmic epoch.

When galaxies are observed at large redshifts, they are observed at significantly earlier stages in their evolution. For example, in the critical world model ($\Omega = 1$, $q_0 = 0{\cdot}5$), the relation between cosmic time, the scale factor and redshift is very simple:

$$R = \frac{1}{(1+z)} = \left(\frac{t}{t_0}\right)^{\frac{2}{3}}, \tag{2}$$

where t_0 is the age of the Universe which for this model is $t_0 = (2/3)H_0^{-1}$. Thus, for the critical model, the Universe was only 35 per cent of its present age at a redshift of 1, 19 per cent of its present age at a redshift of 2 and so on. It is evident that we cannot assume that the galaxies have remained unchanged over such time-scales. Indeed, when we look at the redshift–magnitude relation for the radio galaxies, there is clear evidence that they were considerably brighter in the past. In Fig. 16, the redshift–magnitude relation for a complete sample of bright radio galaxies extending out to redshifts of almost 2 is compared with the expectations of the standard world models. One of the important aspects of this diagram is that the observed intensities are infrared magnitudes measured at 2·2 μm rather than optical magnitudes. This has many advantages over working in the optical waveband, principally because of the absence of the obscuring effect of dust and because the evolutionary effects are much easier to estimate than in the optical waveband. Recalling that, when intensities are measured in magnitudes, the greater the apparent magnitude, the fainter the object, it can be seen that the galaxies at redshifts about one are brighter than is expected for standard world models for which $q_0 = 0$ and $q_0 = 0{\cdot}5$.

Simon Lilly and I interpreted these results as indicating that the radio galaxies at a redshift of about 1 were brighter by about 1 magnitude as compared with their luminosities at the present epoch. It turns out that this is almost exactly the change in luminosity expected if the galaxies at redshifts of 1 have the same numbers of stars as similar objects nearby but that the stellar population was only half its present age or less. The galaxies were brighter in the past than they are now because there were more bright stars populating the giant branch at that time. It turns out that there are great advantages in carrying out this type of analysis in the near infrared waveband because the evolution corrections are remarkably model-independent (Lilly & Longair 1984). Figure 2(a) shows that the more recent data are entirely consistent with our original work. The upshot of these studies is that world models with $\Omega \sim 0-2$, $q_0 \sim 0-1$ are consistent with the redshift–magnitude relation for these galaxies once the corrections for the effects of the stellar evolution of the populations of the galaxies are taken into account.

Thus, in this case, there is strong evidence for the evolution of the properties of the radio galaxies with cosmic epoch. There is, however, much more evidence for changes in the populations of various classes of object with cosmic epoch. The half fact which Peter Scheuer told me about in 1963 has now become part of what I will call Fact 6.

Fact 6. Many different classes of extragalactic system show changes in their average properties with cosmic epoch

Let me summarize some of the evidence for these changes.

(1) The counts of galaxies to faint optical magnitudes show an excess of faint blue galaxies (Fig. 17). The natural interpretation of these results is that there were more blue galaxies in the past than there are now. The problem with this interpretation is that, when the redshifts of the galaxies responsible for the excess are measured, they do not seem to be any more distant than would be expected if there were no blue excess. In other words, there simply seem to be more blue galaxies at redshifts of about 0·5 than there are at the present epoch (Ellis 1992). This programme of observation is at the very limit of capability of the present generation of large telescopes. The galaxies to be observed are very faint indeed and very long exposures using fibre optic multi-object spectrographs are necessary to make even the results described above possible. The other consequence of these observations is just how difficult it is to reach out to redshifts of the order 1 by selecting faint galaxies at random. If I had been asked several years ago how to find normal galaxies at a redshift of 1, I would have said, 'Make a very deep optical survey of galaxies, pick out the blue ones and find their redshifts'. This turns out not to be the case. One of the most important

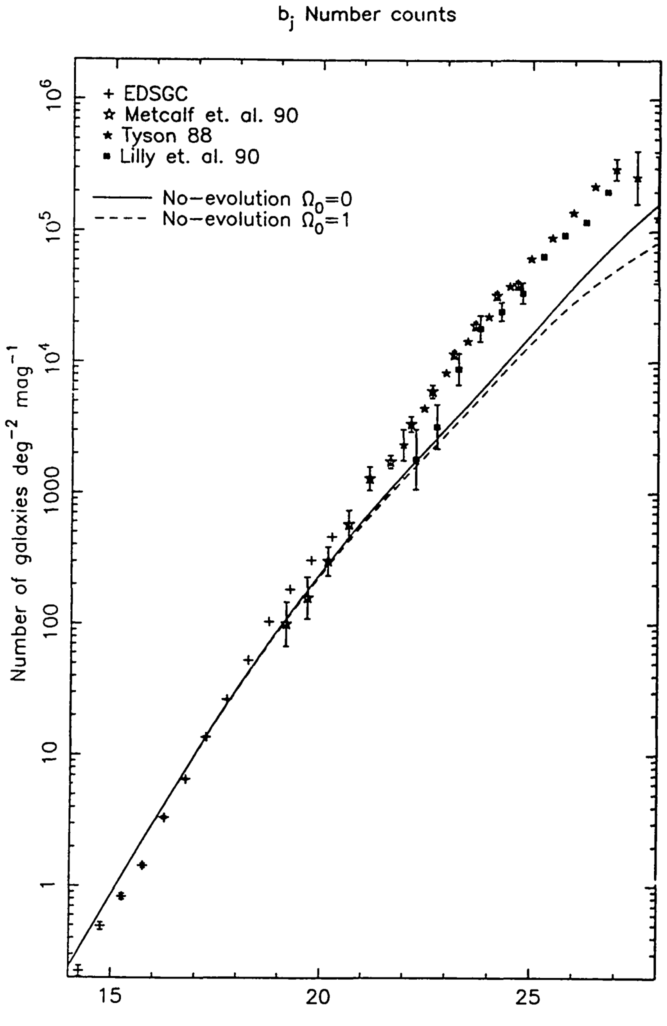


FIG. 17. The counts of faint galaxies observed in the blue (B) waveband compared with the expectations of uniform world models with $\Omega = 0$ and $\Omega = 1$ (after Jones *et al.* 1991, Metcalfe *et al.* 1991).

projects for the next generation of 10-m class optical-infrared telescopes will be to extend the spectroscopic surveys of faint galaxies to yet fainter limits.

(2) The populations of extragalactic radio sources and radio quasars show very strong evolutionary changes with cosmic epoch. The nature of the changes is illustrated in Fig. 18 which shows how the luminosity functions of radio sources with flat and steep radio spectra change with cosmic epoch.

