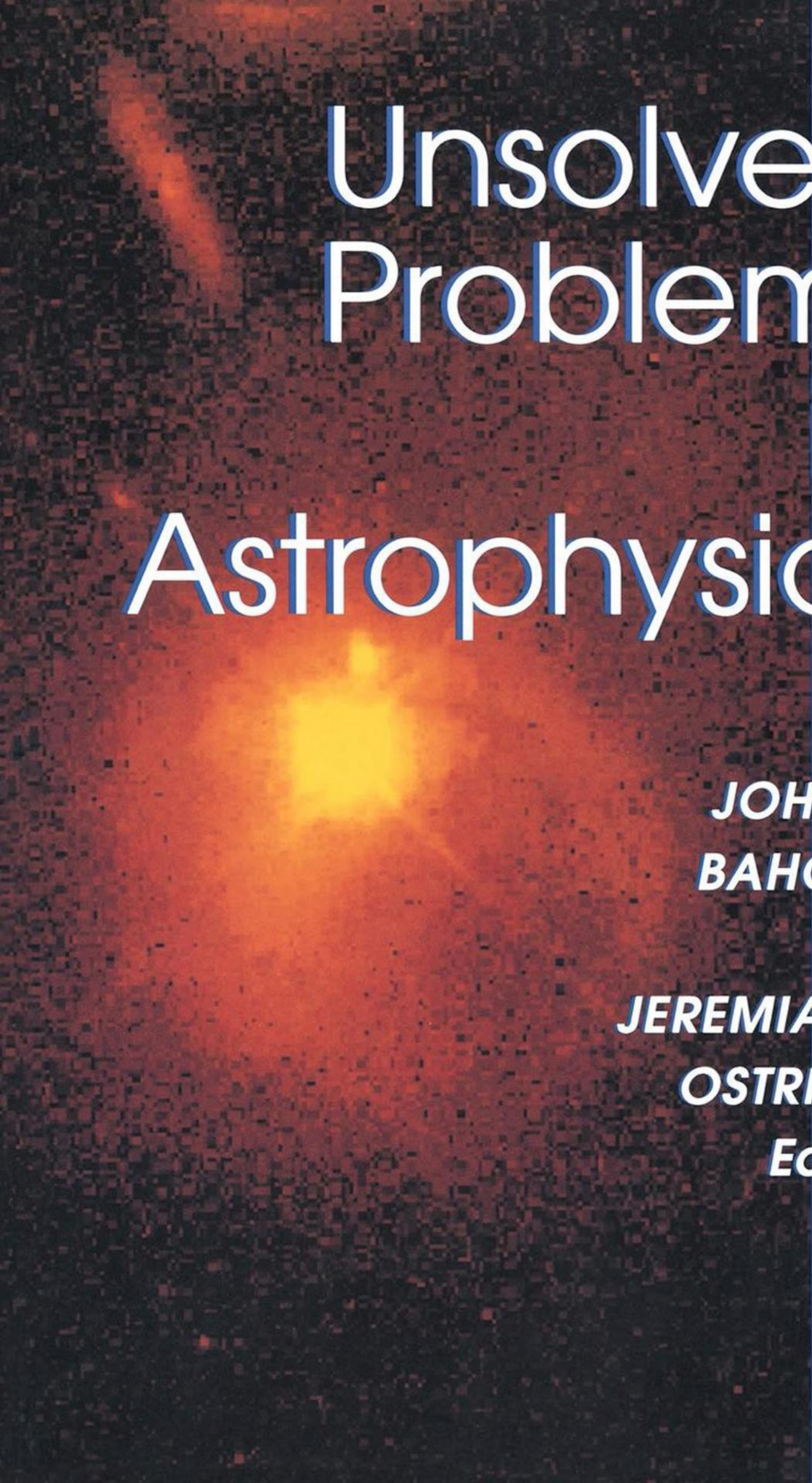
Unsolved Problems in Astrophysics



JOHN N. BAHCALL and JEREMIAH P. OSTRIKER, Editors

John N. Bahcall
and
Jeremiah P. Ostriker, Editors

UNSOLVED PROBLEMS IN ASTROPHYSICS

PRINCETON UNIVERSITY PRESS
Princeton, New Jersey

Published by Princeton University Press, 41 William Street,
Princeton, New Jersey 08540
In the United Kingdom: Princeton University Press,
Chichester, West Sussex

ISBN 0-691-01607-0 (cl)
ISBN 0-691-01606-2 (pa)

The publisher would like to acknowledge the editors of this volume for providing the camera-ready copy from which this book was printed

Princeton University Press books are printed on acid-free paper and meet the guidelines for permanence and durability of the Committee on Production Guidelines for Book Longevity of the Council on Library Resources

http://pup.princeton.edu

Printed in the United States of America

5 7 9 10 8 6 4

Contents

PREFACE

The articles in this volume were written in response to the following hypothetical situation. A second year graduate student walks into the author's office and says: "I am thinking of doing a thesis in your area. Are there any good problems for me to work on?"

Leading astrophysical researchers answer the student's question in this collection by providing their views as to what are the most important problems on which major progress may be expected in the next decade. The authors summarize the current state of knowledge, observational and theoretical, in their areas. They also suggest the style of work that is likely to be necessary in order to make progress. The bibliographical notes at the end of each paper are answers to the parting question by the hypothetical graduate student: "Is there anything I should read to help me make up mind about a thesis?"

As everyone knows who reads a newspaper or listens to the daily news, astrophysics is in the midst of a technologically driven renaissance; fundamental discoveries are being made with astonishing frequency. Measured by the number of professional researchers, astrophysics is a small field. But, astronomical scientists have the entire universe outside planet earth as their exclusive laboratory. In the last decade, new detectors in space, on earth, and deep underground have, when coupled with the computational power of modern computers, revolutionized our knowledge and understanding of the astronomical world. This is a great time for a student of any age to become acquainted with the remarkable universe in which we live.

In order to make the texts more useful to students, each of the papers was "refereed" by cooperative graduate students and colleagues. On average, each paper was refereed four times. We would like to express our gratitude to the referees; their work made the papers clearer, more accessible, and in some cases, more correct. We are grateful to each of the authors for wonderful manuscripts and for their cooperation in what must have at times seemed like an endless series of iterations. The excellent quality of the final texts justifies their hard work.

In looking over the material as it now appears, we believe that these papers may have a wider readership than we originally anticipated. Most of the articles are accessible to junior or senior undergraduate students with a good science background. The book can therefore be useful as an undergraduate introduction to some of the important topics in modern astrophysics. We hope that readers who are graduate students now or in the future will solve many of the problems listed here as unsolved. Anyone, from an undergraduate science major to a senior science faculty member, who would like to know more about some of the active areas of contemporary astrophysics can profit by reading about what these researchers think are the most important solvable problems.

The articles collected here were originally presented as invited talks at a conference entitled 'Some Unsolved Problems in Astrophysics" that was held at the Institute for Advanced Study on April 27-29, 1995. This conference was sponsored in part by the Sloan Foundation, to whom we express our gratitude. The dates for the conference were related to the 60th birthday (and 25th year at the Institute) of one of us (JNB), but nearly every effort was made to focus the meeting on science, not anniversaries. However, a large fraction of the attendees and speakers were alumni of the Institute's postdoctoral program in astronomy and astrophysics.

The manuscript for this book was expertly prepared by Margaret Best. All of us are grateful to Maggie for her exceptional editorial and TeX skills and for her constant good nature.

John N. Bahcall, Jeremiah P. Ostriker
Princeton, June 1996

Unsolved Problems in Astrophysics

CHAPTER 1

THE COSMOLOGICAL PARAMETERS

P. J. E. PEEBLES

Joseph Henry Laboratories, Princeton University, Princeton, NJ

ABSTRACT

The tests of the relativistic cosmological model are well understood; what is new is the development of means of applying them in a broad variety of ways capable of producing a network of consistency checks. When the network is tight enough we will have learned either that general relativity theory passes a highly nontrivial test or that something is wrong with classical physics. I see no reason to look for the latter but we should keep it in mind. In the former case we will have gained the boundary conditions for a deeper cosmology, and a new set of puzzles to study.

1.1 INTRODUCTION

The first thing to know about the measurement of the parameters of the standard relativistic cosmological model is that the problem has been with us for a long time. By the 1930s people understood the physics of the evolving relativistic cosmology and how astronomical observations might be used to test and constrain the values of its parameters. The first large-scale application of the astronomical tests, the count of galaxies as a function of apparent magnitude (Hubble 1936), had already reached redshift $z \sim 0.4$ (inferred from the values of Hubble's counts and more recent measurements of the mean redshift-count relation). The application of the cosmological tests was a "key project" for the 200 inch telescope when it was under construction in the 1930s; now it is a key project for the Hubble Space Telescope and the Keck Telescope.

