by VIRGINIA TRIMBLE

And How We Know That. . . Part III

COBE

& How We Know that the Universe Went through a Hot, Dense Phase

It was a dark and stormy night at the center of the Universe. But this need not concern us, as our story does not take place at the center of the Universe. In fact, the Universe hasn't really got a center, and coming to terms with this is an important part of what is and is not meant by Big Bang.

THE CURRENT "STANDARD MODEL" of cosmology, or hot big bang, is naive and rather proud of it. We find that general relativity is a good description of all the deviations from Newtonian behavior that we see? OK, then the correct (non-quantum) theory of gravity is GR or something very much like it. The observations of wavelength shifts in the spectra of distant galaxies suggest homogeneous, isotropic expansion?
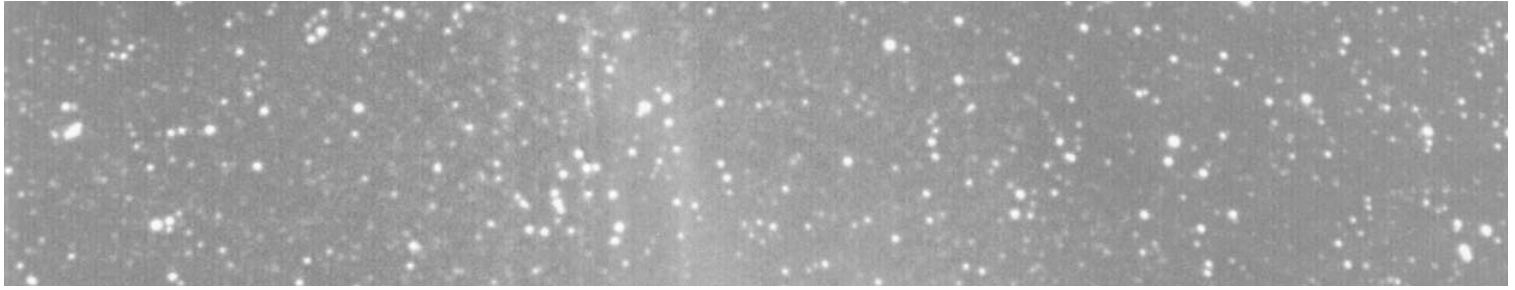
Fine, then the universe is expanding, cooling, and becoming more tenuous. The rate of that expansion, the oldest stars we see, and radioactive elements all suggest time scales of 10–20 billion years? Great! Then we know an approximate age for the expanding universe. Running the expansion backwards implies a past hot, dense state in which an isotropic sea of electromagnetic radiation would naturally be produced and about a quarter of the matter turned into helium? Even better. The radiation and helium we see are then relics from that past and can tell us about its details.

The vast majority of astronomers, including the present author, work happily within this framework, whether we focus on better determinations of the parameters of the standard model and trying to form structure within it or on phenomena like nova explosions, comets, and peculiar A stars, for which the large scale structure of the universe is a sort of Muzak. This is not quite the same as saying we all find standard, general relativistic cosmology intuitively obvious. Everybody is occasionally tempted to think in terms of matter expanding from a point or small region into previously existing space. This is the wrong image. The space itself is expanding and carrying the matter with it. The redshifts we see are not Doppler shifts, caused by relative motion through space, but are rather the stretching out of wavelengths with the metric they propagate on. Thus, whether the universe is finite or infinite, it has no edges and no center, and the only unique point is the time $t = 0$, when the expansion started. This instant is sometimes called the Big Bang, but I think this is a bit misleading.

"Big Bang" was once an insult, thrown out over the airwaves of BBC by Fred Hoyle, then (1950) as now a propounder and advocate of the most robust of the alternative pictures, called Steady State. In it, GR is not the right theory of gravity, the universe is not becoming more tenuous, nothing is a fossil of a hot, dense past state, because there never was one, and the universe is infinitely old. In accordance with common, though not universal, usage, I shall take Big Bang to mean simply the four naive conclusions of the first paragraph.

Carl Wilhelm Wirtz (1876–1939), one of several astronomers who preceded Hubble in attempting to correlate redshifts of spiral nebulae with their apparent brightnesses or angular diameters (both, in principle, distance indicators). Wirtz seems to have been the first to recognize that the correlation would be cleaner if some non-distance-dependent way could be found to identify classes of physically similar objects. He also contributed to the measurements of surface brightnesses of galaxies, of the galaxy luminosity function, and of the correlation between galaxy brightnesses and their location in clusters. (Courtesy of Helmar Duerbeck, Astronomical Institute, University of Muenster and Astronomische Gesellschaft.)

Evolutionary cosmology is a less inflamatory, but also less vivid synonym, and a recent contest to find an alternative name produced nothing better than Calvin and Hobbes' "tremendous space kablooie."

My remaining pages in this issue are devoted to outlining how the community came to agree on the standard model, why the alternatives are rejected, how the results from COBE on the spectrum and fluctuations of the leftover radiation strengthen the case for a hot big bang and help with pinning down some of the remaining uncertainties, and what came before the Big Bang.

