198

8

Did the Universe originate in a big bang?

(Some 'non-standard' views on cosmology)

JAYANT V. NARLIKAR

Modern cosmology is essentially a product of the twentieth century. On the observational side the subject was launched in 1929 by Edwin Hubble's discovery that the nebulae lying outside our Galaxy have redshifts that increase linearly with the nebular distances. Hubble's discovery was interpreted as a consequence of the expanding universe. The cause of expansion? A big explosion that caused particles of matter and radiation to move away from one another! The epoch of this explosion commonly called the 'big bang' is supposed to have marked the creation of the universe.

The notion of a big bang received observational support in 1965 when Arno Penzias and Robert Wilson found that the universe has a radiation background predominantly in the microwaves. With a presently estimated temperature of 3 K this background is believed to be the relic of the early post-big-bang era when the universe was extremely hot.

These observational results went hand in hand with theoretical developments. In 1915 Albert Einstein proposed his remarkable theory linking spacetime geometry to motion and gravity. Known as the general theory of relativity, this theory was used by him in 1917 to construct a simplified model of the large scale universe. Einstein's model was of a static universe and was therefore unable to account for Hubble's findings. However, Alexander Friedman in 1922 constructed expanding models that did anticipate Hubble's discovery.

The Friedman models imply that the universe has been expanding from a singular epoch in the past, an epoch when its space had zero volume and infinite curvature. The adjective 'singular' implies the breakdown of any sensible mathematical description. It was this epoch that later got the popular title of 'big bang'. If we associate a linear scale S with the universe, then in

Friedman's models the scale factor S has been increasing from its value zero at big bang. In the late 1940s, George Gamow argued that explosion implies a universe that was initially 'hot'. Gamow and his younger colleagues Ralph Alpher and Robert Herman tackled the physical problem of how such a hot universe played the role of a nuclear reactor for synthesizing atomic nuclei. One of their predictions was that there should be a relic radiation background in the microwaves surviving today.

It is interesting that both the major discoveries of 1929 and 1965 were anticipated theoretically. It is even more interesting that both theoretical predictions, of Friedman in 1922 and of Gamow *et al.* in 1950 were ignored at the time they were made. Certainly they played no role in inspiring the observations of Hubble and of Penzias and Wilson.

By hindsight the reason can be seen in the overall suspicion with which cosmology was viewed not only by physicists but even by astronomers. Cosmological theories were considered no more than speculative exercises without any possible testable status.

All that has changed dramatically in the last two decades! Not only is the big bang picture considered respectable, it is being used extensively by particle physicists, the high priests of science, as the testing ground for their programme of unification of physical theories. Working jointly, the particle physicists and cosmologists of today are far more daring than Gamow and his colleagues. While Gamow's discussion of the *early universe* during the first 1–200 seconds after the big bang was considered highly speculative four decades ago, the present work on the *very early universe* takes us to the era when the universe was only 10^{-37} second old!

Some reservations

While the cosmologists may rejoice at their subject being elevated to this high pedestal from the earlier status of a poor relation, viewed objectively this transformation has certain disturbing features. The marriage between cosmology and particle physics appears to be motivated by convenience rather than by appreciation of each other's virtues and weaknesses.

For, take the particle physicist's motive. His holy grail is the unification of all physical interactions. Einstein himself had felt deeply that there should be a single unified theory, although his lifelong attempts to deliver one did not succeed. Moreover, while Einstein during his life was in a small minority of physicists who believed in unification, today the situation is dramatically altered. The successful unification of electromagnetic and weak interactions has generated big momentum towards theories of 'grand' unification (GUTs) and 'super' symmetries (SUSY) in which eventually all laws of physics would be brought together.†

Unfortunately, the particle energies at which these theories can be dynamically tested are far higher (10^{15}–10^{17} GeV) than those achieved in manmade accelerators ($\sim 10^3$ GeV). Indeed this energy gap is too wide to be bridged by any foreseeable future technology. But then, theories which cannot ever be tested are no more than speculations, as any physicist will agree. Are GUTs and SUSY, with all their intellectual appeal, mere speculations?