There has been ample time for the development of strong opinions on what the results of the measurements are likely to be, and for a tendency to lose sight of the reasons for measuring the parameters. To my mind there are two main goals: extend the tests of the physics of the standard model, and seek clues to a cosmology based on a deeper level of physics. Despite the sobering record there is reason to believe we may actually be witnessing the end game in finding useful measurements of the parameters.

1.2 WHY MEASURE THE PARAMETERS?

1.2.1 *Testing the Physics*

In the standard cosmological model the universe is close to homogeneous and isotropic in the large-scale average, and it is homogeneously expanding and cooling. The evidence, which most cosmologists agree is quite strong, is summarized for example in Peebles, Schramm, Turner, and Kron (1991) and Peebles (1993). (I refer the reader to these references for details of the following comments.) The evolution of the standard model is described by general relativity theory. This part is not as closely probed, and one goal of the cosmological tests is to broaden the constraints on the underlying gravity physics. Here is the situation.

To begin, I assume the geometry of our spacetime is described by a single metric tensor, that is, a single line element which determines the relations among measured distance or time intervals between events in spacetime. The evidence is that our universe is close to homogeneous and isotropic in the large-scale average, and we know the line element of a homogeneous and isotropic spacetime is unique up to coordinate transformations; the Robertson-Walker form is[1]

$$ds^2 = dt^2 - a(t)^2 dl^2\,, \qquad dl^2 = \frac{dr^2}{1 \pm r^2/R^2} + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2. \quad (1.1)$$

In this equation a comoving observer, who moves so the universe is seen to be isotropic in the large-scale mean, has fixed position r, θ, ϕ. The proper time kept by such comoving observers is t (if the observers' clocks are synchronized so all see the same mean mass density ρ_i at a given time t_i). The parameter R^{-2} in the expression dl^2 for the spatial part of the line element measures the radius of curvature of the three-dimensional space sections of fixed world time t. If the prefactor of R^{-2} is negative space is closed, as in the surface of a balloon. The expansion factor $a(t)$ in the expression for ds^2 means the balloon in general may be expanding or contracting; the evidence is that our universe is expanding. If the prefactor of R^{-2} is negative space sections are open, the circumference of a circle of radius r being larger than $2\pi r$. In this case the nearly homogeneous

[1] Corrections for the departures from homogeneity are important for some of the cosmological tests, and will have to be reconsidered as the tests improve. The issues are discussed in Peebles (1993).

space we observe may extend to indefinitely large distances, or space might be periodic, or conditions beyond the distance we can observe might be very different. If $R^{-2} = 0$ then dl^2 is the familiar Cartesian form for flat space, and spacetime is said to be cosmologically flat.

The proper physical distance between comoving observers, measured at given world time, is $D = la(t)$, where the coordinate distance l is the result of integrating the second part of equation (1.1) along the geodesic connecting the observers at fixed t. The rate of change of the proper distance is

$$v = \frac{dD}{dt} = HD\,, \qquad H = \frac{1}{a}\frac{da}{dt}. \tag{1.2}$$

If v is much less than the velocity of light this gives a good picture for the predicted linear relation between distance and relative velocity in a homogeneous expanding universe.[2] Thus a galaxy at distance D is moving moving away at recession velocity $v = HD$, and the resulting Doppler effect stretches the wavelength of the radiation received from the galaxy by the amount

$$z \equiv \frac{\delta\lambda}{\lambda} = HD/c. \tag{1.3}$$

The constant of proportionality H is Hubble's constant; its present value usually is written as H_o. The linear relation between redshift and distance, which is Hubble's law, has been tested to redshifts on the order of unity; for an example see Figure 7 in McCarthy (1993).

Hubble's law was one of the first pieces of evidence leading to the discovery of the relativistic cosmology, but we see that the functional form follows more generally from the observed large-scale homogeneity of the universe.

When the redshift is comparable to or larger than unity equation (1.2) does not directly apply (because D is measured along a surface of fixed cosmic time t, which is not how light from a distant galaxy travels to us). The easy way to analyze the redshift in this case is to imagine that the electromagnetic field is decomposed into normal modes of oscillation with fixed boundary conditions in the space coordinates of equation (1.1) The boundary conditions mean the physical wavelength λ of a mode stretches as the universe expands:

$$\lambda \propto a(t). \tag{1.4}$$

If the interaction of the radiation with other matter and fields is weak then adiabaticity tells us the number of photons in each mode is conserved, which is to

[2] As in equation (1.1), this assumes perfect homogeneity. In the real world the galaxies are moving at peculiar velocities $\sim$ 500 km s^{-1} relative to the ideal Hubble flow in equation (1.2).

say that equation (1.4) gives the time evolution of the wavelength of freely propagating radiation as measured by comoving observers placed along the path of the radiation. The cosmological redshift factor z is defined by the equation

$$1 + z \equiv \frac{\lambda_{\rm obs}}{\lambda_{\rm em}} = \frac{a(t_{\rm obs})}{a(t_{\rm em})}. \qquad (1.5)$$

The wavelength at emission from a source at epoch $t_{\rm em}$ is $\lambda_{\rm em}$, as measured by a comoving observer at the source, and the radiation is detected at time $t_{\rm obs}$ at wavelength $\lambda_{\rm obs}$. If the time between emission and detection is small we can expand $a(t)$ in a Taylor series to get equation (1.3). It has become customary to label epochs in the early universe by the redshift factor z considered as an expansion factor even when there is no chance of detection of radiation freely propagating to us from this epoch.

The thermal cosmic background radiation (the CBR) detected at wavelengths of millimeters to centimeters is very close to homogeneous — the surface brightness departs from isotropy by only about one part in 10^5 — and the spectrum is very close to blackbody at temperature $T = 2.73$ K. To analyze the behavior of this radiation in an expanding universe suppose the homogeneous space of equation (1.1) at time t contains a homogeneous sea of thermal radiation at temperature T. The photon occupation number of a mode with wavelength λ is given by Planck's equation,

$$N = \frac{1}{e^{hc/kT\lambda} - 1}. \qquad (1.6)$$

If the radiation is freely propagating the occupation number N is conserved, so we see from equation (1.4) that the mode temperature scales with time as

$$T \propto 1/a(t). \qquad (1.7)$$

Since this is independent of the mode wavelength an initially thermal sea of radiation remains thermal, even in the absence of the traditional thermalizing grain of dust.

We know the universe now is transparent to the CBR, at least along some lines of sight, because distant galaxies are observed at CBR wavelengths. At some earlier epoch the universe could have been dense and hot enough to have been opaque and therefore capable of relaxing the radiation to the observed thermal spectrum. When this was happening the interaction of matter and radiation cannot be neglected, of course, but since the heat capacity of the radiation is much larger than that of the matter[3] equation (1.7) still applies to the coupled matter and radiation when sources or sinks of the radiation may be neglected.

[3]The energy density in the radiation is aT^4, where a is Stefan's constant. In a plasma of n protons per unit volume and a like number of free electrons the energy density is $3nkT$. The ratio is $aT^3/(3nk) \sim 10^9$, nearly independent of redshift.

The conclusion is that the CBR is a fossil of a time when our expanding universe was hotter and denser than it is now. This argument does not require general relativity theory, only conventional local physics, a nearly homogeneous and isotropic expansion described by a single line element (as in eq. [1.1]), and an expansion factor large enough to lead back to a time when the universe was opaque enough to have been capable of relaxing to equilibrium. I think all who have given the matter serious thought would agree with this; the issue is the minimum expansion factor needed to account for the observations. Hoyle, Burbidge, and Narlikar (1993) propose a Quasi-Steady State scenario in which the present expansion phase traces back to a redshift only slightly greater than the largest observed for galaxies. Others doubt that the properties of the postulated thermalizing dust grains can be chosen to relax the radiation to blackbody at such low redshifts while still allowing the observed visibility of high redshift galaxies at CBR wavelengths, though the issue certainly could be analyzed in more detail than has been done by either side so far. Most cosmologists accept the evidence for the origin of the light elements as remnants of the rapid expansion and cooling of the universe through temperatures on the order of 1 MeV, at redshift $z \sim 10^{10}$. This model for element formation depends on the rate of expansion through $z \sim 10^{10}$ and thus tests the gravity theory, as follows.