ONCE UPON A TIME

Theory and observation got off to a false start together. Einstein published his first, erroneous theory of gravitation in the same year (1912) that Vesto Melvin Slipher exposed the first spectrogram on which the velocity of a spiral nebula could be measured. Slipher found a blue shift of about 300 km/sec. This is actually the right value, but it is the vector sum of the rotation of our own galaxy and the mutual orbital motion of the Milky Way
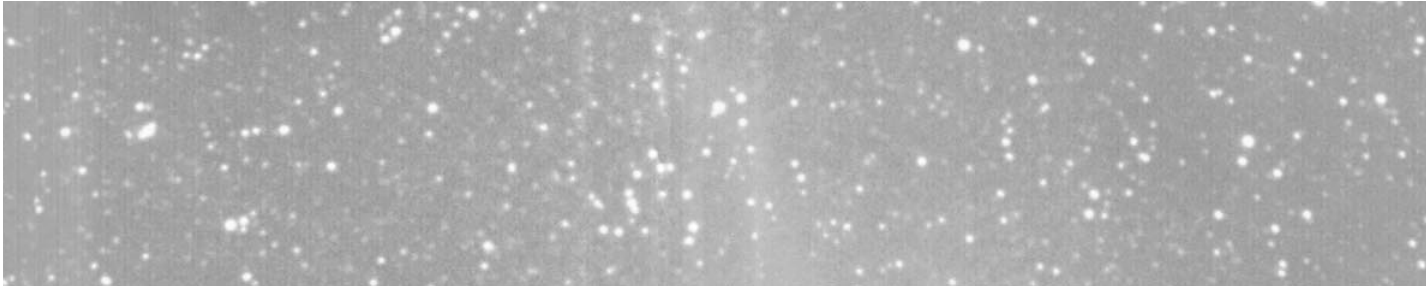
and the galaxy he looked at, our nearest large neighbor, the Andromeda galaxy. This large blue shift has very little to do with cosmology, except that it implies a sizable dark matter component associated with the two galaxies.

During the next decade, Einstein put forward the equations of GR as we now know them (in 1916) and his static solution, including the infamous cosmological constant, in 1917. That same year, Willem de Sitter proposed a different solution of Einstein's equations, in which space is static (and empty), but test particles and photons will show a preponderance of red shifts over blue shifts, with roughly a quadratic relation between distance and redshift. The first decade of cosmology closed in 1922 with Alexander Friedmann's imperfectly derived, but correct and complete, set of solutions to the GR equations (in which the universe could expand or contract uniformly, or oscillate, and have either finite or infinite total volume), and with Slipher's accumulation of enough spectrograms of spiral nebulae to show that redshifts were much commoner than blue shifts and that some were as large as 1000 km/sec. The expanding Friedmann solutions, in contrast to the de Sitter one, predicted a linear correlation of redshift with distance, at least locally.

The community was then still bitterly divided over whether the spiral nebulae were merely gaseous structures within our own galaxy or separate, roughly comparable, structures. The arguments (laid out in the Curtis-Shapley debate just 75 years ago) were tangled intimately with the question of distance scales inside the Milky Way and are too numerous to do justice to here. Edwin Hubble forced the "separate but equal" solution in 1924 by finding Cepheid variable stars and other reliable distance indicators in Andromeda and other nearby galaxies.
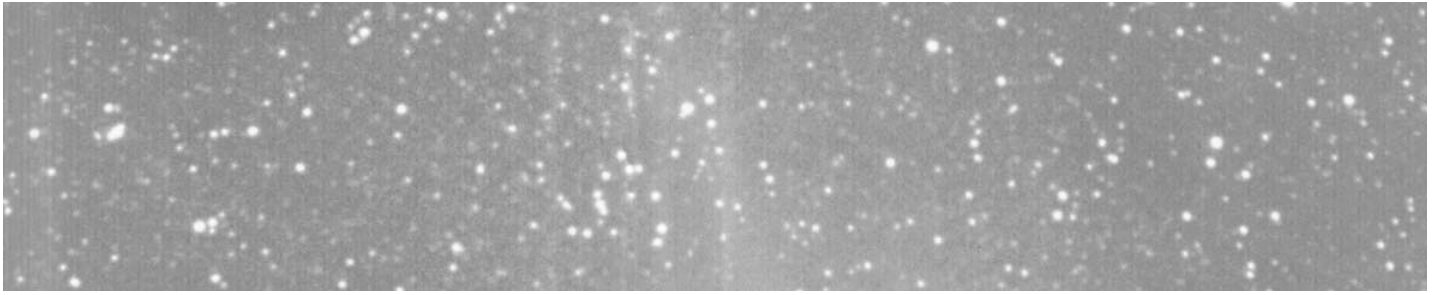
He did not immediately plot a Hubble diagram. Rather, between 1922 and 1926, attempts to correlate measured wavelength shifts with apparent sizes or brightnesses of spirals were made by (at least) the Swede Knut Lundmark, Germans Gustaf Strömberg and Carl Wirtz, Polish-American Ludwik Silberstein, and American