The answer would have been 'yes' but for the big bang universe! It is here and only here that a brief era occurred in the past when the particle energies were so high that GUTs and SUSY had decisive roles to play. These theories can therefore be tested provided it is established that such a laboratory of the early universe ever existed. Hence the particle physicist has no alternative but to take the validity of the big bang for granted.

The arrival of the particle physicist on the scene has been timely from the cosmologist's point of view also. For, there are several questions concerned with the big bang cosmology that cannot be answered without inputs from particle physics. Why does the universe appear to contain predominantly matter rather than antimatter? Why are there so many (10^8–10^{10}) photons for every baryon in the universe? This question could presumably be answered by GUTs or SUSY by outlining the mode of creation of baryons.

There is another outstanding problem of big bang cosmology called the 'flatness' problem. This may be described as follows. The expanding space in a Friedman model can have either a uniform positive curvature, a uniform negative curvature or a zero curvature. The rate of expansion depends on this curvature property and the overall time scale associated with the universe is determined by the curvature mode in which the universe was initially set up. The mode of zero curvature (the flat mode) critically separates the other two. If the space has positive curvature it would eventually contract whereas if it has negative curvature it would quickly disperse to infinity. The adjective 'eventually' and 'quickly' are to be judged against any fundamental time scale associated with the universe. The only time scale that was available to the very early universe was the so-called Planck time

$$t_{\mathrm{P}} = \frac{\sqrt{(G\hbar)}}{c^5} \approx 5 \times 10^{-44}\,\mathrm{s}.$$

How in spite of such a short time scale did the universe manage to 'last' so

---

† It is an indication of how 'sell-oriented' science has become in today's consumer age that adjectives like 'grand' and 'super' are needed to qualify theories. Newton, Maxwell and Einstein who relied on the quality of their product than on such superlative adjectives would probably have fallen by the wayside in the race for funds.

long as its present age of $\sim 10^{10}$ years? For it to have lasted so long without contraction or dispersal to infinity much sooner, the universe has to be set up in the flat mode. How did this come about?

Then there is the 'horizon' problem arising from the fact that in a universe that has finite age, sufficiently remote parts have not had time to communicate with each other since the limit on communication speed is finite, the speed of light. At early epochs, the radius of the sphere of communication, the so-called particle horizon, is very small. For example, at $t = 10^{-37}$ s (when GUTs or SUSY operated) this radius was $\sim 3 \times 10^{-27}$ cm. Any relics of such an early era coming from regions well separated from one another should therefore show inhomogeneities. If the radiation background presently observed in the microwaves was generated very early in the universe how did it manage to acquire such homogeneity?

It is hoped that by understanding the physics of the very early universe we may be able to understand the answers to such questions. This is the motivation that drives the cosmologist to work with the particle physicist.

In a sense all the above problems may be considered relics of the very early epochs: for it is unlikely that these questions could be answered from the knowledge of the state of the universe at later epochs even if they were as early as the epochs considered by Gamow *et al.* (1–200 s) after the big bang.

There is one subtle difference between the early universe considered by Gamow and the very early universe discussed by the particle physicists and cosmologists today. To calculate the synthesis of nuclei during 1–200 s, Gamow was using laws of physics already tested in the laboratory. The reaction rates, decay rates, particle masses, the rules of statistical physics, etc. were all taken over from *known* physics. For this reason, the estimates of primordial abundances of light nuclei like helium provide a reliable check on the early universe scenario. In the case of the very early universe on the other hand, the particle physics being used has *not* been tested independently in the terrestrial laboratory; on the contrary, as emphasized earlier, it is supposed to be tested in the laboratory of the very early universe (which is itself under scrutiny!)