(3) Optically selected samples of quasars show similar evolutionary changes with cosmic epoch to that of the radio quasars and radio galaxies. The surveys of Boyle and his colleagues (Boyle *et al.* 1987) show that the optically selected quasars were much more populous at redshifts of the order 2 than they are now. The simplest way of accounting for these changes quantitatively is to assume that the luminosities of all the quasars were on average about an order of magnitude greater at

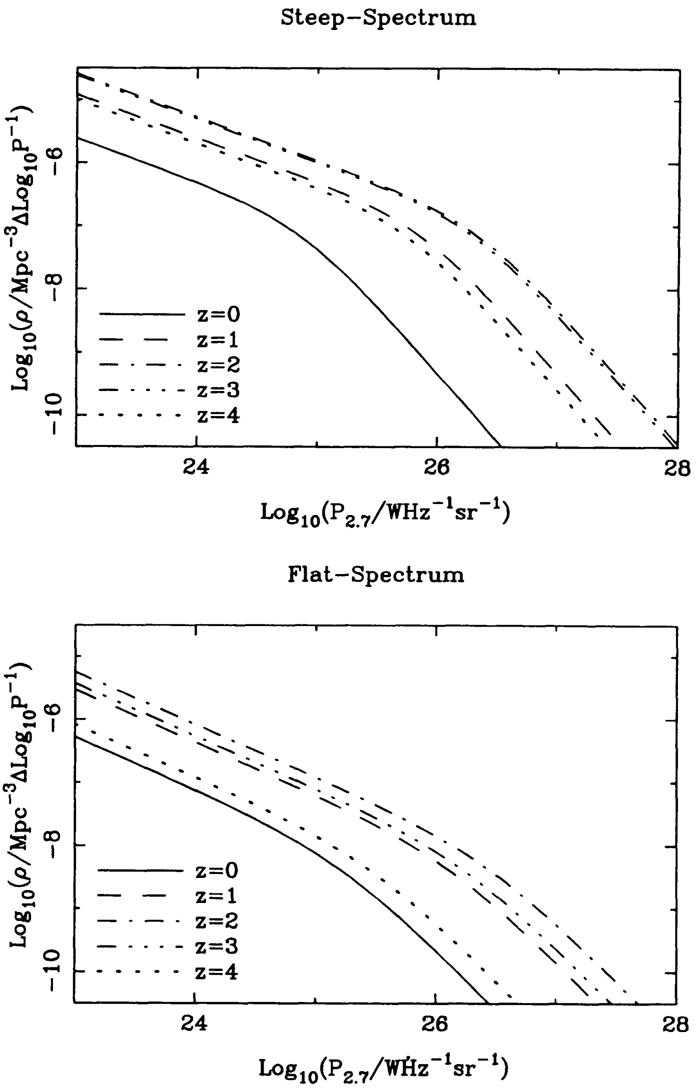


FIG. 18. Illustrating the evolution of the luminosity function of extragalactic radio sources with steep and flat radio spectra as a function of cosmic epoch. The luminosity function describes the number of radio sources of different radio luminosities per unit comoving volume, that is, in a coordinate system which expands with the Universe so that the figure shows the changes over and above the changing density due to the expansion of the Universe. (Dunlop & Peacock 1990.)

redshifts $z \sim 2-3$ than they are at the present day. According to Peacock (1992), these evolutionary changes can be described by functions of the following form:

$$\left.\begin{aligned} L(z) &= L(0)(1+z)^3 \text{ for } 0 < z < 2 \\ L(z) &= \text{constant} = 27L(0) \text{ for } z > 2, \end{aligned}\right\} \qquad (3)$$

where $L(0)$ is the luminosity of the quasar at zero redshift.

It turns out that the evolution of the radio properties of those quasars which are bright radio sources can be explained by exactly the same type of model. Unfortunately, no one knows exactly what the significance of this result is. Apparently, the epoch when the quasars were at their most active is not the present epoch but the time when the Universe was about 20 per cent of its present age. The behaviour of the population at even larger redshifts is not understood but the increase in luminosity does not continue beyond redshifts of about 2–3. Although Fig. 18 suggests that the numbers of radio galaxies and radio quasars decrease at the largest redshifts, $z \sim 4$, the statistics are limited and it is not certain whether or not the overall population decreases at the largest redshifts. The problem is that there are fewer and fewer quasars at very large redshifts and it requires a huge effort to find the largest redshift quasars. Nowadays, however, there are about 20 quasars known with redshifts greater than 4 and so eventually we may know the answer but it is a very long and difficult task.

(4) The physical sizes of double radio sources are smaller at redshifts $z \sim 1$ than they are at the present epoch, $z \sim 0$.

(5) In the spectra of distant quasars, absorption lines are observed associated with the intergalactic clouds along the line of sight to the quasar. Studies of the space distribution of these clouds with cosmic epoch have shown that the clouds which show only hydrogen absorption lines increase in number with increasing redshift while those systems which possess metals increase in number as the redshift decreases. Again, the cause of these changes is not understood but the reality of these changes has been clearly demonstrated.

In one way or another, these observations provide information about the sequence of events which must have taken place as the Universe evolves to its present state. Unfortunately, the evidence is still too sparse to build these separate pieces of evidence into a convincing self-consistent picture. Part of the trouble is that there are only limited classes of object which can be observed at large redshifts. It is still an impossible task to study normal galaxies at the same distances as the most distant quasars and radio galaxies. Optimistically, we may be able to extend these studies for normal galaxies to redshifts of about 1 with the next generation of 10-m class telescopes and this is an essential step before we are able to tackle the evolution of normal systems at the very largest redshifts.

4.3 *The density parameter* Ω

It might seem that we should be able to make better estimates of the density parameter Ω since we only have to count up all the galaxies in the Universe and work out the total amount of mass associated with them. If we do this calculation, it is found that the average density of matter in the Universe corresponds to only about 1 per cent of the critical density. We know, however, that this is a serious underestimate of the total amount of mass present. If we make observations of the outer regions of giant spiral and elliptical galaxies, it is found that there must be about ten times as much mass present as would be inferred from the optical light. The reason is that with increasing distance from the centre of a galaxy, the light falls off much more rapidly than the mass so that the visible galaxies are surrounded by dark haloes. The same problem is found in clusters of galaxies. The total mass of clusters of galaxies can be estimated from their internal velocity dispersions and this exceeds by a factor of about 10–20 the mass which would be inferred from the optical light alone. These are aspects of the famous *dark matter problem* and it is so pervasive in extragalactic astronomy that I will elevate it to the status of a fact.

Fact 7. Most of the mass of the Universe is in some dark form and it exceeds the amount of visible matter by a factor of at least ten

It will be noted that I have been careful not to make any statement about: (1) How much there actually is? and (2) What is it?

We only possess firm lower limits to the amount of dark matter present. It is likely that there is at least 10–20 times the amount of visible matter but a key issue is whether or not there could be about a factor of 100 times more matter so that the Universe approaches its critical density. In recent work, it is claimed that, when the Universe is observed on scales much greater than clusters of galaxies, more dark matter is found and that the density parameter approaches 1. It is somewhat disturbing that, if this really is the case, most of the dark matter must reside in the regions in which it is most difficult to observe it – namely, in the regions between clusters of galaxies. There remains a great deal to be done to establish precisely how much dark matter there really is. A personal concern I have is that, in their enthusiasm for the critical model of the Universe, which is one of the predictions of the inflationary model of the early Universe, some of the interpreters have been more than anxious to show that our Universe does indeed have the critical density.