In general relativity the expansion factor $a(t)$ in equation (1.1) satisfies

$$H^2 = \left(\frac{1}{a}\frac{da}{dt}\right)^2 = \frac{8}{3}\pi G\rho \pm \frac{1}{a^2R^2} + \frac{1}{3}\Lambda. \tag{1.8}$$

The mean mass density is ρ, Λ is Einstein's cosmological constant, and the constant R^2 appears with the same algebraic sign as in equation (1.1). (I simplify the equations by choosing units so the velocity of light is unity.) The equation of local energy conservation is $\dot{\rho}/\rho = -3(\dot{a}/a)(\rho + P)$, where P is the pressure. If the pressure is not negative the mass density varies with the expansion parameter at least as rapidly as a^{-3}, meaning it is the most rapidly varying term in equation (1.8), and hence the dominant term at high redshift. Thus the predicted expansion rate through the epoch of light element production is very well approximated as

$$\left(\frac{1}{a}\frac{da}{dt}\right)^2 = \frac{8}{3}\pi G\rho. \tag{1.9}$$

For a reasonable value of the baryon number density, and assuming the baryon distribution at high redshift is close to homogeneous and the lepton numbers are small, the predicted values of the light element abundances left over from the rapid expansion and cooling of the early universe are close to the observed abundances (with modest and reasonable corrections for the effects of nuclear burning in stars).

This is a non-trivial test of general relativity theory, through equation (1.9). It is not difficult to arrive at equation (1.9) in a quasi-Newtonian picture, however, so the test arguably is not very specific.

The program of cosmological tests may extend the checks of the gravity theory. One part of the program uses galaxies or other markers to measure the curvature of spacetime. In the description of our universe by the homogeneous and isotropic line element in equation (1.1) this amounts to measuring the value and sign of the parameter R^2 and the expansion parameter a as a function of world time t. These results with equation (1.8) predict the present value of the mean mass density ρ_o, which we can hope to compare to what is deduced from studies of the dynamics of systems of galaxies, and the expansion time $t(z)$ as a function of redshift, which we can hope to compare to radioactive decay ages and stellar evolution ages applied at the present epoch and, from stellar evolution times derived from the spectra of distant galaxies, at earlier times. In equation (1.8) the parameter R^{-2} plays the role of a constant of integration or a conserved energy for the effective value $\dot{a}^2/2$ of the kinetic energy of expansion per unit mass, while in equation (1.1) it is a measure of the curvature of space sections. It will be fascinating to see a check that the same parameter plays both roles, as predicted in general relativity.

It may be useful to note in a little more detail the theory behind the use of astronomical observations to constrain $a(t)$ and R^2. We can define two distance functions, the proper rate of radial displacement of a light packet with respect to redshift,

$$s(z) = \frac{dt}{dz}, \tag{1.10}$$

and the angular size distance,

$$r(z) = R \sinh\left[\frac{1}{R}\int \frac{dt}{a(t)}\right], \tag{1.11}$$

which is the coordinate distance r in equation (1.1). This expression assumes an open model; in a closed model the hyperbolic sine is replaced by a trigonometric sine. Here are examples of the uses of these functions.

The integral of $s(z)$ in equation (1.10) over redshift is the expansion time, to be compared to other measures of time. The function $s(z)$ also enters the analysis of the absorption lines in quasar spectra that give such a remarkably detailed picture of the distribution of gas along the line of sight. If a class of clouds has proper number density $n(z)$ and cross section $\sigma(z)$ at redshift z then the density of absorption lines produced by these clouds, measured as the probability of finding a line in a quasar spectrum at redshift z in the range dz, is

$$dP = \sigma(z)\, n(z)\, s(z)\, dz. \tag{1.12}$$

A galaxy with angular diameter θ appearing at redshift z has proper diameter

$$d = a(z)r(z)\theta, \tag{1.13}$$

which is why r is called the angular size distance.

In a metric theory the surface brightness of a galaxy integrated over wavelength varies as $(1 + z)^{-4}$. It is an interesting exercise for the student to derive this from Liouville's theorem, and to check that the angular size distance in equation (1.13) thus predicts the detected energy flux density f from a galaxy with luminosity L isotropically radiating at redshift z:

$$f = \frac{L}{4\pi a_o^2 r(z)^2 (1 + z)^2}. \tag{1.14}$$

The count of objects per steradian depends on the radial and angular size distances:

$$\frac{dN}{dz} = [a(z)r(z)]^2 s(z)n(z). \tag{1.15}$$

The first factor is the proper area per steradian subtended at redshift z, the second factor is the proper radial displacement of a light packet per increment of redshift, and $n(z)$ is the number of objects per unit proper volume.

Similar expressions follow for other measurements that might be done at least in principle. A real measurement is quite another matter, of course, but there has been impressive progress, a few examples of which will be considered in the next section.

1.2.2 *How Will It All End?*

It is fascinating to think we can discover how the world ends; the standard cosmological model offers a few definite possibilities. One is a violent collapse back to a "Big Crunch," a sort of time-reversed "Big Bang." Another is expansion into the indefinitely remote future, a "Big Chill" marked by the eventual deaths of all the stars and by the eventual relativistic collapse of gravitationally bound islands of matter and the slow evaporation and dissipation of the mass of the resulting black holes. All this is romantic, but the science is debatable: why should we trust an extrapolation into the remote future of a theory we know can only be an approximation to reality? Surely the more likely outcome of a successful application of the cosmological tests would be the discovery of the boundary conditions for some deeper physical theory which most of us hope will not be blemished by the singular initial and final states of the universe or parts of it that follow within general relativity theory. According to this way of thinking an examination of the history of ideas on how the universe ought to end is valuable for the hints it might offer to a deeper theory, and we hope a knowledge of how the world ends within the present model will focus our minds on the search for a deeper picture. While we wait to see whether anything comes of this we can contemplate some interesting examples in the sociology of science.

If the curvature term does not stop the expansion then within the standard model in equation (1.8) the long-term future of the universe depends on the value of Λ. A negative value, however small, eventually returns the universe to a Big Crunch. If Λ is identically zero and the curvature term does not stop the expansion then spacetime becomes arbitrarily close to Minkowski in the remote future. Dyson (1979) points out that this allows interesting things to continue happening for a very long time, albeit operating at ever increasing timescales. If Λ is positive and the curvature term does not stop the expansion then the universe approaches the static de Sitter spacetime (where the cosmological constant term dominates the right-hand side of eq. [1.8]). In this limit spacetime returns to initial conditions like those of the cosmological inflation scenario described in these proceedings by Paul Steinhardt, who was one of the leaders in its development. Ome might imagine that this signals another round of inflation that will be followed by a phase transition that converts the present Λ into entropy. The characteristic mass corresponding to an astronomically interesting Λ is only a milli-electron volt, but it does not seem inconceivable that a complex new world could develop out of what is to us an exceedingly small energy density (Misner 1992). If this does seem inconceivable one might consider the idea that the Big Chill ends in a burst of creation of new material, as in the Quasi-Steady State scenario (Hoyle, Burbidge, & Narlikar 1993), the material being created at densities and temperatures large enough to satisfy the observational constraints (assuming that what comes after resembles what we see).

Lemaître (1933) considered the idea that the collapse of a universe to a Big Crunch might be followed by a bounce and a new expanding phase. To Lemaître this picture has (in my loose translation) "an incontestable poetic charm, bringing to mind the Phoenix of the legend," and others certainly have agreed. For example, Dicke was led to the idea that the universe might contain an observable thermal cosmic radiation background (the CBR) through the idea that starlight from the previous phase of an oscillating universe would be thermalized during a deep enough bounce and would be capable of evaporating the stars and heavy elements to provide fresh hydrogen for the next cycle (Dicke, Peebles, Roll, & Wilkinson 1965). There are no generally accepted ideas on the physics of a bounce, but evidence that there will be a Big Crunch might be expected to concentrate our thoughts.

Einstein's original static cosmological model requires nonzero values for space curvature and the cosmological constant, and it was natural therefore that the first discussions of the evolving case, by Friedmann and by Lemaître, included these terms. Einstein and de Sitter (1932) noted that neither is required to account for the expansion of the universe, and neither was needed fit the available constraints on the values of the expansion rate, the mean mass density, and space curvature. They proposed therefore that one might pay particular attention to the case in which the mass density dominates the right-hand side of the expansion equation (1.8), as in equation (1.9). This has come to be called the Einstein-de Sitter model. Einstein and de Sitter concluded by remarking that the "curvature is, however, essentially determinable and an increase in the precision of the data derived from observations will enable us in the future to fix its sign and to determine its value." I do not get the impression from this that the authors felt the only reasonable possibility is the Einstein-de Sitter model.