An analogy will help in understanding the logical status of the above procedure. Given the validity of Ohm's law, we can use it to calibrate resistances, ammeters or voltmeters. Or, given previously calibrated resistances, ammeters and voltmeters we can test Ohm's law. But we cannot do both at the same time. Likewise, if we have complete faith in the very early universe scenario we can test GUTs and SUSY; or we can take GUTs and SUSY for granted and discover what the very early universe was like. At best we may hope to arrive at a consistent picture in which a given particle theory provides a satisfactory explanation of present-day relics when used in a specific scenario of the very early universe. What we cannot claim is that our self consistent picture has *absolute* validity. It could very well be the case that *both* the particle physics and cosmology as depicted in the above solution are wrong!

Unfortunately this logical alternative is either forgotten or ignored in the definitive statements made about the origin of the universe. In the rest of this article I will discuss ways in which the above standard picture could go wrong at various levels.

The problem of the relics

An artist once exhibited a blank picture in an impressive frame. When asked what the picture was, he said that it was about a cow grazing. To the question as to where was the grass his answer was that the cow had finished it all. Where was the cow? The cow had left because it could no longer find grass there!

The scientific analogue of the above story is the following. Make a hypothesis H_1 to predict the existence of a certain relic R. If you don't find R, make another hypothesis H_2 to argue that after creation R could not have survived long enough to be seen. Thus the non-observation of R allows you (apparently!) to confirm the validity of two hypotheses H_1 and H_2.

We see examples of this type in the so-called relics of the very early universe. Some GUTs models predicted the existence of massive magnetic monopoles which are not only *not* seen today but whose existence (if detected) would be positively embarrassing for cosmology. For example, a typical monopole would have a mass of $\sim 10^{-8}$ g and the expected number density of such monopoles would be $\sim 10^{-6}\,\mathrm{cm}^{-3}$. Thus the mass density of monopoles of $\sim 10^{-14}\,\mathrm{g\,cm}^{-3}$ would be far higher than the *total* mass density of $\sim 10^{-29}\,\mathrm{g\,cm}^{-3}$ predicted for the universe by most big bang models! Moreover the existence of free monopoles would have played havoc with the magnetic field in our Galaxy and outside it. So as to get rid of the unwanted monopoles a new hypothesis is needed.

The existence of massive neutrinos is another example. GUTs and SUSY predict a variety of new particles of which neutrinos are expected to survive from the very early epochs, because they only weakly interact. However, massive neutrinos tend to lower the age of the universe (a problem I will come to later), introduce too much patchiness in the distribution of galaxies and possibly destroy the good agreement between the observed and calculated values of primordial helium. So numerous constraints are needed to ensure that most of the massive neutrinos do not survive for long.

The problem of monopoles and neutrinos has acquired additional interest because there are indications that the universe might possess non-luminous ('hidden') matter in a substantially greater measure than the luminous part in the form of galaxies. Could this matter be in the form of black holes or made up of exotic particles like photinos, gravitinos, axions etc?

It is certainly not beyond the combined ingenuity of particle physicists and cosmologists to produce scenarios in which the awkward monopoles do *not* get created, the massive neutrinos do *not* survive and the more exotic particles fulfill the necesssary requirements of hidden matter by remaining unseen. The overall self-consistency of this picture, as and when it eventually emerges, can at best be *a* plausible demonstration of how the universe began: it can never be a scientific proof that this was *the* way the universe began.

The best bet for a consistent scenario of the very early universe is offered at present by the so-called inflationary model. Here too, the original idea proposed in 1981 by A. Guth, though elegant and ingenious, did not work. It had to be replaced by the 'new' inflationary model which had its own crop of difficulties. One of the difficulties currently faced by the inflationary models is the predicted very large density fluctuation in the distribution of matter and radiation as seen today in the universe. The observed small scale smoothness of the microwave background clearly contradicts this prediction. Here too, the difficulty may be eliminated in an inflationary model, Mark n ($n > 2$), by exploiting some hitherto unexplored parameters, of a particularly exotic particle theory.

I think the present day cosmologists should pause now and then to ask themselves if they are going the same way that their ancient Greek counterparts did two millennia ago. Are the parameter fitting exercises needed to prop up a given scenario for the very early universe any different from the epicycles of Hipparchus and Ptolemy?