If estimating the total density of the Universe is difficult, determining what the dark matter is, is even more uncertain. The basic problem is that, if some constituent of the Universe does not emit or absorb much radiation, it is very difficult to detect and so it is remarkably easy to hide dark matter. The example which I like to quote is that, if the critical density were made up of standard bricks, we would not know about it from any observation because they would not emit detectable radiation and would not absorb background radiation either. In my opinion, this is such an important measurement that there is an urgent need for more observations to tie down more precisely exactly what forms the dark matter could take.

One form which I am sure must exist is what might be called ordinary (or baryonic) dark matter in the form of cold dust, rocks, cool solid bodies and all the way up to planet-sized objects and brown dwarfs. Objects such as brown dwarfs are expected to be cool but none of them has yet been definitively found in infrared surveys despite intensive efforts. I am sure these searches must continue because this issue is crucial for cosmology. Another form of dark matter are the black holes, either massive black holes, solar mass black holes or mini-black holes. One very beautiful idea, which is the subject of a very large survey at the present time, is to search for dark matter candidates by searching for the rare gravitational lensing effects expected when a black hole, brown dwarf or planet passes in front of a background star. When this occurs, there is a characteristic brightening and dimming of the light from the star and many of the properties of the intervening dark object can be found from the characteristic signature of the changing brightness of the star. This is a very demanding survey because lensing events are very rare but they must occur at some level if the dark matter in our Galaxy is in some discrete form (Alcock 1992).

The forms of dark matter which have caused the greatest excitement in the theoretical community are ultraweakly interacting particles, as yet unknown to science. These are predicted to exist according to different versions of those theories which seek to unify the forces of nature – grand unified theories, supersymmetry, superstring theories and so on. Examples of these types of hypothetical particle include photinos, gravitinos, axions and so on. None of these has yet been observed in laboratory experiments. According to current speculation, the early Universe was hot enough for these particles to be produced and so the particle physicists invert the whole process and use the very early Universe as a laboratory within which to test theories of elementary processes. The methodological problem is that cosmology and particle physics then boot-strap their way to a self-consistent solution and there is no independent way of constraining the theories by observation or experiment.

One important possibility is that the least massive of these hypothetical particles is stable and so might well be present in the Universe now as relics of the very hot early phases. As we will argue in the next section, any such particles would have to be rather massive. It is now possible to search for these dark matter particles by laboratory searches. The idea is that, if the dark matter in our own Galaxy were made up of these exotic particles, we know roughly what their velocities would have to be because they have to form a bound self-gravitating halo about the Galaxy. Therefore, a considerable number of them would be passing through terrestrial laboratories each second. The new generation of very low temperature crystal detectors can detect the very rare collisions between one of these particles and the crystal lattice resulting in a tiny but measurable temperature increase in the crystal. Such experiments are now underway in a number of countries and they should enable constraints to be placed upon a wide range of possible candidates for the exotic dark matter. The optimists argue that this approach is no different from that of Newton and Einstein in that astronomical discoveries and problems result in new physical concepts and suggest experiments which lead to the discovery of phenomena which could not be detected by purely laboratory experiments. We shall have to wait and see. I will return to this interface between particle physics and cosmology later.

What can we say about the value of the density parameter Ω? I believe it is best to treat the statement in three parts. It is certain that Ω is greater than about 0·01 because that corresponds to the amount of visible matter present in the Universe. Probably Ω is at least 0·1 because of the presence of the dark haloes about giant galaxies and the dark matter in clusters of galaxies. Possibly Ω is about 1 from the peculiar velocities of galaxies on a large scale but, in my view, this is not established with any real certainty.

4.4 *The comparison of* Ω *and* q_0

As for our comparison of the deceleration parameter q_0 and the density parameter Ω, it is probable that they are equal within a factor of about 10 although I believe the measurement of q_0 is quite uncertain in the range 0–1. If this is the case, some cosmologists argue that they are very likely to satisfy the equality $q_0 = \Omega/2$ but we would really like to know this with much more certainty from improved measurements of both Ω and q_0.

Another way of looking at these numbers is to note that Ω is probably within a factor of 10 of the critical value $\Omega = 1$. The cosmologist can argue that this surely cannot be a coincidence. There is nothing in the standard model of the Universe which determines what the value of Ω should be but there is one important aspect of the evolution of the standard models which leads to one of the basic problems of cosmology. It can be shown very easily that, if the Universe has a density which is different from the critical world model at any epoch, then the value of Ω diverges rapidly so that after a long enough time, the Universe would have density parameter either very much greater or very much less than unity. Since the Universe is within a factor of ten of the critical model now, this means that the Universe must have been very close indeed to the critical density in the very distant past. The cosmologist argues that, since it is very unlikely that the Universe was set up with Ω just infinitesimally different from 1 in the very early Universe, the only stable value for the density parameter is exactly one. To many theorists, this is a very persuasive argument and they take the point of view that there is no value which Ω can take other than 1. It is not clear to me how strongly this argument has influenced theorists either consciously or unconsciously in their analyses of cosmological problems.

5 THE HOT EARLY UNIVERSE

Despite the problems of determining the cosmological parameters described in the last section, we can have considerable confidence that the basic physical picture is correct. The case for the essential correctness of the Big Bang comes from the study of the early evolution of these models. It turns out that radiation was the dominant form of 'mass' which determined the early dynamics of the Universe – the Universe was radiation-dominated and the Cosmic Background Radiation we observe today is the cooled remnant of these early stages. As a consequence, we can work out rather precisely the early dynamical evolution of the Universe, in fact, with greater accuracy than we can work out its present dynamics. One of the remarkable results of these studies is that, as the Universe cools down through temperatures of about 10^8–10^9 K, the light elements helium, deuterium and lithium are synthesized from the primordial plasma which at that stage consisted mostly of protons and neutrons. The results are dramatic. It turns out that we can account for the observed abundances of deuterium, lithium and the isotopes of helium for a single value of the ordinary matter density in the Universe now (Fig. 19). This is a great triumph for the Big Bang picture since it has been a great astrophysical problem to account for the observed abundances of these elements by nucleosynthesis in stars. These are very fragile elements and they are destroyed rather than synthesized in stars.