### Radial Velocities
of 25 Spiral Nebulae

Nebula	Velocity km (sec^{-1})	Nebula	Velocity km (sec^{-1})
N.G.C. 221	-300	N.G.C. 4526	+580
224	-300	4565	+1,100
598	-260	4594	+1,100
1023	+300	4649	+1,090
1068	+1,100	4736	+290
2683	+400	4826	+150
3031	-30	5005	+900
3115	+600	5055	+450
3379	+780	5194	+270
3521	+730	5236	+500
3623	+800	5866	+650
3627	+650	7331	+500
4258	+500		

Velocities for the 25 spiral nebulae thus far observed. In the first column is the New General Catalogue number of the nebula and in the second the velocity. The plus sign denotes the nebula is receding, the minus sign that it is approaching. (From "A Spectrographic Investigation of Spiral Nebulae," by Vesto M. Slipher, Proceedings of the American Philosophical Society **56**, 403-409 [1917], reprinted in A Source Book in Astronomy and Astrophysics, 1900–1975, ed. Kenneth R. Lang and Owen Gingerich, Harvard University Press, 1979.)

"....It has for a long time been suggested that the spiral nebulae are stellar systems seen at great distances. This is the so-called 'island universe' theory, which regards our stellar system and the Milky Way as a great spiral nebula which we see from within. This theory, it seems to me, gains favor in the present observations...."
—Op. Cit.

The Hubble Diagram
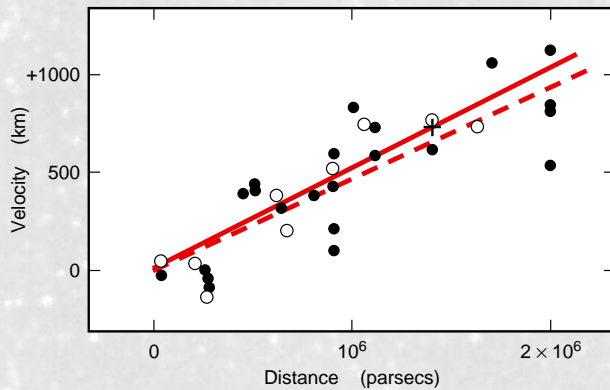
The Hubble diagram as first plotted by Hubble. He used apparent brightnesses of whole galaxies as his primary distance indicators (and his distance scale differed by a factor of 5–10 from the present one, in the sense of being too small). Notice that the vertical coordinate is incorrectly labeled as kilometers, when it should be km/sec. His fit to the data yielded H = 536 km/sec/mpc with error bars of less than 10% (a mistake being made right down to the present time). [Figure adapted from "A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae," Edwin P. Hubble, Proceedings of the National Academy of Sciences **15**, 168–173 (1929)].

"....The results establish a roughly linear relation between velocities and distances among nebulae for which velocities have been previously published, and the relation appears to dominate the distribution of velocities...."

—Op. Cit.

Howard P. Robertson. The motivation was typically the de Sitter solution, and the results fell somewhere between a scatter diagram and the expected quadratic, which has still at least one living defender.

The winning diagram, presented by Hubble in 1929, was linear enough to persuade most who saw it. Slipher's redshifts had been augmented by the first few to be measured by Hubble and Milton Humason with the new 100-in. telescope at Mt. Wilson, but the main improvement was better distance indicators and more careful selection of a sample of galaxies of uniform type for study. Hubble himself oscillated several times in the following years between universal expansion and tired light as his preferred explanation of what was soon dubbed Hubble's Law.* His value for the proportionality constant implied a time scale for the universe, $1/H \approx 2$ Gyr, at any rate comparable with the age of the earth as then understood.

Secondary literature leaves the impression that respectable scientists did not take any of this very seriously for some time. In this context, the latter sections of Richard C. Tolman's **Relativity, Thermodynamics, and Cosmology** (published in 1934 by Oxford University Press) repay study. He is ready to prefer Friedmann models to static or de Sitter ones and to compare the measured mass density of our region of space (extending only to 10^8 light years) with the unique ones of some models. And he is already worried about the discrepancy between $1/H$ and the longer time scales of stellar evolution. But he does not consider extrapolating back in time or farther away than the observed volume of space.