I end this section by discussing a difficulty of the inflationary model that will not go away by any amount of juggling with the parameters of GUTs and SUSY. The model, while 'resolving' the flatness problem also predicts that the age of the universe must be given by $\frac{2}{3}H_0$, where H_0 is the present value of the Hubble constant. The Hubble constant is measured observationally, as the ratio of the redshift (z) of a relatatively nearby galaxy multiplied by the speed of light (c) to its distance (D):

$$H_0 = \frac{cz}{D}.$$

The values of H_0^{-1} are considerably uncertain, lying in the estimated range of $10^{10} - 2 \times 10^{10}$ yr. There are observational reasons in favour of both values but the 'middle of the way' consensus among the astronomers today is to take the value at $\sim\frac{4}{3} \times 10^{10}$ yr.

Corresponding to this value the age of the universe is $\sim 9 \times 10^9$ yr. This value is too small to accommodate the age of the Galaxy and in particular the ages of globular clusters. These ages are variously estimated at (10–18) × 10^9 yr. Even taking Hubble's constant at the lowest value in the above range, the predicted age of the universe is not able to exceed the estimated ages of various astronomical systems in it.

This difficulty is usually pushed under the rug by taking $H_0^{-1} = 2 \times 10^{10}$ yr and stating that this is the 'Hubble age' of the universe. Nothing can be more misleading than this statement. If one is to take the current band-wagon of the very early universe seriously, then there is no escape from the conclusion that the 'true age' of the universe is two thirds of the 'Hubble age' and that it falls woefully short of other astronomical ages.

Quantum cosmology

Maxwell's electromagnetic theory coupled with the Lorentz–Einstein inputs of special relativistic electrodynamics admirably describes macroscopic phenomena. Yet the theory is found to be inadequate for microscopic purposes like, for example, the study of the simplest atom, that of hydrogen.

Classical electrodynamics tells us that the hydrogen atom should have a very short term existence. For, an electron circling a proton radiates continually and this energy loss brings it closer and closer to the proton until it merges with it. The time scale for this phenomenon is none other than the classical electromagnetic time scale $e^2/mc^3 \sim 10^{-23}$ s where e and m are the charge and mass of the electron. That the hydrogen atom is of a stable nature and that whenever it radiates it does so in discrete packets of energy rather than in a continuous manner are sufficient to tell us that the classical description is inadequate.

Quantum theory solves this problem in a satisfactory way. Given the Planck's constant, the mass of the electron and the electronic charge, we can construct a length scale $\hbar^2/me^2 \sim 10^{-8}$ cm, at which quantum ideas become relevant. Detailed atomic theory then tells us that this is the characteristic size of the atom and that the discreteness is the result of stationary states.

The existence of stationary states could be discovered even without going into the full details of quantum theory by simply concentrating on the quantization of the radial separation r between the electron and the proton. Such an approach does not tell us about the numerous stationary states due to the angular momentum quantum numbers. But it gives us the crucial information that the 'singular' fate of the classical H-atom arising from $r \rightarrow 0$ is avoided by its quantum counterpart.

All this does have relevance to cosmology! Like the Maxwell–Lorentz–Einstein electrodynamics the general theory of relativity is also a successful theory of gravity at the classical level. It has, of course, not been tested in the situations where gravity is very strong and concepts such as the 'black hole' and the 'big bang' are logical extrapolations of the theory into untested domains.

While physics in general would not progress but for such extrapolations, the example of the classical H-atom warns us to be cautious. In particular, can we trust the conclusion of classical general relativity that the big bang origin of the universe is inevitable? Can we trust the theory at time scales less than the Planck time t_P mentioned earlier? Was there a big bang at all, according to quantum theory?

Obviously the quantum theory of gravity holds the key to this important mystery. Unfortunately, in spite of many continuing attempts, a formally satisfactory and at the same time practically workable theory of quantum gravity still remains unattainable. Can we, however, expect to capture the flavour of the quantum inputs by a less ambitious but mathematically more manageable approach?