We obtain two essentially independent pieces of information about the dynamics of the Universe during the period when the light elements were synthesized. The amount of helium produced is remarkably independent of the density of matter at that time because the ratio of the numbers of

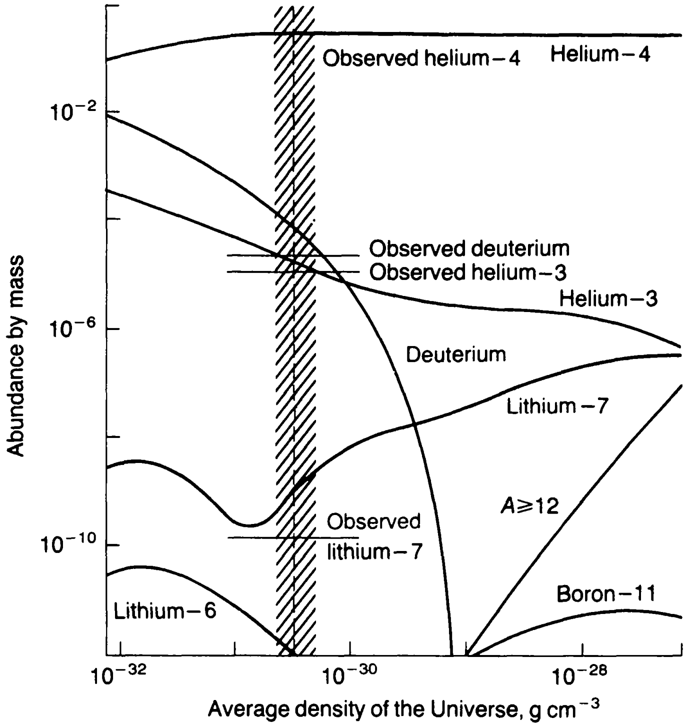


FIG. 19. The predicted abundances of the light elements synthesized in the standard Big Bang model of the early Universe. In the standard model, it is assumed that the lepton number is zero and that the model evolves from an equilibrium state at a temperature of about 10^{12} K. The predicted abundances depend only upon the mean density in ordinary matter now. The abundance of deuterium is a sensitive measure of the density of the Universe now since, if the mean density is high, all the synthesized deuterium is converted into helium. (After J.Audouze 1982.)

neutrons to protons depends only upon the temperature of the Universe when the neutrinos decoupled from the reactions which maintained them in thermal equilibrium with the neutrons and protons. Subsequently, almost all the neutrons combine with protons to create helium nuclei. Thus, the helium abundance is essentially a thermometer for the early Universe. This fact enables us to set further limits on physical processes in the early Universe. If the early expansion of the Universe were any faster than it is in the standard model, the neutron-to-proton ratio would freeze out at a higher temperature resulting in the over-production of helium. This argument enables us to rule out models in which the gravitational constant was stronger in the past and also excludes the possibility that there are more than three neutrino species. In the latter case, the energy density of the early Universe would have been greater resulting in a more rapid expansion than is permitted by the observed abundance of helium. The recent experiments at the Large Electron-Positron collider (LEP) at CERN have confirmed that there are only three neutrino species. This same experiment also shows that there cannot be any unknown weakly interacting particles with rest masses less than about 40 GeV and so the cold dark matter cannot be any neutrino-like particle with rest mass less than about 40 GeV.

The second aspect of these calculations is that the light elements, deuterium and helium-3, are density probes of the early Universe. This is because these elements are by-products of the synthesis of helium from protons and neutrons. In the case of deuterium, the amount produced depends upon the density of the matter content of the Universe. The first step in the synthesis of helium is the formation of deuterium nuclei by combining a proton and a neutron and these deuterons then combine successively with protons and other helium-3 nuclei to form helium. Thus, if there is a high density of protons and neutrons, almost all of the deuterium is converted into helium but, if the density is low, fewer deuterium nuclei are used up in the production of helium-4. Thus, the greater the density, the lower the expected abundance of deuterium.

One of the important results of this study is that it provides an upper limit to the mean density of ordinary matter in the Universe. This turns out to be less than about 10 per cent of the critical density. Now, as discussed above, the theorists have a strong preference for models of the Universe which have exactly the critical density. Therefore, the hard-line inflationist has to assume that most of the mass in the Universe is in some extraordinary form which does not disturb the excellent agreement between the observed abundances of the light elements and the predictions of the standard Big Bang – the matter would have to be in some exotic form of dark matter – black holes, ultraweakly interacting particles, axions, photinos, gravitinos and all the other exotica discussed by the particle physicist. These must not change the early dynamics of the Universe significantly from the canonical picture.

I find the agreement between the observed light element abundances and the predictions of the standard Big Bang Picture so impressive that I will elevate these results to the status of Fact 8.

Fact 8. The light elements, helium, deuterium, helium-3 and possibly lithium, were created primordially

It is the combination of the observational Facts 2, 3, 4 and 8 and the success of General Relativity (Fact 5) which gives us such confidence that the Big Bang model can provide an excellent description of the evolution of the global properties of the Universe. It is therefore the best framework within which to study the more difficult problems of astrophysical cosmology. Let me emphasize two points. The first is that the Facts 2, 3, 4 and 8 are independent pieces of observational evidence, all of which find a natural explanation within the context of Big Bang models of the Universe, although which of them is the best description is not uniquely defined.

The second point is that the reason we can be confident about the success of the model is that we do not need to extrapolate the physics beyond what has been tested in the laboratory. Thus, we are still dealing with known physics from the time the Universe was about 1 millisecond, or even less, old to the present epoch. When we extrapolate to earlier times, we rapidly run out of known physics, the maximum energies for which laboratory experiments have been carried out corresponding to about 80 GeV. Thus, it is only safe to regard many of the inferences about the physics of the Universe earlier than about 1 millisecond as speculative. On the other hand, it is a very respectable aspect of theoretical physics to use the early evolution of the Universe as a constraint upon possible theories of elementary processes. In the strict Popperian sense, we can use the class of Big Bang models as a constraint upon physical theories in that, if the latter result in

1993QJRAS..34..157L

universes which bear no resemblance to our Universe, they can be discarded. Thus, I believe a healthy approach to speculations about the very early Universe is to regard them as hypotheses which can be disproved but to be wary of taking them seriously as real physics until there is some form of independent experimental or observational validation.

For me, the real significance of these remarkable developments is that we can ask meaningful questions about the very early evolution of the Universe. How far one is prepared to extrapolate our present understanding of particle physics is, in my view, a matter of taste. What is unquestionable is the fact that, if we accept the essential correctness of the Big Bang picture, there are some basic problems which we cannot avoid.

Basic Problem 1. This is often known as the *Horizon Problem.* As we go further and further back in time, the distance over which information can be communicated gets smaller and smaller. For example, when we observe the Microwave Background Radiation in opposite directions on the sky, we detect radiation from the last scattering surface when the Universe was contracted by a factor of about 1000 compared to its present size. We can work out easily how far a light ray could have travelled since the beginning of the Universe and convert that into an angular scale on the sky. This turns out to correspond to an angle of only 5°. This means that there is no way in which regions in opposite directions on the sky could have communicated. In other words, according to the standard picture, it is a puzzle why regions in opposite directions on the sky are so precisely the same. How could the different regions of the Universe know that they had to end up looking the same in all directions?

Basic Problem 2. This is often referred to as the *Flatness Problem* and we have already discussed it in Section 4.4. Why is the Universe so close to its critical density $\Omega = 1$ when, *a priori* it could have taken any value at all? Furthermore, all values of $\Omega \neq 1$ are unstable.