Prior to the discovery of the inflation scenario others were similarly cautious. Robertson (1955) stated that the Einstein-de Sitter case is "of some passing interest." In the addendum to the second edition of his book, *Cosmology*, Bondi (1960) was a little more enthusiastic: he wrote that, "in addition to its outstanding simplicity, this model has the remarkable property (unique amongst relativistic models) that $\gamma\rho R^2/\dot{R}^2$," which is the density parameter defined in equation (1.16) up to a numerical factor (and $\gamma = G$ is Newton's constant), "is constant. It may well be argued that as important a simple pure number as $\gamma\rho R^2/\dot{R}^2$ should be constant during the evolution of the universe so as to provide in some sense a constant background to the application of the theory."

Bondi's remark is close to the coincidences argument I believe originated with Dicke. He noted that if the universe did expand from a dense state then there would have been an earliest time in the past when the physics of the expanding universe can be well approximated by classical field theory. Whatever had happened earlier would have set the initial conditions for the subsequent classical evolution, and in particular would have set the values of the time t_1 at which the mass density ceases to be the dominant term in the right-hand side of the expansion equation (1.8) and the time t_2 at which we came on the scene and measured the parameters in this equation. We know t_2 is not much larger than t_1, because the known masses of the galaxies contribute at least 10% to the present value of H^2 in equation (1.8). It would be a curious coincidence if t_1 and t_2 had similar values. The likely possibility therefore is that t_1 is much larger than t_2, meaning we still are in the Einstein-de Sitter phase. I think I first learned this argument from Dicke in about 1960, when he led me into research in gravity physics and cosmology; he got around to publishing it a decade later (Dicke 1970).

If others knew the Dicke coincidences argument before the discovery of the inflation scenario it did not prevent discussions of alternatives. For example, Petrosian, Salpeter, and Szerkes (1967) and Shklovsky (1967) discussed the possible role of a nonzero cosmological constant Λ in interpreting the quasar redshift distribution, and Gott et al. (1974) argued that the estimates of the mean mass density and the expansion rate H_o, and the theory of light element production in the early universe, seemed to be most readily understandable if the mean mass density is less than the Einstein-de Sitter value.

During the mid-1980s there was a sharp swing toward the opinion that the Einstein-de Sitter model is the only reasonable possibility. This was driven by an idea, cosmological inflation (Guth 1981).

The main point of inflation for our purpose is the concept that the universe might have evolved through an epoch of rapid expansion that would have stretched the length scales of the wrinkles of the primeval ooze into values much larger than those we can explore: the universe may be disorganized and without form in the large yet very close to homogeneous and isotropic in the bit we can see. The stretching that ironed out the wrinkles would have ironed out any mean space curvature too, so in this picture the curvature term in equation (1.8) for the present value of H^2 is negligibly small.

This leaves the cosmological constant Λ, a term for which the particle physics community demonstrates a love-hate attitude. In the modern and successful theories for the weak, electromagnetic, and strong interactions there very naturally appear a set of contributions to the stress-energy tensor that act like a time-variable Λ. This was one of the element that led Guth to inflation: he postulated that a large effective Λ from particle physics drives the rapid expansion during inflation. The decay of this term into ordinary matter and radiation would end inflation and begin evolution according to the standard cosmological model. A residual nonzero Λ could be left over after inflation, and present now, but the expected value of a Λ term coming out of standard particle theory is ridiculously large compared to what is acceptable for cosmology. The natural presumption is that the cosmological constant has settled down to the only reasonable and observationally acceptable value, $\Lambda = 0$. It would follow from all this that inflation predicts the Einstein-de Sitter model.

During the past decade many people, including respected and thoughtful observational astronomers, were led to conclude that the observable part of the universe likely is well described by the Einstein-de Sitter model. I have mentioned the three main drivers: the Dicke coincidences argument, the belief that an astronomically interesting value for a cosmological cosmological is not likely to come out of particle physics, and the inflation scenario for the early universe. All three are worth serious consideration, and inflation in particular has been very influential in the development of ideas of what our universe might have been like before it was expanding. Theory was the driver, however: there was little observational evidence for the Einstein-de Sitter case and a nontrivial case against it. I can say this without excessive resort to hindsight because as inflation was becoming popular my reading of the evidence was leading me to abandon my earlier enthusiasm for the Einstein-de Sitter model: it seemed difficult to see where the large mass required by this model might be located (for the reasons discussed in Peebles 1986). As described in the next section, where some of the observational arguments pro and con $\Omega = 1$ are discussed, I have seen no reason to change my mind, but the case certainly is not yet closed.

The observational pressures on the Einstein-de Sitter model have led people to explore alternatives. Inflation's original explanation for the near homogeneity of the observable space requires that the curvature term R^{-2} be negligibly small, but it certainly would allow a nonzero Λ. N. Bahcall and Ostriker describe in these proceedings the benefits this parameter offers in interpreting the observations. Another possibility is inflation in an open universe (Ratra and Peebles 1994; Bucher, Goldhaber, & Turok 1995). Here we need another explanation for homogeneity; the best bet seems to be Gott's (1982) picture for the growth of an open universe out of an event in a de Sitter spacetime.

My impression is that research on inflation is in a healthy state: the science is being driven by advances in the observational evidence that are leading people to consider new ideas. It is too soon to decide whether the rich flow of ideas on how the world begins and ends within inflation and other scenarios is leading us toward a deeper understanding of physical reality; perhaps that will depend on the outcome of the measurements of the cosmological parameters.

1.3 THE STATE OF THE MEASUREMENTS

The first thing to understand is that we have no measurement presently capable of unambiguously distinguishing a Big Crunch from a Big Chill. We do have some promising lines of evidence, however, and the reasonable hope that research in progress will show us how this evidence can be tied together in a concordance tight enough to be believable.

The situation is illustrated in Table 1.1. The density parameter Ω is the fractional contribution of the mass density to the present value of the expansion rate in equation (1.8):

$$\Omega = \frac{8\pi G\rho}{3{H_o}^2}. \tag{1.16}$$

The case in the second column of the table, labeled $\Omega = 1$, is meant to be the Einstein-de Sitter model. It is possible that the curvature and Λ terms cancel each other at the present epoch, leaving $\Omega = 1$ with nonzero values of R^{-2} and Λ, but until we are driven from it it is sensible to operate under the assumption the Nature would not have been so unkind. By the same hopeful reasoning the third column assumes there are just two significant terms in the expansion rate equation, the dominant one being space curvature or else a cosmological constant, with almost all the rest in the mass density.

The mark $\surd$ means there is significant evidence in favor of the case, the mark X significant evidence against. I have attempted to give some indication of the degree of significance by assigning question marks to the more debatable evidence.

Table 1.1: Scorecard 1995

Observation	$\Omega = 1$	$\Omega \sim 0.1$
Dynamics and biasing on scales $\lesssim 3$ Mpc	X	$\surd$
Dynamics on scales $\gtrsim 10$ Mpc	$\surd$	$\surd$
Expansion time $H_o t_o$	?	$\surd$
Radial and angular size distances	X?	$\surd$?
Plasma mass fraction in clusters	X	$\surd$
Models for structure formation	$\surd$?	$\surd$?

The first entry in the table is based on the dynamical studies of groups and the central parts of clusters of galaxies, on scales less than a few megaparsecs. The derived mass per galaxy multiplied by the mean galaxy number density yields the contribution to the mean mass density from the material that is concentrated around galaxies; the result is equivalent to the density parameter

$$0.1 \lesssim \Omega \lesssim 0.2. \tag{1.17}$$

Another way to put it is that the small-scale dynamical measurements of the mass per galaxy agree with the mass found in the dark halos of spiral galaxies within radii

$$r_{\rm halo} \sim 300 \text{ kpc}. \tag{1.18}$$

This is discussed further in these proceedings by N. Bahcall.

The measurement in equation (1.17) seems to be secure; the open issue is whether there might be a good deal more mass in a more broadly distributed component. The galaxy masses derived from the relative gravitational accelerations in samples of close pairs of galaxies are measures of the mass concentrated around apparent brightnesses and intrinsic luminosities of stars that are luminous enough to be observationally interesting at interesting distances, and maybe also periodic or exploding. New directions have been found. One involves the measurement of time delays in the gravitational lensing events Blandford discusses in these proceedings. Another uses the physics of plasmas in clusters of galaxies, as discussed recently by Herbig et al. (1995). Yet another uses the remarkably precise measurements of the velocities and accelerations of massing interstellar clouds close to the nuclei of galaxies (Miyoshi et al. 1995). Perhaps out of all of this we may hope to know H_o to 10% by the turn of the millennium.