The first great extrapolater was George Gamow, beginning with a 1935 paper in the *Ohio Journal of Science*, addressing the role of neutrons in nucleosynthesis. His

The linear relation is best written as ($\Delta\lambda/\lambda$) c = Hd, where the left side has units of velocity, but can exceed the speed of light, d is distance, and H is the Hubble constant. H has units of velocity/distance, or reciprocal time. In a Friedmann model, expansion has been going on for 1/H in an empty universe, less than that with matter present (e.g., 2/3H for the density that would just stop expansion in infinite time) and longer than 1/H for positive values of the cosmological constant.
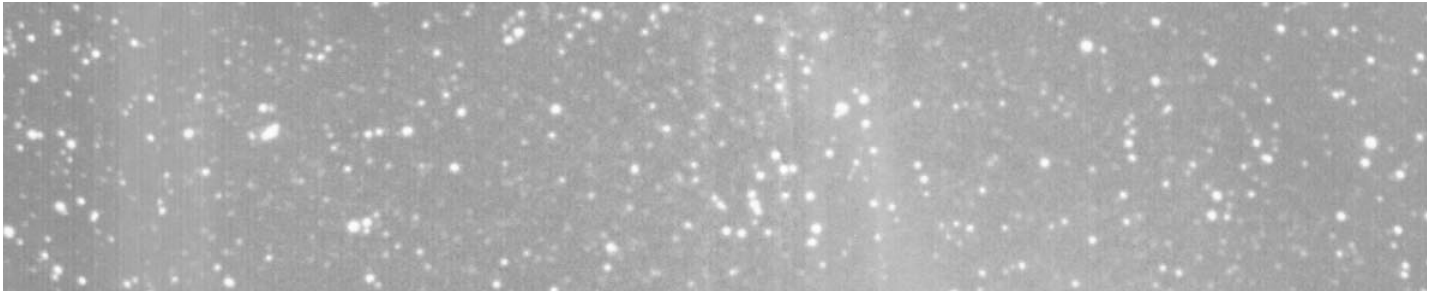
collaboration (1948–53) with Ralph Alpher and Robert Herman produced the first quantitative estimates of the current cosmic temperature and of cosmological nucleosynthesis. They are justly regarded as heroic pioneers. They began, however, from a false premise, with a pure neutron "ylem" at $t = 0$. Neutron decay was then blamed for additional heating as well as for protons and electrons. They had also a hopeless goal—the production of the full range of elements and nuclides in the early universe. Such progression from A = 1 (hydrogen) to A = 238 (uranium) is, of course, interrupted by the extreme instability of all possible A = 5 and 8 nuclides (and the gaps can be bridged only at higher density, by three-particle interactions, such as occur in evolved stars).

The first correct calculation of cosmological nucleosynthesis, starting with thermal equilibrium at early times, came from C. Hayashi in 1950. He also hoped one might still hop across the A = 5 and 8 holes with some suitable mix of protons, deuterons, tritons, helium nuclei, and so forth. Actually one does, a bit. Traces of lithium-7 are among the products of modern standard model calculations (and even a bit of beryllium and boron if one considers sufficiently inhomogeneous initial conditions). That you get about 25 percent helium out of either a Gamovian or a thermal equilibrium sort of hot big bang continues to strike me as slightly mysterious. Perhaps it says nothing more profound than that all nuclear energies are around 1 MeV per amu, but I would be interested in alternative perspectives from those who have thought about the issue.

The same 1948–53 period saw the flowering, largely in England, of the Steady State alternative model, in which new matter is continuously created to keep the density of an expanding universe constant. That American students of the time took it much more seriously than did their senior, professional mentors is perhaps mostly a tribute to the writing skills of its propounders, Hermann Bondi, Thomas Gold, and Fred Hoyle and early supporters, including Dennis Sciama, William H. McCrea, and Raymond A. Lyttleton. Apart from philosophical considerations, the primary motivation for
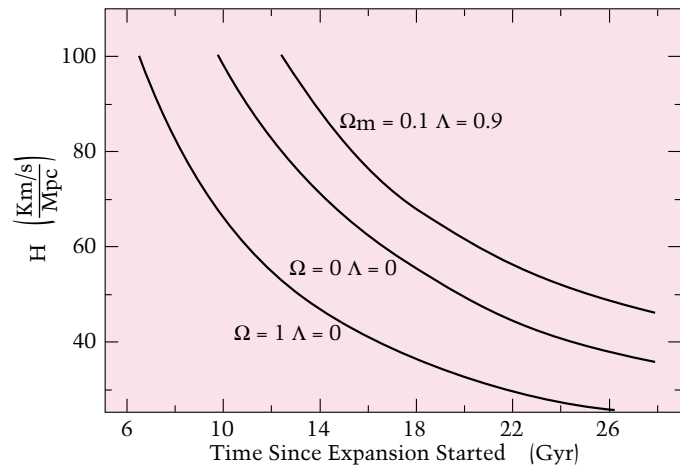
Time since the universe began expanding ("age of the universe") as a function of the correct, global Hubble constant, H, and other parameters whose values are also uncertain. The far left curve is for total density in matter (baryonic plus dark etc.) equal to the critical density and zero cosmological constant. The middle one is for density and cosmological constant both zero (or anyhow small), and the right one for flat space-time geometry with contributions from a matter density (10% of the critical density) and cosmological constant (90%).

steady state was the time-scale problem noted already by Tolman. This became acute when stellar evolution, based on nuclear energy sources, was put on a firm footing, but it quickly ameliorated after the 200-in. telescope was turned to cosmological problems (also starting in 1948) and began to provide evidence for larger distances in the universe and so longer expansion times.