Such an approach is fortunately available. Just as the essence of the problem (but not the full details) could be captured by quantizing r for the H-atom, we can similarly try to quantize the scale factor S of the expanding universe: for, the classical big bang singularity is obtained by letting S go to zero.

While S can be quantized, a more convenient way (and also physically a more satisfactory one) is offered by the method of *conformal quantization*. According to this picture, from any classical spacetime geometry that satisfies Einstein's equations we can generate new geometries by arbitrarily scaling all spacetime intervals at every point by a factor Ω. This factor, known as the conformal factor, could change from one spacetime point to another. By quantizing Ω we therefore obtain a fairly general description of how scale fluctuations of spacetime occur around the Planck epoch.

Recent work by myself and T. Padmanabhan shows that such quantum fluctuations describe a host of spacetime geometries, the vast majority of which are *not* singular. In other words, the classical big bang turns out to be more the exception than the rule that characterizes the pre-Planck-time universe. The quantum mechanical probability that the universe arose from a big bang is almost zero. Thus the notion of an 'origin' a finite time ago is almost ruled out. Instead we have a universe that exists for ever and whose behaviour is determined largely by classical gravity. The exceptions arise when it happens to pass through highly compact states when quantum gravity comes into play. In the quantum regime there will be transitions of the universe from one state to another until it again expands and emerges in a classical state.

The removal of the concept of 'origin' also removes the horizon problem since there is now no limit on the range of communication. Likewise, if we compute the probability that through quantum conformal fluctuations the universe got into a Friedman-like model from an initial empty (vacuum) state then we find that the flat Friedman mode is overwhelmingly preferred. Thus the universe is without a singular origin and without the problems of horizon and flatness that beset the classical big bang models.

So it appears that there is a prima-facie case for quantum cosmology significantly altering the conclusions of classical cosmology. If the problems of singularity, horizon and flatness are resolved in the quantum era, then much of the motivation for the inflationary phase occurring later disappears.

Is the microwave background a relic of the big bang?

I next turn to another non-standard concept that questions the main evidence for the hot big bang. The concept has roots in the question posed above.

The main expected signatures of the microwave background as relic radiation were the following:

(i) Its spectrum should be Planckian.
(ii) It should show small scale fluctuations that relate to the era of galaxy formation.
(iii) Its energy density should be deduced from the physical conditions prevailing soon after the big bang.

Had these signatures been confirmed there would have been no difficulty in accepting the microwave background as the relic of the big bang. In reality the situation has been otherwise: only the first signature appears to be confirmed.

The extraordinary smoothness of the radiation background poses problems for scenarios of galaxy formation. If galaxies formed after the radiation background was made, why did the process of their formation leave no apparent marks on the background? Here again, after the realization that the original and more natural theory of galaxy formation through an adiabatic process leaves too large a fluctuation on the background, the less persuasive isothermal process was invoked. The latter leaves so small a fluctuation of temperature that it cannot be detected in the foreseeable future. This certainly saves the scenario, but at the same time it deprives it of predictive vulnerability that is the hallmark of a good scientific theory (this is another example of the cow grazing scenario!)

The energy density of the radiation, at the present temperature of 3 K is $\sim 6 \times 10^{-13}\,\mathrm{erg\,cm^{-3}}$. Why is the present temperature $\sim 3\,\mathrm{K}$? Why not 1 K or 10 K? Clearly, according to the relic radiation hypothesis the answer to this question must lie in the history of the very early universe. So far this answer has not come from the current ideas in big bang cosmology.

This is where one is tempted to look elsewhere for the origin of this background. Could it have been of a much recent origin, having nothing to do with a hot big bang? There are several astrophysical processes (of no connection with cosmology) that do produce energy densities of this order. To name a few, the galactic magnetic field, the cosmic rays and starlight all produce energy densities of the same order. With regard to the starlight, it was pointed out in 1968 by F. Hoyle, N.C. Wickramasinghe and V.C. Reddish that if all helium observed in the universe were made in stars, the starlight so generated would have the same energy density as the microwave background.