Basic Problem 3. The third problem is called the *Asymmetry Problem.* In the Universe now, there are about 10^9 photons of the Cosmic Microwave Background Radiation for every proton. In the very early Universe, at temperatures greater than about 10^{12} K, particle-antiparticle pair production flooded the Universe with protons and antiprotons, neutrons and antineutrons and so on with one particle-antiparticle pair for each pair of photons. Therefore, when the clocks are run forward, a slight asymmetry in favour of matter as opposed to antimatter has to be built in at the level of one part in 10^9 or the present observed ratio of photons to particles is not obtained. Why was there this very slight asymmetry in the initial conditions in favour of matter?

Basic Problem 4. Finally, what was the origin of the fluctuations from which galaxies and the large scale structure of the Universe formed? We will analyse this problem in section 6 but suffice it to say that density perturbations grow so slowly in the expanding Universe that there have to be some 'seed' fluctuations introduced in the very early Universe to produce the large scale features of the Universe we see now. Where did these 'seeds' come from?

Within the framework of the standard Hot Big Bang model of the Universe, these are all *ad hoc* initial conditions which have to be introduced arbitrarily in order to 'explain' the large scale properties of the Universe now. Let me suggest five ways of attempting to solve these problems:

(1) That is just how the Universe is – the initial conditions were set up that way.
(2) There are only certain classes of Universe in which intelligent life can evolve. For example, the fundamental constants of nature should not be too different from the values we observe them to have or else there would be no chance of life ever forming as we know it. This approach is known as the *Anthropic Principle* and it asserts that the Universe is the way it is because we are here to observe it (see Barrow & Tipler 1987, Gribbin & Rees 1991).
(3) Seek clues from particle physics and extrapolate that understanding beyond what has been confirmed by experiments with large accelerators to the earliest phases of the Universe.
(4) The inflationary scenario for the early Universe.
(5) Something else which we have not yet thought of. This will certainly involve some new physics.

There is some merit in each of these positions. In approach (1), it might just be too hard a problem to decipher what it was that set up the initial conditions from which our Universe has evolved. I am particularly mindful of McCrea's Proposition (C). How can we possibly check that the physics adopted for the very early Universe is correct? In approach (2), there is certainly truth in the statement that the mere fact that we can ask questions about the origin of the Universe must say something about the sort of Universe we live in. Whilst the Cosmological Principle asserts that we do not live in any special location in the Universe, we are certainly privileged in that we are able to ask the question at all. I do not like this line of reasoning, however, because it means that we could never seek any physical reason for the relations between the fundamental constants of nature. I regard the Anthropic Principle as the last resort if all other physical approaches fail.

The third approach provides many important clues to possible physical solutions to the basic problems, one of them being approach (4), the inflationary scenario for the early Universe. I like to think of the inflationary model in three stages. The first is *inflation without physics*, in the sense that, if the Universe expanded exponentially by an enormous factor in its very early phases, for whatever reason, we are able to eliminate the first and third of the basic problems. This occurs for two reasons. First, if the scale factor of Universe expanded exponentially in its very earliest phases, regions which were originally very close together are separated exponentially rapidly by the inflationary expansion. Thus, causally connected regions are swept beyond their local horizons by the inflationary expansion. The second effect is that the very rapid expansion has the effect of straightening out the geometry of the early Universe, however complicated it was to begin with. The geometry of the Universe is driven towards flat Euclidean geometry and so, when the inflationary expansion ceases and the Universe transforms over to the 'normal' Universe, the geometry is flat and consequently the Universe must readjust to the critical density. These ideas can be formulated without specific reference to any particular physical realization of the process of inflation. Ironically, this description is no more than de Sitter's model of the dynamics of an empty Universe which he developed to show that there can exist solutions of Einstein's field equations including the cosmological constant even if there is no matter present (de Sitter 1917).

In the second stage, if we put in a little bit of physics, similar to that which accounts for the asymmetry of elementary processes in the 'low-temperature limit', specifically the charge asymmetry of K^0 decay, we can explain the matter-antimatter asymmetry. In the third stage, if one is much bolder, one uses the best theories we have of elementary particles to identify forces which could cause the exponential expansion of the early Universe. The intriguing discovery has been that processes required by the theories of elementary particles bear a close resemblance to what is needed to produce the exponential expansion of the very early Universe. Specifically, the equation of state at very high energies has to be a negative energy equation of state $p = -\rho c^2$ and this is a property of the scalar fields needed to account for the masses of elementary particles. Similar processes are assumed to have taken place in the very early Universe. All the inflationary action is supposed to take place before the Universe was 10^{-35} sec old at extremely high energies (Guth & Steinhardt 1989).

It will be noted that I have referred to the fourth possibility as the *inflationary scenario* since it has been designed to solve specifically problems (1) and (3) in the above list. There is certainly no other evidence for the inflationary picture beyond the need to solve these four problems and the physical realizations of the physics of inflation cannot be tested in the laboratory. I therefore applaud the endeavours of the theorists to give a proper physical basis for the inflationary scenario but it is not clear how we are to find independent evidence that the physics is along the right lines.

Part of the concern is tied up with the fifth approach – the need for new physics. I have shown schematically a popular representation of the evolution of the Universe from the Planck era, when the Universe was only 10^{-44} sec old, to the present epoch (Fig. 20). Halfway up the diagram, from the time when the Universe was only about a millisecond old, to the present epoch, we can be reasonably confident that we have the correct picture for the Big Bang despite the four basic problems described above. However, it will be seen that, at times earlier than about 1 millisec, we very quickly run out of known physics. Indeed, the models of the very early Universe suppose that we can extrapolate across that huge gap from 10^{-3} sec to 10^{-44} sec using our understanding of laboratory physics. The theorists may be correct but one must have some concern that there may be some fundamentally new physics to be understood at higher and higher energies before we reach the Planck era, $t \sim 10^{-44}$ sec.

The one thing which is certain is that at some stage we will have to understand how to quantize gravity. The singularity theorems of Penrose and Hawking show that, according to classical theories of gravity under very general conditions, there is inevitably a physical singularity at the origin of the Big Bang (see Hawking & Ellis 1973). One of the possible ways of eliminating this problem may be to find a proper quantum theory of gravity. This remains an unsolved problem and we can be certain that our understanding of the very earliest stages will remain seriously incomplete until it is solved. Thus, there is no question but that there is some new physics needed if we are to develop a serious physical picture of the very early Universe.