Quite early on, Bondi asked a paleontological question—if the universe was different in the past, where are the fossils? Helium is clearly one such (if and only if there is no other way to account for its ubiquity). Radio galaxies and, soon after, quasars (much commoner at redshifts of 1–2 than here and now) were increasingly accepted as such fossils through the period 1955–65. And the definitive fossil, "A measurement of Excess Antenna Temperature at 4080 Mc/s," was announced in 1965 by Arno Penzias and Robert Wilson (as if you didn't
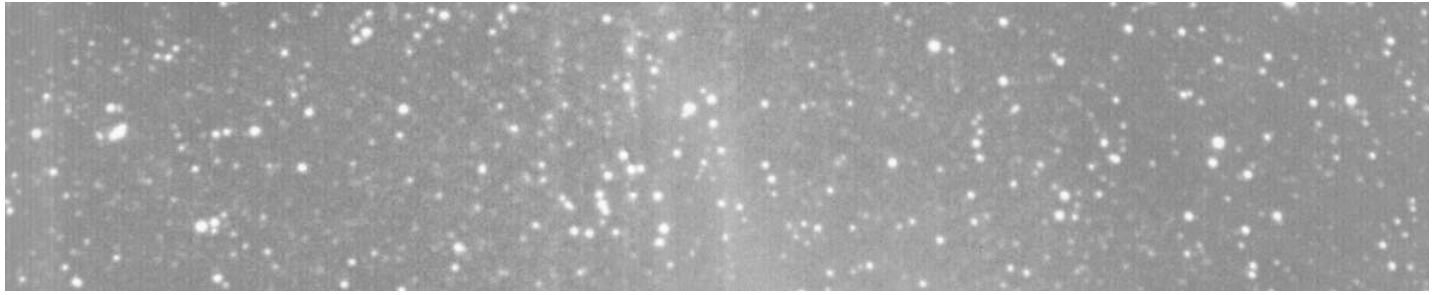
The Microwave Background Radiation

....Measurements of the effective zenith noise temperature of the 20-foot horn-reflector antenna (Crawford, Hogg, and Hunt 1961) at the Crawford Hill Laboratory, Holmdel, New Jersey, at 4,080 Mc/s have yielded a value about 3.5K° higher than expected. This excess temperature is, within the limits of our observations, isotropic, unpolarized, and free from seasonal variations (July 1964–April, 1965). A possible explanation for the observed excess noise temperature is one given by Dicke, Peebles, Roll, and Wilkinson (1965) in a companion letter this issue....

already know). George Gamow's reaction, expressed at the 1967 "Texas" symposium on relativistic astrophysics was, "Well, I lost a nickel, and you found one. Who is to say it's the same nickel?"

For what it is worth, we are now as far away in time from the Penzias and Wilson measurement as they were from the first nucleosynthesis paper by Gamow. And, despite periodic alarums and excursions, the status of big bang nucleosynthesis calculations remains that the observed amounts of ordinary hydrogen, deuterium, helium-3 and -4, and lithium-7 can be simultaneously produced in the standard model provided that the nuclear and particle physics is taken to be what we see in the laboratory and that the density of baryonic material is between 1 and 10 percent of the total density that would just halt expansion in infinite time ($\rho_c = 3H^2/8\pi G$).

RULING OUT THE ALTERNATIVES

Disciples of Sherlock Holmes (and perhaps Karl Popper, but of this I am not sure) require both that their explanations take care of all the facts and that all conceivable alternatives be ruled out. The previous section addressed the positive side of the hot big bang. Here we look briefly at the flaws of some counter-proposals, though not all of them, as it is already getting on for tea time.

Tired light is the idea that photons lose energy and so are shifted to longer wavelengths simply by traveling long distances. Admittedly, the Feynman graphs for this sum to zero. But it is remarkably difficult to rule out on purely observational grounds. We are still a few orders of magnitude away from being able to see the expected redshift over laboratory distances. (Mössbauer spectroscopy seems to come closest.) The surface brightnesses of distant galaxies as a function of redshift are different in expanding and tired-light cosmologies, but the difference is easily lost in observational uncertainties. The cleanest test comes from the behavior of time intervals. In relativistic expansion, all of them are stretched out, just like the time between wave crests of radiation. In tired light, time intervals are the same for all observers. A couple of supernovae have now been seen at sufficiently large redshifts (0.3 and 0.5) that the time dilation of their light curves should show up. It probably does, though some of the data remain unpublished.