The angular size distance (eq. [1.11]) enters the analysis of other cosmological tests, including galaxy counts, the rate of lensing of quasar images by the mass concentrations in foreground galaxies (e.g., Fukugita & Turner 1991), and the magnitude-redshift relation applied to distant supernovae (e.g., Goobar & Perlmutter 1995).

A recent example of the first of these tests is the deep K-band (2.2 micron) counts of Djorgovski et al. (1995). Many experts advise against interpretation until the counts as a function of apparent magnitude are better established and we have a better understanding of the time evolution of galaxy luminosities and colors, but we can take note of the following indications. If the giant galaxies that tend to dominate a sample selected by apparent magnitude are evolving only through the stellar evolution of populations formed at high redshifts, the predicted counts at the faint end of the Djorgovski et al. sample differ by nearly an order of magnitude in the high and low density cases with $\Lambda = 0$. That is, the observations seem to be reaching redshifts where the effects of the cosmology are large. The effects of galaxy evolution can be large too, but we do have checks. For example, the galaxies selected by the Bergeron effect at redshift $z \sim 1$ show little evolution from the present (Steidel, Dickinson, & Persson 1994). If evolution in galaxies selected in the K-band is modest then the counts favor low Ω, with $\Lambda = 0$, and I have accordingly entered this result as tentative good news for the low density case.

The gravitational lensing of quasar images by foreground galaxies also samples large redshifts, and again the effect of the cosmological parameters on the predicted lensing rate is large (Fukugita & Turner 1991). The analysis by Maoz and Rix (1993) indicates that, if space curvature vanishes, Ω is no less than about 0.3. An interesting problem in this analysis is that it assumes elliptical galaxies have massive dark halos, in analogy with the dark halos needed in Newtonian mechanics to account for the rotation curves in spiral galaxies. Massive dark halos are needed to gravitationally contain the pools of plasma observed around giant ellipticals in clusters of galaxies. At issue here are massive halos around the more numerous less luminous ellipticals that would be responsible for lensing cases where the angular separations of the lensed images is less than about one second of arc. Tests of the mass distributions in such ellipticals based on the distribution and motions of the stars are difficult to apply, and the constraints on massive halos in these galaxies are still subject to debate. There is one case, the galaxy M 105, where a detailed study of the motions of planetary nebulae (Ciardullo, Jacoby, & Dejonghe 1993) and of a conveniently in placed ring of atomic hydrogen (Schneider 1991) yield a well-motivated mass model that requires little dark matter. If this were a common situation among the less luminous ellipticals, it would reduce the bound on Ω.

The fifth entry in the table is an elegant new test based on the standard model for the origin of the light elements and the measurements of the baryonic mass fraction in clusters of galaxies (White et al. 1993; White & Fabian 1995). A reasonable fit to the observed light element abundances follows out of the computed production of elements as the young universe expands and cools, if the density parameter in baryons is $\Omega_B \sim 0.01h^{-2}$, where $H_o = 100h$ km s^{-1} Mpc^{-1} (Walker et al. 1991). If the Hubble parameter is no less than about $h = 0.5$ and $\Omega = 1$ this would say the cosmic baryonic mass fraction is no more than about 0.05, the rest being some kind of nonbaryonic dark matter. In the central parts of rich clusters the mass fraction in observable baryons — stars and intracluster plasma — exceeds this number if $h \gtrsim 0.5$. Numerical studies of cluster formation indicate that the baryons are not likely to have been able to settle relative to the dark matter. The straightforward interpretation is that the cosmic baryonic mass fraction is significantly larger than 0.05 because the mean mass density is less than the Einstein-de Sitter value.

The last entry in the table refers to theories of the origin of structure: intergalactic gas clouds, galaxies, and the large-scale galaxy distribution. The theories depend on the cosmological model, because the cosmology determines the relation between the mass distribution, the peculiar velocities relative to the general expansion, and the perturbations to the angular distribution of the CBR (as discussed in these proceedings by N. Bahcall, Ostriker, and Steinhardt). The unraveling of clues to structure formation thus will guide our assessment of the cosmological parameters. Of particular interest now is the relation between the parameters of the cosmological model and the angular irregularities in the thermal cosmic background radiation. As Steinhardt describes, the measurements of the spectrum of angular fluctuations as a function of angular scale already significantly constrain the cosmological parameters. We should bear in mind that the constraints depend on the model for structure formation, however, and theorists have shown great ingenuity in tuning models to fit the measurements. Perhaps the improved precision of measurement of the angular fluctuation spectrum to be expected in the next few years will allow only model that will prove to be consistent with all the other observational constraints. But we should not underestimate the devious natures of astronomy and theorists.

It certainly is premature to settle bets on the value of Ω, whether consistent with the Einstein-de Sitter model or significantly lower, but we do have quite a few promising lines of evidence and good reason to expect that improvements from measurements in progress will yield enough cross checks with sufficient consistency to make a convincing case. If so, we are at last near the end of the trail Hubble (1936) pioneered six decades ago.

1.4 COSMOLOGY FOR THE NEXT GENERATION

When (or if) we arrive at a convincing interpretation of all the lines of evidence indicated in Table 1.1, along with the related tests people may invent, in terms of a set of values of the cosmological parameters, it will not mean cosmology has come to an end; we will turn to other problems. I expect we will still be debating the issue of what really happened in the remote past and what is really going to happen in the distant future. We may hope the debate will be sharper because we will know what the present standard model predicts, but I would not count on any more direct help from the observational evidence. I discuss here examples of what I expect will be observationally-driven problems for the next decade or so.

Finding a believable set of values of the cosmological parameters is likely to depend on a demonstration that the cosmology admits a successful cosmogony, a theory of the origin of galaxies. This successful cosmogony will include a description of the dominant mass components and the nature of their initial distributions at high redshift, but, unless we are lucky, there need not be an unambiguous picture of where these initial conditions came from in terms of some deeper theory. If not, we will continue to sort through the observational evidence for clues to the origin of the initial conditions for the standard model. Present research along these lines concentrates on large-scale structure, because that is close to linear and therefore relatively easily interpreted. A next step will have to be the study of the origin of the small-scale structure of the universe, which in hierarchical cosmogonies ties to the early history of structure formation. The analyses will be a good deal more difficult than for large-scale structure, but there are some openings to explore.

The drag of the CBR inhibits structure formation out of diffuse baryons at $z \gtrsim 1000$. At $z \sim 1000$, when the mean baryon density is about 100 protons cm^{-3}, the CBR is cool enough to allow the plasma to combine to neutral atomic hydrogen that can slip past the CBR to form gravitationally bound structures. Galaxies as we know them could not have formed then, because the density is at least two orders of magnitude too large, but there could have been star clusters. Could objects from this epoch have survived more or less intact as parts of present-day galaxies? My favorite candidate has been the globular star clusters, but other ideas would be welcome.

At $1000 \gtrsim z \gtrsim 100$ matter reionized by massive stars is recoupled to the CBR and radiation drag tends to force the plasma to rejoin the general expansion of the universe. That is, very early baryonic structure formation tends to be self-limiting. A closer analysis of this effect would be feasible and welcome.

If the mass of the universe were dominated by nonbaryonic pressureless matter, as in the CDM model, then mass clustering at high redshift would have grown independent of what the baryons were doing, maybe producing potential wells into which the baryons fell when radiation drag or matter pressure at last allowed it. This effect is taken into account in numerical simulations of the cosmic evolution of the distributions of baryons and dark matter, if the spatial resolution allows an adequate representation of the early development of small-scale mass clustering. I am looking forward to seeing analytic studies of what happens to the early small-scale evolution of the distributions of mass and baryons. An example of what might be found is the presence of dark massive halos around globular star clusters, at least those which formed at the baryonic Jeans mass and escaped disruption.

Galactic winds from supernovae are capable of stripping the gas from a young galaxy, leaving the dark matter halo, if it had one, and whatever stars had already formed. It has been suggested that such failed galaxies may be present now in the voids between the giants, along with most of the mass of an Einstein-de Sitter universe. Failed galaxies could be visible by the starlight from the generations produced before the gas was expelled. Perhaps evidence for this picture comes from the weaker clustering of galaxies selected for strong emission lines (Salzer 1989), or for low surface brightness (Mo, McGaugh, & Bothun 1994). In the former case one sees from maps of the galaxy distributions that the clustering is weaker because the emission line galaxies avoid dense regions. This has a natural interpretation in terms of environment: collisions and the ram pressure of the intergalactic plasma in dense regions tend to strip the galaxies of the interstellar gas responsible for the emission. Missing so far is evidence of a population that lives in the voids between the concentrations of bright galaxies. I expect to see continued discussion of this observational issue and the related theoretical one: would failed galaxies be expected to have accreted enough gas to be detectable at low redshift in quasar absorption line studies?