This last coincidence suggests a stellar origin of this background provided it can be thermalized subsequently. M.J. Rees proposed in 1978 that Population III stars at the epoch of $\sim 10^7$–10^8 yr could process the helium. Even earlier, in 1974–75 Wickramasinghe and others (including this author) had proposed that even ordinary starlight in more recent epochs would be efficiently thermalized by long slender grains of graphite. The details of such a process were studied extensively by N.C. Rana in 1980–81, who found that a plausible model of the microwave background can be made up within the observational constraints.

It is early days yet to assess such attempts, but their value as alternatives to the hot big bang scenario (since it is no longer as attractive as before) cannot be discounted.

How universal is Hubble's law?

I have discussed alternatives to the standard hot big bang cosmology, in an order that is progressively more radical. The last in this series questions the very basis of the idea of the expanding universe. As is well known, the redshift/distance relation of Hubble applies to galaxies over a wide range of apparent brightness. But, the notion of the expanding universe implies that Hubble's law must apply to *all* extragalactic objects.

We may express the above requirement thus. Given an extragalactic object at distance D, its redshift should ideally be given by a *unique* relation of the kind

$$z = f\left(\frac{DH_0}{c}\right)$$

where f is a function determined by the cosmological model. For small distances $f(x) \sim x$ and the above formula reduces to the linear relation first found by Hubble. However, *whatever* be the cosmological model, all objects at the same distance must have the same redshifts.

In practice small departures from this idealized situation are permitted, largely because galaxies in a cluster may have random motions of the order of $\leqslant 1000\,\mathrm{km\,s^{-1}}$. Translated into spectral shifts by the Doppler formula, these motions may generate departures in z as given by the Hubble law, of the order of $\Delta z \sim 0.003$. Indeed historically speaking, the very first example of extragalactic spectral shift to be investigated was one of blueshift!

Barring such small fluctuations therefore we expect all extragalactic objects to obey the above generalized Hubble relation. Do they?

Over the last two decades there has been a steady accumulation of data that seem to cast doubts on the universality of the Hubble relation. The data have been collected by many observers but by far the lion's share comes from the observations of H.C. Arp. The data are of the following kind.

(i) There are pairs or larger groups of quasars that appear to be near neighbours in space but with individual members having very different redshifts.
(ii) There are quasar-galaxy associations wherein typically a high redshift quasar is found near a low redshift galaxy.
(iii) There are galaxy–galaxy associations wherein typically a companion galaxy with larger redshift (vastly exceeding $\Delta z \simeq 0.003$) appears to be dominated by a main large galaxy of lower redshift.

It should be mentioned that all these cases are at present highly controversial, the dispute arising because we do not know the distances of the objects. There is no independent way of measuring distances, especially of the quasars concerned. Physical nearness of two quasars or of a quasar and a galaxy is usually argued on the basis of their projected nearness on the sky. Statistical arguments are needed to decide whether two objects at vastly different distances from us would happen by chance to be projected very near each other as seen by us. If the probability of chance projection is very low (say $< 10^{-2}$) then there is reason to suspect that the objects in question are physically near each other. The controversy is centred round the way the probabilities are computed.

These statistical arguments apart, Arp has produced in case (iii) examples of filamentary structures linking galaxies with discrepant redshifts. Are these structures proofs of physical connection between the galaxies? It seems hard to discount such evidence as 'projection effect' or 'photographic artefacts'.

Quasars being highly unusual objects might have other causes for their redshifts; but to argue that Hubble's law does not hold for galaxies would shake the very foundations of modern cosmology.

Conclusions

Nonstandard cosmology, by definition, includes ideas not conforming to the standard hot big bang model and as such it includes much more than what is presented here. For example, it is possible to work in the frame of other gravity theories, to assume that fundamental constants are changing or to propose completely new laws of physics.

To avoid entering into an open-ended field I have concentrated on motivations provided by observational cosmology. The aim here was not to offer a cut and dried alternative to the standard hot big bang but to emphasize that the standard model by no means provides the last word on the origin of our universe.