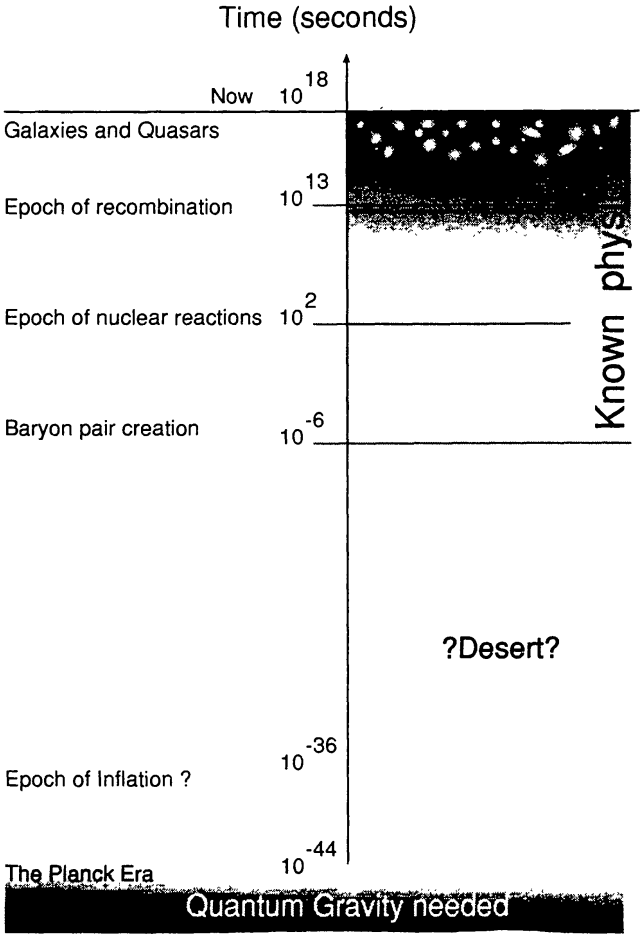


FIG. 20. A schematic diagram illustrating the evolution of the Universe from the Planck era to the present time. The shaded area to the right of the diagram indicates the regions of known physics.

6 THE ORIGIN OF GALAXIES

There is not a great deal that the astronomers can do to help solve the first three of the great problems which I listed above. We must continue striving to determine the basic cosmological parameters with greater precision but there is not much more that can be done concerning the isotropy of the Universe and the Asymmetry problem. Where the astronomers have a direct link to the very early Universe is in the fourth problem, the study of the origin of the fluctuations from which galaxies formed. Let me show what can be done and the problems of interpretation.

The big problem in understanding the origin of galaxies is to reconcile the quite remarkable smoothness of the Cosmic Background Radiation (Fig. 8) with the gross irregularity in the distribution of galaxies, most vividly portrayed in the three-dimensional picture of their large-scale distribution (Fig. 21). If the distribution of galaxies were uniform, the sectors would be uniformly filled with points. It can be seen that the distribution of galaxies is very non-uniform with connected structures on a scale much greater than clusters of galaxies. There are huge two-dimensional sheets of galaxies as well as filaments and great 'voids'. One of the big advances in these studies has been in quantifying the clustering of galaxies on different scales and in describing the large-scale topology of the galaxy distribution.

One measure of the clustering of galaxies can be obtained from the 2-point correlation function $\xi(r)$ which describes the excess likelihood of finding a galaxy at distance r from any given galaxy over a uniform distribution. It turns out that on scales up to about 20 Mpc, the two-point correlation function $\xi(r)$ can be described by a simple power-law of the form

$$N(r)\,dV = N_0[1+\xi(r)]\,dV \qquad \xi(r) = \left(\frac{r}{r_0}\right)^{-1\cdot 8}. \tag{4}$$

The scale on which ξ is equal to 1 is about 8 Mpc. The intriguing point is that there is no preferred scale in this power-law distribution. Thus, although we can identify clearly groups and giant clusters of galaxies, these are no more than parts of a continuous scale of clustering which extends from small groups of a few galaxies right up to the giant clusters and beyond. On the very largest scales the two-point correlation function begins to turn over. Notice that the two-point correlation function is spherically symmetric and therefore cannot give a proper description of the full three-dimensional complexity of the distribution of galaxies.

Another approach taken by Gott and his colleagues has been to evaluate the topology of the distribution of holes and structures in the local distribution of galaxies (Gott 1987). What they find is that the distribution of the galaxies on the large scale is 'sponge-like'. The material of the sponge represents the location of the galaxies and the holes represent the large voids seen in Fig. 21. Thus, both the holes and the distribution of galaxies can be thought of as being continuously connected throughout the local Universe. These are facts which have to be explained by any satisfactory theory of galaxy formation.

Fact 9. The distribution of galaxies on large scales in the Universe, although uniform on the cosmological scale, possesses large-scale irregularities on a scale much greater than that of clusters of galaxies

The problem arises because whereas, in a static medium, any small density perturbation on a large enough scale grows exponentially with time, in an expanding medium, the rate of growth of the perturbation is only *algebraic.* The result first derived by Lifshitz can be written

$$\frac{\Delta\rho}{\rho} \propto R = \frac{1}{(1+z)} \tag{5}$$

provided $\Omega z \gg 1$. In this expression $\Delta\rho$ is the density enhancement relative to the mean background density ρ. In the limit $\Omega z \gg 1, R \propto t^{\frac{2}{3}}$ and the above relation describes the algebraic time development of the density perturbations. The problem is that the matter and radiation decoupled when the scale factor R was about $1/1000$ of its present value and so, since galaxies and clusters of galaxies exist now with $\Delta\rho/\rho \gg 1$, there must have been significant density perturbations in the matter distribution at the time when the photons of the Cosmic Background Radiation were last scattered. There should therefore be fluctuations in the intensity of the Cosmic Background Radiation in different directions on the sky. This is the line of reasoning which leads to

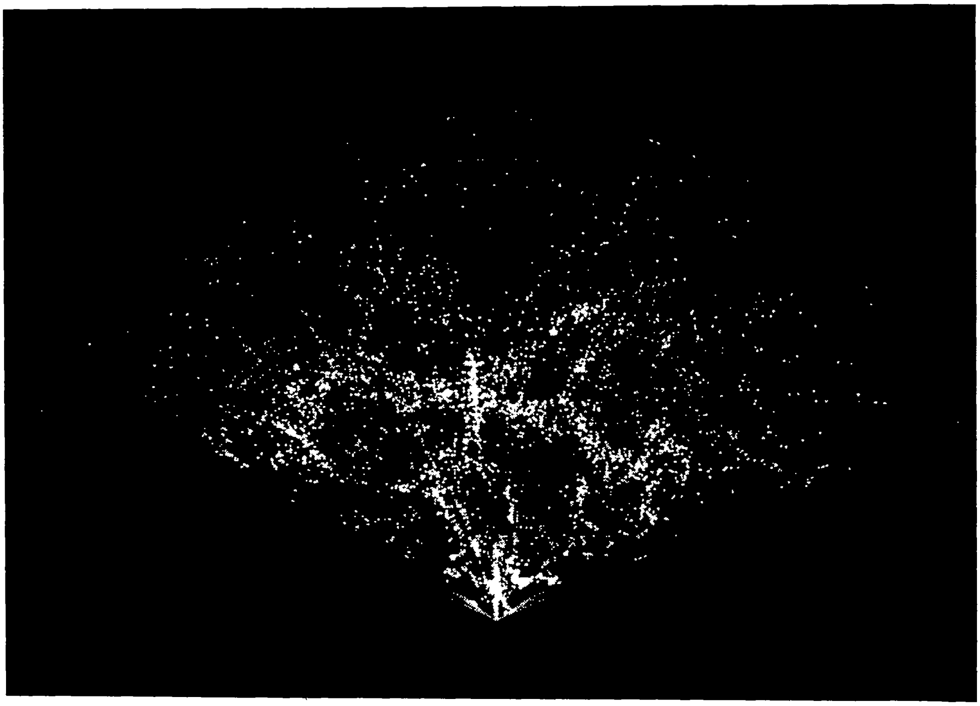


FIG. 21. The distribution of galaxies in the nearby Universe as derived from the Harvard–Smithsonian Center for Astrophysics survey of galaxies. This picture is a projection of the three-dimensional distribution of galaxies. Our Galaxy is located at the apex of the segments in which the distances of the galaxies have been measured. If the distribution of galaxies were uniform, the points would be distributed uniformly and at random within these segments. In fact, the distribution is grossly non-uniform with huge sheets, filaments and voids in the distribution of galaxies. (Courtesy of Margaret Geller, John Huchra and the Harvard–Smithsonian Center for Astrophysics.)