What about a sort of Newtonian or Galilean universe, in which matter explodes outward from a point or small region into previously existing space? Such a universe has a hot, dense phase in its past to make helium and the microwave background radiation. And, of course, after a while velocities and distances will be linearly related. But we must be remarkably close to that central point or region to see as nearly an isotropic universe as we do. This was a sort of 99 percent conclusion even when all we had to go on was the Hubble diagram in different directions. It becomes more powerful with the inclusion of 3K radiation measurements. Even if you blame off-center location for the dipole anisotropy (normally attributed to ordinary motion of our galaxy), we are still restricted to the central 0.1 percent or so.
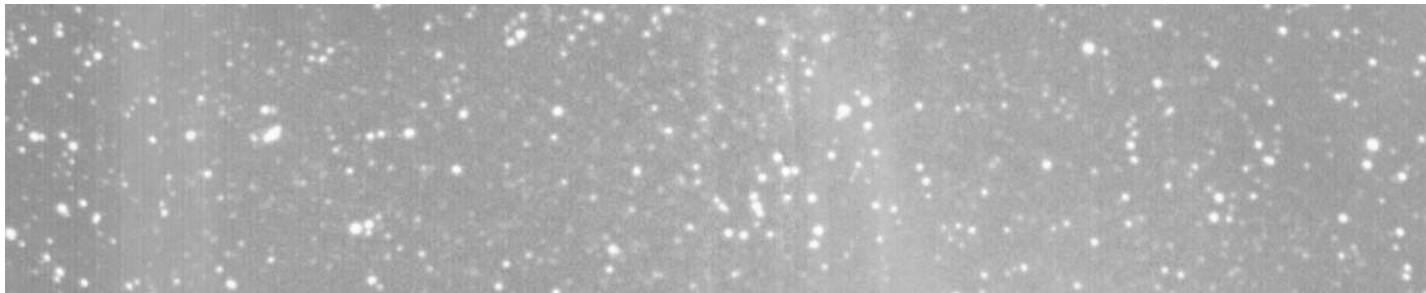
Friedmann's original set of solutions to Einstein's equations included some that oscillate between expanding and contracting phases. Such a universe could contain objects (very sturdy ones, anyhow) older than $1/H$. Whole conferences have worried about whether there are thermodynamic objections to this. But it doesn't matter. If you accept the general relativity that suggested these solutions in the first place, then they are impossible.

Theorems dating from the 1960s, due to Stephen W. Hawking, G.F.R. Ellis, Roger W. Penrose, W. Israel, B. Carter, and others, establish that, first, if a system once gets into a singular state from non-singular conditions (goes through a collapse phase), it can never get out again, and, second, that we have a singular state (or at least a trapped surface, which is closely related) in our past, and also in our future if the density of the universe exceeds the critical density. Thus a closed universe can expand and contract once, but only once. This conclusion is about the same age as the 3K radiation, but it has taken much longer to find its way into textbooks.

Finally, let us deal collectively with steady state, its modifications, and any other scenarios that do not provide a hot, dense phase 10 or 20 billion years ago. First, unless matter is somehow created as 3/4 hydrogen and 1/4 helium you must fuse protons (or protons and neutrons) into alpha particles. This happens in stars—and a good thing, too, since it keeps them shining. The ratio of helium production to shine is a laboratory number. The density of starlight in the present universe (including that absorbed and reradiated by dust as infrared) is reasonably well measured. And, putting the numbers together, you discover that real galaxies can turn only about 2 percent of their baryonic mass into helium in a Hubble time. Or, looking at it the other way around, if you want to start with pure hydrogen and no hot, dense stage, galaxies have to be 10 times as bright as the ones we see to reach 23 percent helium now.
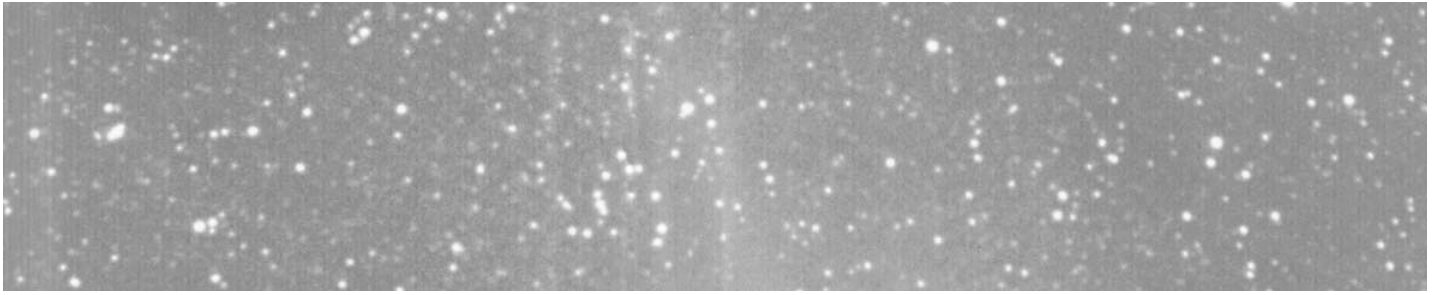
Second, you must come up with the microwave background radiation and make it both very accurately a black body and very accurately the same in all directions in space. A perfectly homogeneous, isotropic, isothermal big bang makes perfectly isotropic, thermal radiation; and we then have only to worry about not messing it up later (next section). But if you produce the radiation in sources distributed through time and space, you must isotropize and thermalize it. The energy requirements are not beyond possibility, being comparable with what you would get by turning another 25 percent of the hydrogen to helium instantaneously [though fusion spread over time won't quite do, since you lose energy density with redshift, z, in proportion as $(1 + z)^{-4}$]. But the energy from the hypothetical sources must be continuously absorbed and reradiated by something that is at 2.7K and blankets the sky. We have no independent evidence for such absorbing material, and strong evidence against it. Radio sources with redshifts of two and more are seen as bright points against the background at centimeter and millimeter wavelengths. The universe is not optically thick at the present time and cannot be continuously isotropizing and reradiating energy, whatever class of sources you care to postulate.