The baryon densities of objects starting to form at $z \sim 100$, when drag by the CBR becomes unimportant, would be about 1 proton cm^{-3}. Might it be significant that this is the density characteristic of the bright parts of a giant galaxy?

The massive dark halos are less dense, but they could have been added at lower redshifts.

In the past decade there has been considerable interest the idea that even the bright parts of the giant galaxies were assembled much later, at $z \lesssim 2$. This grew out of early numerical studies of the biased $\Omega = 1$ CDM model, and it was reenforced by the observations of substantial galaxy evolution at $z \lesssim 0.5$.[5] Quasars at $z \sim 5$ might be accommodated in this late assembly picture, but it certainly looks like a tight fit. The same is true of the observations of galaxies at redshifts as high as $z \sim 3$ with relaxed-looking morphologies and spectra characteristic of star populations. There have been dramatic recent improvements in such observations, as Ellis and Sargent describe in these proceedings, but it is worth noting that the subject is not entirely new: consider Oke's (1984) remark, "When one looks at the spectra of first-ranked cluster galaxies over the whole range of z covered", to redshift $z \sim 0.8$, "one is impressed by the fact that the vast majority are very similar to each other and to nearby ellipticals." Nature certainly has been inviting us to consider the idea that galaxy assembly occurred well before $z = 1$. It will be interesting to see whether the current evidence for the assembly of the giant elliptical galaxies $z \gtrsim 3$ holds up, and if so whether any of the variants of the CDM model now under discussion can account for this early formation.

There is a plateau in the abundance of quasars at $2 \lesssim z \lesssim 3$. Is this telling us this is when the L_* galaxies were assembled? Can we imagine instead that the galaxies were assembled earlier and the time needed to produce a central engine brought the age of the quasars to $z \sim 2.5$?

At $z \sim 4$ there is about as much diffuse baryonic mass in the intergalactic medium in the Lymanα forest clouds as there is in the young galaxies. If the galaxies had been assembled at $z \gg 4$ they would have tended to gravitationally accrete the intergalactic matter, removing the Lyα forest. Because the forest clearly is present at $z = 4$ this line of argument suggests galaxies were not assembled much earlier. A counter argument uses the abundant evidence of mass exchange between present-day galaxies and the intergalactic medium (e.g., Irwin 1995). Could it be that galactic winds in gas-rich young galaxies born at $z \gg 4$ fed the intergalactic medium, accounting for the presence of the Lymanα forest without overly perturbing the CBR or overly polluting the forest clouds with heavy elements?

Which formed first, galaxies or clusters of galaxies? Hubble (1936) argued for the latter.[6] Lemaître (1934) argued that galaxies and clusters of galaxies may have been produced by the gravitational instability of the expanding universe. The

[5] A good sample of the debate is in Frenk et al. (1989)

[6] The argument is mainly of historical interest. Following Jeans, Hubble considered the idea that elliptical galaxies evolve into spirals. Since spirals prefer less dense surroundings than ellipticals, this suggested the field galaxies might have evolved out of ellipticals that escaped from clusters.

[7] Butcher, M., Goldhaber, A. S., & Turok, N. 1995, Phys. Rev. B, in press.

[8] Ciardullo, R., Jacoby, G. H., & Dejonghe, H. B. 1993, ApJ, 414, 454.

[9] Dekel, A., et al. 1993, ApJ, 412, 1.

[10] Dekel, A., & Rees, M. J. 1994, ApJ, 422, L1.

[11] Dicke, R. H. 1970, Gravitation and the Universe (Philadelphia: American Philosophical Society).

[12] Dicke, R. H., Peebles, P. J. E., Roll, P. G., & Wilkinson, D. T. 1965, ApJ, 142, 414.

[13] Djorgovski, S., et al. 1995, ApJ, 438, L13.

[14] Dyson, F. 1979, Rev. Mod. Phys., 51, 447.

[15] Efstathiou, G., & Rees, M. J. 1988, MNRAS, 230, 5P.

[16] Einstein, A., & de Sitter, W. 1932, Proc. NAS, 18, 213.

[17] Freedman, W. L., et al. 1994, Nature, 371, 757.

[18] Frenk, C. S., et al. 1989, The Epoch of Galaxy Formation (Dordrecht: Kluwer).

[19] Fukugita, M., & Turner, E. L. 1991, MNRAS, 253, 99.

[20] Goobar, A., & Perlmutter, S. 1995, ApJ, 450, 14.

[21] Gott, J. R. 1982, Nature, 295, 304.

[22] Gott, J. R., Gunn, J. E., Schramm, D. N., & Tinsley, B. M. 1974, ApJ, 194, 543.

[23] Guth, A. H. 1981, Phys. Rev. D, 23, 347.

[24] Herbig, T., Lawrence, C. R., Readhead, A. C. S., & Gulkis, S. 1995, ApJ, 449, L5.

[25] Hoyle, F., Burbidge, G., & Narlikar, J. V. 1993, ApJ, 410, 437.

[26] Hubble, E. 1936, Realm of the Nebulae (New Haven: Yale Univ. Press).

[27] Irwin, J. I. 1995, PASP, 107, 715.

[28] Lanzetta, K. M., Bowen, D. V., Tytler, D., & Webb, J. K. 1995, ApJ, 442, 538.

[29] Lemaître, G. 1933, Ann. Soc. Sci. Brussels, A53, 51.

[30] Lemaître, G. 1934, Proc. NAS, 20, 12.

[31] Maoz, D., & Rix, H.-W. 1993, ApJ, 416, 425.

[32] McCarthy, P. J. 1993, ARA&A, 31, 639.

[33] Misner, C. W. 1992, in the 13th International Conference on General Relativity and Gravitation, Univ. Nacional de Cordoba, Argentina, ed. P. W. Lamberti, & O. E. Ortiz, p. 417.

[34] Miyoshi, M., et al. 1995, Nature, 373, 127.

[35] Mo, H. J., McGaugh, S. S., & Bothun, G. D. 1994, MNRAS, 167, 129.

[36] Oke, J. B. 1984, in Clusters and Groups of Galaxies, ed. F. Mardirossian, et al. (Dordrecht: Reidel), 99.

[37] Partridge, R. B., & Peebles, P. J. E. 1967, ApJ, 147, 868.

[38] Peebles, P. J. E. 1986, Nature, 321, 27.

[39] Peebles, P. J. E. 1993, Principles of Physical Cosmology (Princeton: Princeton Univ. Press).

[40] Peebles, P. J. E. 1995, ApJ, 449, 52.

[41] Peebles, P. J. E., Schramm, D. N., Turner, E. L., & Kron, R. G. 1991, Nature, 352, 769.

[42] Petrosian, V., Salpeter, E. E., & Szekeres, P. 1967, ApJ, 147, 1222.

[43] Ratra, B., & Peebles, P. J. E. 1994, ApJ, 432, L5.

[44] Robertson, H. P. 1955, PASP, 67, 82.

[45] Salzer, J. J. 1989, ApJ, 347, 152.

[46] Schneider, S. E. 1991, in Warped Discs and Inclined Rings Around Galaxies, ed. S. Casertano, P. D. Sackett, & F. H. Briggs (Cambridge: Cambridge Univ. Press), 25.

[47] Shaya, E. J., Peebles, P. J. E., & Tully, R. B. 1995, ApJ, 454, 15.

[48] Shklovsky, I. S. 1967, Astr. Tsirk., No. 429.

[49] Steidel, C. C., Dickinson, M., & Persson, S. E. 1994, ApJ, 437, L75.

[50] Walker, T. P., et al. 1991, ApJ, 376, 51.

[51] White, D. A., & Fabian, A. C. 1995, MNRAS, 273, 73.

[52] White, S. D. M., Navarro, J. F., Evrard, A. E., & Frenk, C. S. 1993, Nature, 366, 429.

[53] Zaritsky, D., Olszewski, E. W., Schommer, R. A., Peterson, R. C., & Aaronson, M. A. 1989, ApJ, 345, 759.

CHAPTER 2

IN THE BEGINNING ...