the difficulty of reconciling the extraordinary smoothness of the background radiation with the irregularity of the distribution of galaxies. If there existed only ordinary baryonic matter in the Universe now, the expected levels of fluctuation in the Cosmic Background Radiation would be much greater than the very low upper limits to the intensity fluctuations, $\delta I/I \leqslant 10^{-5}$. Although this poses a big problem for cosmologists, it is also a blessing because, were it not for the fact that the fluctuations grow so slowly, we would not have the possibility of asking questions about the nature of the initial fluctuations.

The most popular solution to this problem is to build the dark matter into the model universe so that it is dominant gravitationally. This means that all the interesting action by which the large scale structure of the Universe formed took place in some form of matter of which we are wholly ignorant. These fluctuations are assumed to be present in the dark matter at the moment when the matter and radiation decouple but these are not reflected in the distribution of the ordinary matter which is closely coupled to the background radiation as in the standard model. Although the nature of the dark matter is unknown, this is the current orthodoxy and the question is whether or not astrophysical arguments can cast more light on its nature.

The most popular picture is the *Cold Dark Matter* model in which the dark matter fluctuations on a wide range of scales collapse and begin to form galaxies by a process of hierarchical clustering. Large scale structures are all rather recent features of the Universe. A less popular version is the *Hot Dark Matter* model in which the neutrino has a finite rest mass. In this case, all fine scale structure is obliterated by the free streaming of the neutrinos and the large scale structures form first. Galaxies form by fragmentation of the largest scale structures into smaller-scale objects like clusters and galaxies. Neither picture gives exactly the right answer without some additional astrophysical considerations (Fig. 22). The cold dark matter picture can account for the correlation function for galaxies but has difficulty producing enough structure on the very largest physical scales. The hot dark matter picture produces too much large scale structure.

Recently, there has been great excitement because at last fluctuations in the Cosmic Microwave Background have been discovered by the COBE satellite at an intensity level about one part in 100000 on angular scales of 10° (Fig. 8). These angular scales correspond to physical scales in the Universe very much greater than the large holes in the distribution of galaxies. What has encouraged many theorists is that this is roughly the level of intensity fluctuations expected in the most popular variant of the Cold Dark Matter picture, if the initial spectrum of fluctuations is extrapolated in the most natural way to these large physical scales. This, however, is not the end of the story but rather the beginning. We are just about to enter the epoch when the study of fluctuations in the Cosmic Background Radiation becomes an astrophysical and cosmological discipline in its own right.

Whilst this work is very impressive, it is important to remember that there are many assumptions which have to be made. To list just a few of them:

1. It assumed that the Universe can be described by the critical model.
2. The spectrum of initial fluctuations has to be assumed. These are assumed to arise from quantum fluctuations in the early Universe, from phase transitions as symmetries are broken, from cosmic strings and so on.
3. The relation between the distribution of visible and dark matter has to be defined.
4. Some biassing of the galaxy formation process has to be assumed to obtain sufficiently clumped galaxies on an intermediate scale.
5. All the important action has taken place in the invisible, unknown dark matter – the Universe we all know and love was more or less an afterthought in the grand scheme of things.

I make this list not to criticize but to indicate that there are many areas of astrophysical study which are needed to underpin the assumptions made. But I have another worry and that is that the theorists have 'looked up the answer in the back of the book'. In other words, the input physics has been determined to a considerable extent by the need to get the correct answer in the end. Now, this is a perfectly respectable scientific procedure and saves a lot of time seeking solutions which are of no interest. However, it must be emphasized that the models of the origin of structure in the Universe are designed to give the right answer. The encouraging thing is that they give a flavour for the type of physics and constraints upon physical theory which can be obtained from studies of the large scale distribution of galaxies in the Universe.

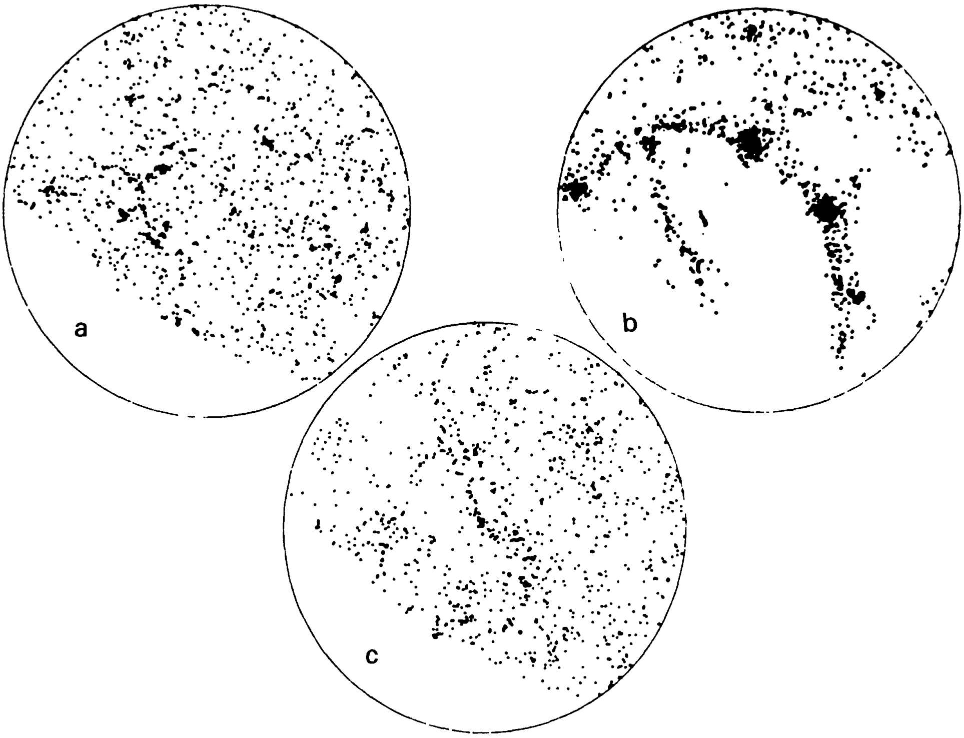


FIG. 22. Equal area projections of the galaxy distribution on the northern sky and in artificial catalogues made from n-body simulations. (a) The standard cold dark matter model including biased galaxy formation; (b) the neutrino-dominated model in which galaxy formation began at a redshift of 2·5; (c) the galaxies observed in the Harvard-Center for Astrophysics northern sky survey. The outer circle represents galactic latitude +40° and the empty regions lie at declinations below 0°. (From M.Davis, G.Efstathiou, C.S.Frenk and S.D.M.White (1992).)