COBE AND THE PROBLEM OF STRUCTURE FORMATION

Can cosmologists close up shop and go home? Not quite. While our standard model does a great job with things (expansion, nucleosynthesis, background radiation) that are homogeneous and isotropic, it has no galaxies, clusters, or other structure. Such a universe has only individual atoms for observers, and we would not be here to talk about it. Structure must form early enough to account for the quasars we see back nearly to a redshift of five, and it must form without ruffling up the passing $2.7(1 + z)$K radiation.

Structure formation is arguably the single most important outstanding problem in modern astrophysics. It is genuinely insoluable with the most obvious choice of initial conditions—baryonic material (only) at the density implied by nucleosynthesis, growing its structures from adiabatic perturbations (ones in which matter

$-150\,\mu K$ ▮▮▮▮ $+150\,\mu K$
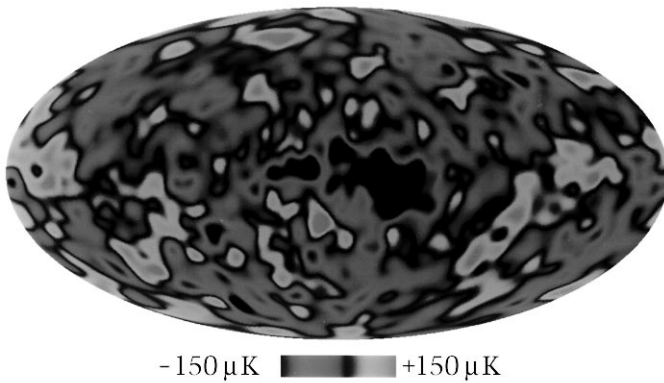
Residual fluctuations in the nearly-isotropic microwave background, after contributions from the solar system, galaxy, and ordinary Doppler shifts have been subtracted. The data are now accurate enough that the experiment group can say with confidence that particular bits of the sky are warmer or cooler than average, based on looking separately at two years of data. (Courtesy George Smoot, LBL.)

and radiation remain coupled until the universe is too cool to keep hydrogen ionized and becomes transparent).

To get density perturbations of order unity now (so they grow non-linearly into galaxies), you must have at least $\Delta p/p = 10^{-3}$ at $z = 1000$, when baryons and photons part company. In an adiabatic lump (where photon number density goes as T^3, and $3^2 = 10$ in astronomical convention), the associated temperature lumps must be close to $\Delta T/T = 3 \times 10^{-4}$. And these will still be with us. But limits in data obtained from the ground were already below 10^{-4} and shrinking when COBE took off. Measured values, at several wavelengths and on several angular scales, are now all in the ball park of 1–2×10^{-5}.

This conflict has driven most studies of large-scale structure formation for the last decade and more. Fruits of the struggle include models with many different kinds of non-baryonic dark matter, non-adiabatic (plus non-Gaussian, and non-scale-invariant) perturbations as starting conditions, and cosmic strings and other seeds for galaxies to condense around. The point of

non-baryonic dark matter is that it stops speaking to photons electromagnetically long before $z = 1000$, so lumps can form without dragging radiation with them. But gravitational communication remains between radiation and any kind of matter. Thus you can buy only about one additional order of magnitude between the necessary $\Delta p/p$ and the minimum resulting $\Delta T/T$.

This is, however, all we need, since two years of data from the Diffuse Microwave Radiometer on the COsmic Background Explorer, as well as data sets coming from ground-based observatories from California to the South Pole, now show brightness fluctuations in the background radiation at the "predicted" lower level. COBE and one ground-based observation even find the same warm and cool patches on the sky.

A couple of other happy conclusions have transpired. The relative amplitudes of the fluctuations on different angular scales are, at least, not inconsistent with most people's favorite guess (equal power on all scales, otherwise known as a Harrison-Zeldovich spectrum). The shape of the radiation spectrum is as precisely a black body as anything ever measured, and severely limits any stray energy inputs from, e.g., helium synthesis in pre-galactic stars.