PAUL J. STEINHARDT

Department of Physics and Astronomy,
University of Pennsylvania, Philadelphia, PA

ABSTRACT

New observational technologies are transforming cosmology from speculative inquiry to rigorous science. In the next decade, observations will be made which will provide redundant, quantitative tests of theoretical models for the origin and evolution of large-scale structure in the universe. Using the inflationary theory as an example, this paper discusses some of the crucial tests and how they can be combined to discriminate among different models and measure essential parameters.

2.1 THE FUTURE FATE OF COSMOLOGY

In the beginning ...

... long before there was science, there was cosmology. As soon as the first intelligent eyes were capable of peering up at the heavens, questions began to arise concerning the origin and evolution of the Universe: How big is the Universe? How old is the Universe? How did it begin and how did it develop into what we see today? Many answers have been proposed over the centuries, yet these remain the fundamental, unsolved problems of cosmology today (see the contribution by P. J. E. Peebles to these proceedings).

What has changed is that there is now the technology needed to test our proposed answers to these questions. The transformation began near the turn of this century with the construction of the first, giant, optical telescopes. It was first discovered that the Universe is composed of galaxies and that the galaxies are receding from another. New observational technologies in successive decades have brought further breakthroughs.

Now, as the beginning of a new millennium approaches, the number and diversity of new technologies is increasing at an incredible pace. The forthcoming decade will bring the first galaxy surveys with red shifts of millions of galaxies allowing the first extensive three-dimensional reconstruction of large-scale structure in the Universe; the first high-resolution maps of the cosmic microwave background, providing a snapshot of the universe in its infancy; detections of cosmic sources of gamma-rays, x-rays, infra-red radiation, and perhaps even gravitational waves; and, the first extensive catalogs of supernovae and gravitational lenses (see, e.g., the contributions by N. Bahcall and R. Blandford to these proceedings). Each of these new measurements represents a fundamentally new probe of the cosmos. Together, they form a powerful, quantitative, and redundant suite of tests that will make it possible to decisively discriminate among competing cosmological models.

It appears that, for the first time in human history, cosmology will be as much observation-driven as theory-driven, finally achieving the balance required for a true, healthy science. Young scientists selecting directions for their research careers should be aware of the lessons of history which suggest that, once a field reaches this epochal stage, a heroic age will ensue in which some of the fundamental problems of the ages may be resolved.

What is discovered in the next decade may determine whether cosmology ever achieves its grand ambition. The nominal purpose of cosmology is to weave together a story, a history of the evolution of the Universe. In this sense, the science of cosmology is most similar to archaeology or paleontology. Its approach is similar, too. The basic methods entail gathering fossil relics from different epochs and tracing the evolutionary links between them. In addition, though, cosmology also has a not-so-hidden, grand ambition: to reduce this history by explaining it as a natural consequence of some simple, predictive and explanatory model based on symmetries, general physical principles, and known physical processes. It has been, to a large degree, an article of faith that there exists such a simple predictive and explanatory theory. This faith has remained resolute as long as one has found that the Universe is a relatively simple place. But, with the torrent of new observations, it is conceivable that the Universe will be found to be a more complex place than has been supposed.

Only a tiny patch of the Universe can be observed. Causality forbids seeing beyond 15 billion light years or so. Most likely, the total Universe is much larger. The hope has been that the Universe is simple enough that, by observing just our one, tiny patch, one can understand the Universe entire. Instead, it may be found tested by diverse experiments. An important advantage is that the anisotropy is a linear response to well-understood physical processes. Consequently, the physical interpretation of anisotropy measurements is less subject to model-dependent assumptions than other cosmological tests. For these reasons, there is real hope that microwave background measurements in the next decade will make a historic contribution to our understanding of the universe.

The cosmic fingerprint [6] is obtained from a temperature difference map (Fig. 2.1) which displays $\Delta T(\mathbf{x})/T_\gamma$ as a function of sky direction $\mathbf{x}$. The map represents the deviations in temperature from the mean value, $T_\gamma = 2.726 \pm .010$. The root-mean-square deviation from average in the COBE map is of order $\Delta T/T_\gamma = 0.001\%$. The temperature autocorrelation function, $C(\theta)$, compares the temperature at points in the sky separated by angle θ:

$$\begin{aligned} C(\theta) &= \left\langle \tfrac{\Delta T}{T_\gamma}(\mathbf{x}) \tfrac{\Delta T}{T_\gamma}(\mathbf{x}') \right\rangle \\ &= \tfrac{1}{4\pi} \textstyle\sum_\ell (2\ell+1) C_\ell P_\ell(\cos\theta), \end{aligned} \quad (2.1)$$

where $\langle\rangle$ represents an average over the sky and $\mathbf{x} \cdot \mathbf{x}' = \cos\theta$. The coefficients, C_ℓ, are the *multipole moments* (for example, C_2 is the quadrupole, C_3 is the octopole, etc.). Roughly speaking, the value of C_ℓ is determined by fluctuations

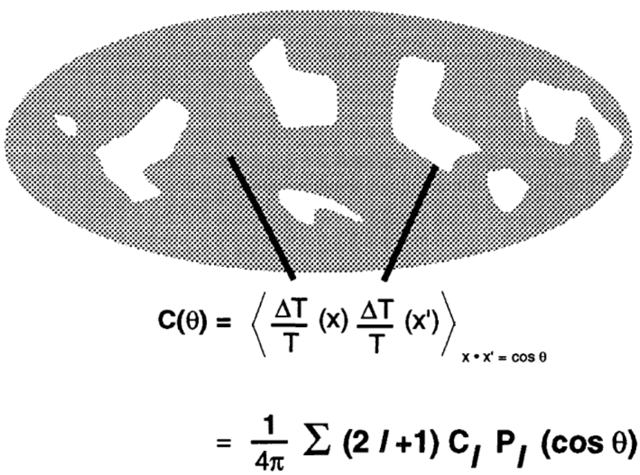


Figure 2.1: The temperature autocorrelation function, $C(\theta)$, is obtained from a map of the sky (here represented by the oval) displaying the difference in the microwave background temperature from the average value, $\Delta T/T$. $C(\theta)$ is computed by taking the map-average of the product of $\Delta T/T$ measured from any two points in the sky separated by angle θ. If $C(\theta)$ is expanded in Legendre polynomials, $P_\ell(\cos\theta)$, the coefficients C_ℓ are the *multipole moments*.

on angular scales $\theta \sim \pi/\ell$. A plot of $\ell(\ell+1)C_\ell$ is referred to as the *cosmic microwave background (CMB) power spectrum.* An exactly scale-invariant spectrum of primordial energy density fluctuations, if there were no evolution when the fluctuations pass inside the Hubble horizon, produces a flat CMB power spectrum curve (*i.e.*, $\ell(\ell+1)C_\ell$ is independent of ℓ).

There is valuable information in the cosmic microwave background anisotropy in addition to the CMB power spectrum that will be extracted some day. Higher-point temperature correlation functions (entailing three or more factors of $\Delta T/T_\gamma$) could be obtained from the temperature difference map and be used to test if the fluctuation spectrum is gaussian, as predicted by inflation. However, the fact that statistical and systematic errors in $\Delta T/T_\gamma$ compound for higher-point correlations makes precise measurements very challenging. Polarization of the microwave background by the last scattering of photons from the anisotropic electron distribution is another sky signal that provides quantitative data that can be used to test models. However, for known models, the predicted polarization requires more than two orders of magnitude better accuracy than anisotropy measurements alone in order to discriminate models [16]. For the coming decade, the most reliable information will be the CMB power spectrum. Fortunately, the CMB power spectrum is packed with information that can be used, by itself, to discriminate the leading cosmological models.

Here we will focus on how measurements of the CMB power spectrum can be used to test the inflationary paradigm and measure cosmological parameters. For alternative theories of large-scale structure formation, the model tests and criteria for constraining parameters are similar, but differ in detail. The CMB power spectrum for inflationary models is a direct reflection of the key, generic features of inflation outlined in the previous section. These can be seen by studying Figure 2.2, which displays a representative power spectrum for inflation [17]. For this example, the spectral index is $n = 0.85$ with equal scalar and tensor contributions to the quadrupole moment, within the range allowed by inflation. Left to right in the figure spans large- to small-angular scale fluctuations.

It is useful to imagine a vertical line at $\ell \approx 100$. To the left of $\ell \approx 100$ are multipoles dominated by fluctuations with wavelengths much larger than the size of the Hubble horizon at the time of last scattering, spanning angles $> 1^\circ - 2^\circ$ on the sky. According to the inflationary model, these wavelengths did not have a chance to evolve before last scattering and the beginning of the photon trek towards our detectors. Hence, these fluctuations preserve the imprint of whatever fluctuations were there in the first place. If the fluctuations are remnants of a nearly scale-invariant inflationary spectrum, then the low-ℓ part of the CMB power spectrum should be featureless, just as shown in the figure.