7 CONCLUSIONS

I have described 9 facts about the large scale properties of the Universe which I believe are now secure. This represents remarkable progress since 1963 when there were only $2\frac{1}{2}$ facts. In my opinion, the best bits of the story of modern cosmology are those in which laboratory and cosmological physics come together to form a unique and convincing synthesis.

I have also described four major problems of cosmology and it is a matter of speculation, in my opinion, how far we will be able to understand the solution to them in convincing physical terms. I described five ways of approaching the problem, some of which are more physical than others. I will defend to the last the right of theorists to probe more deeply into the structure of the theories of particle physics and quantum gravity and the hope is that these studies will cast new light on these problems. I would not be surprised if this remained a difficult problem for some time.

My own position is that there is so much to be done in consolidating the observational and experimental foundations of the subject that I hope a very major effort can be made to advance the discipline of observational cosmology over the next 30 years. It would be a wonderful thing if the next 30 years were to contribute as many new real facts about the Universe as the last 30.

REFERENCES

Alcock, C., 1992. *From Protostars to Protogalaxies: The Neugebauer 60th Birthday Symposium*, ed. Soifer, T. ASP conference series, (in press).

Audouze, J., 1982. In *Astrophysical Cosmology*, p. 409, eds Bruck, H.A., Coyne, G.V. & Longair, M.S., Vatican City, Pontifica Academia, Scientiarum Scripta Varia.

Barrow, J.D. & Tipler, F.J., 1987. *The Anthropic Cosmological Principle*, Oxford, Clarendon Press.

Bondi, H., 1952. *Cosmology*, London, Cambridge University Press.

Bondi, H., 1989. *Let Newton Be*, p. 241, Oxford, Oxford University Press.

Boyle, B.J., Shanks, T., Fong, R. & Peterson, B.A., 1987. In *Observational Cosmology* eds Hewett, A., Burbidge, G.R. & Fang, L.Z., IAU Symposium No. 124, p. 643. Dordrecht: D.Reidel Publishing Co.

Branch, D., 1988. *The Extragalactic Distance Scale*, eds van den Bergh, S. & Pritchet, C.J., p. 146. San Francisco: Astron. Soc. Pacific.

Davis, M., Efstathiou, G., Frenk, C.S. & White, S.D.M., 1992. *Nature*, **356**, 489.

de Sitter, W., 1917. *Mon. Not. R. astr. Soc.*, **78**, 3.

Dunlop, J. & Peacock, J.A., 1990. *Mon. Not. R. astr. Soc.*, **247**, 19.

Einstein, A., 1917. *Sitz. Preuss. Akad. d. Wiss.*, **142**.

Ellis, R.G., 1992. *From Protostars to Protogalaxies: The Neugebauer 60th Birthday Symposium*, ed. Soifer, T.. ASP conference series, (in press).

Gamow, G., 1970. *My World Line*, New York, Viking Press.

Gott, J.R., 1987. *Observational Cosmology*, IAU Symposium No. 124, eds Hewett, A., Burbidge, G.R. & Fang, L.Z., p. 433, D. Reidel, Dordrecht.

Gribbin, J. & Rees, M.J., 1991. *Cosmic Coincidences: Dark Matter, Mankind and Anthropic Cosmology*, London, Black Swan.

Guth, A.H. & Steinhardt, P.J., 1989. In *The New Physics*, ed. Davies, P.C.W., p. 34. Cambridge, Cambridge University Press.

Hawking, S.W. & Ellis, G.F.R., 1973. *The Large Scale Structure of Space-Time*. Cambridge, Cambridge University Press.

Hesser, J.E., Harris, W.E., Vanden Berg, D.A., Allwright, J.W.B., Schott, P. & Stetson, P., 1989. *Publ. Astron. Soc. Pacific*, **99**, 739.

Jones, L.R., Fong, R., Shanks, T., Ellis, R.S. & Peterson, B.A., 1991. *Mon. Not. R. astr. Soc.*, **249**, 481.

Lilly, S.J. & Longair, M.S., 1984. *Mon. Not. R. astr. Soc.*, **211**, 833.

Longair, M.S., 1989. In *The New Physics*, p. 94, ed. Davies, P.C.W., Cambridge, Cambridge University Press.

Longair, M.S., 1992. *Theoretical Concepts in Physics*, p. 345, Cambridge, Cambridge University Press.

Mather, J.C., Cheng, E.S., Eplee, R.E. Jr., Isaacman, R.B., Meyer, S.S., Shafer, R.A., Weiss, R., Wright, E.L., Bennett, C.L., Boggess, N.W., Dwek, E., Gulkis, S., Hauser, M.G., Janssen, M., Kelsall, T., Lubin, P.M., Moseley, S.H. Jr., Murdock, T.L., Silverberg, R.F., Smoot, G.F. & Wilkinson, D.T., 1990. *Astrophys. J.*, **354**, L37.

McCrea, W.H., 1970. *Nature*, **228**, 21.

Metcalfe, N., Shanks, T., Fong, R. & Jones, L. R., 1991. *Mon. Not. R astr. Soc.*, **249**, 498.

Milne, E. A. & McCrea, W.H., 1934. *Q. J. Math.*, **5**, 73.

Panagia, N., Gilmozzi, R., Macchetto, F., Adorf, H-H. & Kirshner, R.P., 1991. *Astrophys. J.*, **380**, L23.

Peacock, J.A., 1992. In *The Nature of Compact Objects in Active Galactic Nuclei*, NATO Advanced Study Institute, Cambridge (in press).

Penzias, A. A. & Wilson, R. W., 1965. *Astrophys. J.*, **142**, 419.

Sandage, A.R., 1968. *Observatory*, **88**, 99.

Seldner, M., Siebars, B., Groth, E.J. & Peebles, P.J.E., 1977. *Astron. J.*, **82**, 249.

Smoot, G.F., Bennett, C.L., Kogut, A., Wright, E.L., Aymon, J., Boggess, N.W., Cheng, E.S., DeAmici, G., Gulkis, S., Hauser, M.G., Hinshaw, G., Lineweaver, C., Loewenstein, K., Jackson, P.D., Janssen, M., Kaita, E., Kelsall, T., Keegstra, P., Lubin, P., Mather, J.C., Meyer, S.S., Moseley, S.H., Murdock, T.L., Rokke, L., Silverberg, R.F., Tenorio, L., Weiss, R. & Wilkinson, D.T., 1992. *Astrophys. J. Letts*, **396**, L1.

Wall, J.V. & Peacock, J.A., 1985. *Mon. Not. R. astr. Soc.*, **216**, 182.