Third, the same fluctuations are seen at several wavelengths, with relative amplitudes that are just the derivatives of a black body spectrum that you expect from "colorless" Doppler shifts and gravitational redshifts due to small variations around isotropic expansion of a big bang universe. In contrast, a "sum of sources" model has fluctuations due to slightly different source numbers and reradiation temperatures in different parts of the sky, and the thermal nature of the fluctuations would be a complete coincidence. This last point seems to have been widely appreciated only within the last six months or so. I first heard it in a talk by Martin Rees at the August 1994 general assembly of the International Astronomical Union. It is a specific objection to recent, quasi-steady-state revivals.

All careful readers of the New York *Times* will remember a recent flurry of responses to a Hubble Space
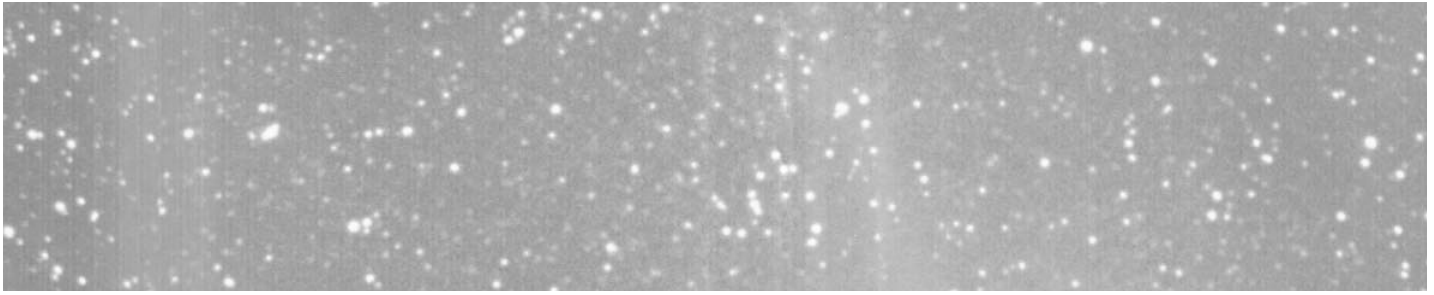
Telescope measurement of H = 80 km/sec/Mpc, based on the period-luminosity relation of Cepheid variable stars in one galaxy in the Virgo cluster (12–22 Mpc away). If you accept this as the correct global value, are sure that the cosmological constant is zero, and have confidence that the mass density of the universe is closer to the critical density than to zero, then the expansion time scale is necessarily rather less than 10 Gyr. Pushing along the evolution of galaxies and stars so that they look like the ones we see after so short a time presents certain difficulties. It is, however, probably obvious from my tone of ink that I have doubts about the need to accept the triple assumption that leads to the contradiction.

WHAT CAME BEFORE THE BIG BANG?

If you mean by Big Bang a state of exact thermal equilibrium, then all evidence of what come before that state will, by definition, have been wiped out, and this is a silly question. It would then deserve only a silly answer, along the lines of the response to "What was God doing before he created heaven and Earth?" "He was creating hell for people who ask questions like that."*

But we have just persuaded ourselves that the real universe was not precisely in such equilibrium. Rather, there were fluctuations in the density (of something) that have grown into the galaxies, clusters, and voids that we see.

Candidates for non-equilibrium entities which might, therefore, carry traces of what came before the hot, dense phase include many kinds of non-baryonic dark matter and assorted topological and non-topological singularities and solitons, left behind by symmetry breakings and phase transitions, including, of course, the particles and fields responsible for a possible pre-nucleosynthetic epoch of inflationary (exponential in time) expansion. I am not sufficiently knowledgeable to cherish any strong opinions about the existence or nature of these entities.

This exchange is widely attributed to St. Augustine. But more careful reading indicates that he really quoted it as an example of the kind of nonsense up with which he would not put. Bohr probably didn't really have a horse shoe over his barn, either.

Probably no one would claim that the evidence for and about them is anywhere near as strong as that for the hot, dense phase, of which helium and the thermal background radiation are fossils.

Further work is needed, as the old saying goes. But, meanwhile, us non-experts can respond to questions from the even less expert by saying, firmly, "Yes, there was a big bang! The evidence is overwhelming that the universe was very hot and very dense 10–20 billion years ago, and has been expanding and cooling every since."

What To Read Next

A good introduction to historical materials is N.S. Hetherington, Ed., **Encyclopedia of Cosmology**, Garland Publishing, NY, 1993.

A great place to start if you want to learn to do real calculations and evaluate real observations is P.J.E. Peebles, **Physical Cosmology**, 2nd edition, 1993.

Ralph Alpher and Robert Herman have told their story in *Phys. Today* **41**, 24 (1988).

Some recent words on nucleosynthesis and COBE results are, respectively, C.J. Copi et al., *Science* **267**, 192 (1995), and G. Smoot et al., *Astrophys. J.* **437**, 1 (1994).

The very latest word at any given instance must be sought at conferences and on your preprint shelves.

One view of what came before the Big Bang is expressed by A. Linde, "The Self Reproducing Universe," *Scientific American*, November 1994, p. 32.