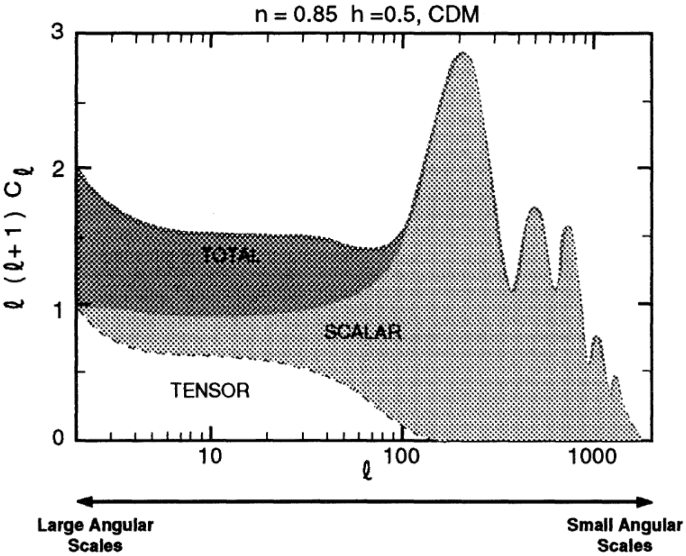


Figure 2.2: A plot of $\ell(\ell + 1)C_\ell$ vs. multipole moment number ℓ is the cosmic microwave background power spectrum. For a given ℓ, C_ℓ is dominated by fluctuations on angular scale $\theta \sim \pi/\ell$. In inflation, the power spectrum represents the sum of two independent, scalar and tensor contributions.

The spectrum includes, in general, both scalar and tensor contributions, as indicated in the example of Figure 2.2. For inflationary models, the two contributions are predicted to be statistically independent. Consequently, the total power C_ℓ is simply the sum of the scalar and tensor contributions, $C_\ell^{(S)}$ and $C_\ell^{(T)}$, respectively. The fluctuations in $\Delta T/T$ are also predicted to be Gaussian-distributed for inflationary models. The C_ℓ's, which are an average over $2\ell + 1$ Gaussian-distributed variables, have a χ^2-distribution.

To the right of $\ell \approx 100$ are multipoles dominated by fluctuations with wavelengths smaller than the horizon at last scattering. This side of the power spectrum appears different from the left because there are additional physical effects once inhomogeneities enter the Hubble horizon and begin to evolve. Gravitational waves inside the horizon red-shift away. For energy density fluctuations, the baryon and photon begin to collapse and oscillate acoustically about the centers of high and low energy density, adding to the net microwave background perturbation.

Each wavelength laid down by inflation begins its acoustic oscillation shortly after entering the Hubble horizon. Hence, there is a well-defined phase-relation between the acoustic oscillations on different wavelengths. Waves just entering the horizon and smaller-wavelength waves which have completed a half-integral number of oscillations by last scattering will be at maximum amplitude. Wavelengths in-between are mid-phase and will have smaller amplitudes. In a plot of C_ℓ's, increasing ℓ corresponds to multipoles dominated by decreasing wavelengths. The variations of the oscillation phase with wavelength results in a sequence of peaks as a function of ℓ. These peaks are sometimes referred to as Doppler peaks or acoustic peaks.

The position of the first Doppler peak is of particular interest. Its position along the ℓ-axis, left or right, is most sensitive to the value of Ω_{total}. The peak moves to the right in proportion to $1/\sqrt{\Omega_{\text{total}}}$ [18], for large Ω_{total}. There is only weak dependence on the Hubble constant and other cosmological parameters [19]. Decreasing Ω_{total} to 0.2, say, causes the first Doppler peak to shift to $\ell \approx 600$ instead of $\ell \approx 200$, a dramatic and decisive difference. As a test of Ω_{total}, the first Doppler peak has the advantage that it is relatively insensitive to the form of the energy density, whether it be radiation, matter, or cosmological constant and it is relatively difficult to mimic using other physical effects.

In sum, Figure 2.2 illustrates how all three key features of inflation can be

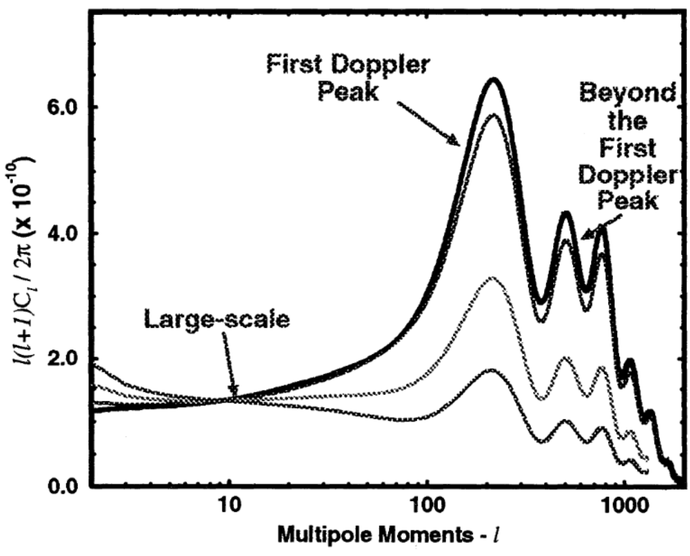


Figure 2.3: A band of microwave background power spectra allowed by inflation. The uppermost curve is a pure-scalar, scale-invariant spectrum, and the lower curves have tilt ($n < 1$) and gravitational waves. Inflationary models with spectra somewhat higher than the uppermost curve are also possible. The common features among these curves—the prime targets for experimental tests of inflation—are a plateau at large angular scales, a prominent first Doppler (or acoustic) peak, and subsequent acoustic oscillation peaks at small angular scales.

tested by the microwave background power spectrum. Large-angular scale fluctuations are consequences of scale-invariance and the combination of scalar and tensor perturbations. A nearly flat CMB power spectrum at small ℓ is the signature of being nearly scale-invariant. Small-angular scale fluctuations, especially the position of the first Doppler peak, are consequences of Ω_{total} being unity. The presence of a combination of scalar and tensor fluctuations can be determined from more subtle features, such as the ratios of the Doppler peaks to the plateau at small ℓ.

The inflationary prediction for the CMB power spectrum is not unique, since there are undetermined, free parameters. Figure 2.3 is a representative band of predicted curves for different values of the spectral index. Each example lies within the parameter space achievable in inflationary models. Although the band is wide, there are common features among the curves which can be used as the critical tests of inflation. All have a plateau at small ℓ, a large first Doppler peak at $\ell \approx 200$, and then smaller Doppler hills at larger ℓ.

The key microwave background tests of inflation for the next decade break down into two key battles. The first is the "Battle of the Primordial Plain," illustrated in Figure 2.4. Displayed is a blow-up of the power spectrum over the

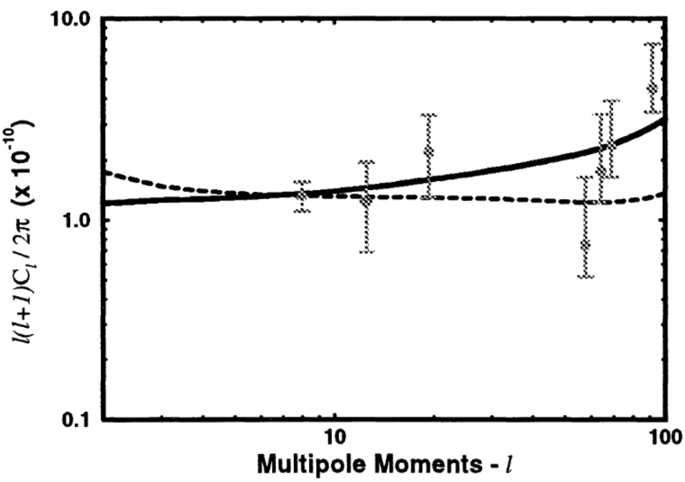


Figure 2.4: Blow-up of the power spectrum focusing on large angular scales. The upper curve corresponds to a pure-scalar $n_s = 1$ (scale-invariant) spectrum, and the lower curve has $n_s = 0.85$ and equal scalar and tensor contributions to the quadrupole. Superimposed are the experimental flat band power detections with one-sigma error bars. Left-to-right, these correspond to: (a) COBE; (b) FIRS; (c) Tenerife; (d) South Pole 1991; (e) South Pole 1994; (f) Big Plate 1993-4; and (g) PYTHON.