



Fermi National Accelerator Laboratory

FERMILAB-Conf-95/151-E

Cosmological Models at the Millennium

Joshua A. Frieman

*Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510*

June 1995

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Cosmological Models at the Millennium

Joshua A. Frieman

NASA/Fermilab Astrophysics Center, Fermi National Accelerator Laboratory, Batavia, IL 60510

Department of Astronomy and Astrophysics, The University of Chicago, Chicago, IL 60637

This chapter provides an overview of cosmological models, starting with a review of the standard big bang cosmology and ending with scenarios for the early universe. We introduce the expanding, homogeneous and isotropic Friedmann-Robertson-Walker universe and briefly summarize the observational evidence which has led to its adoption as the standard cosmological model. We then discuss scenarios for the very early universe, focusing on the inflationary universe and topological defect models, and their possible role in seeding the observed large-scale structure of the Universe. We emphasize the exciting prospects for testing early universe models, and thus for probing particle physics at an energy scale of order 10^{16} GeV, with astronomical and terrestrial observations in the coming decade.

§1 Introduction

Over the last fifteen years, theoretical cosmology has undergone a renaissance: extrapolating concepts from elementary particle physics, in particular the standard electroweak gauge theory, to very high energies, a framework has emerged in which one can meaningfully speculate about the evolution of the very early universe. This marriage of particle physics and cosmology has led to a number of remarkable developments, including models for the generation of the baryon asymmetry, the inflationary scenario, the notion that topological defects could be created in cosmological phase transitions, and predictions for non-baryonic particle dark matter, to name just a few. In addition, both inflation and topological defects have provided causal frameworks for generating primordial density perturbations in the early universe, thought to be the seeds for galaxy formation.

In recent years, observational cosmology has been undergoing its own rebirth. There has been an explosion of information on the large-scale clustering of galaxies from redshift, peculiar velocity, and photometric surveys gathered by ground-based telescopes. Studies of rich clusters of galaxies via their gravitational lensing effects as well as X-ray emission from hot intracluster gas have started to provide new clues to the distribution of dark matter. The detection of large-angle anisotropies in the cosmic microwave background radiation by the COBE satellite has provided the first probe of structure on very large scales, while a series of ground-based and balloon-borne anisotropy experiments on smaller scales now have positive results as well. On the scale of the universe itself, there has been steady progress in attempts to measure cosmological parameters (in particular, the age, expansion

rate, and mean density) as well as the light element abundances more precisely, using the Hubble Space Telescope and the new generation of large ground-based telescopes.

As a consequence of these observational advances, cosmology is becoming data-driven in an unprecedented way: we can now confront cosmological models of the early universe with an impressive array of observations. There is still debate about the reliability and interpretation of much of the data, but we are definitely entering what one might call the ‘scientific’ age of cosmology: those theories which are sufficiently predictive (and therefore interesting) are becoming increasingly falsifiable and will stand or fall in the coming years. This is a very healthy development for the field. It also implies that a larger portion of cosmologists’ time in the future will be spent in the realm of ‘phenomenology’, developing more precise methods for testing theoretical predictions against observational results and discriminating among competing models.

It is safe to say that at present the big bang framework for the large-scale evolution of the universe rests upon solid foundations. We also have in hand rough architectural sketches of a model which may account for the origin and evolution of structure within this framework: primordial, quasi-scale-invariant perturbations from inflation or defects, with some variant of cold dark matter, which subsequently grow by gravitational instability. However, while the picture is becoming clearer, we still lack firm answers to basic questions about the large-scale universe, among them: what is the origin of the perturbations responsible for large-scale structure? once perturbations are generated, how do luminous galaxies form in detail? what is the universe made of, that is, what is the dark matter(s), how much is there, and how is it distributed? when and how was the baryon (matter/anti-matter) asymmetry

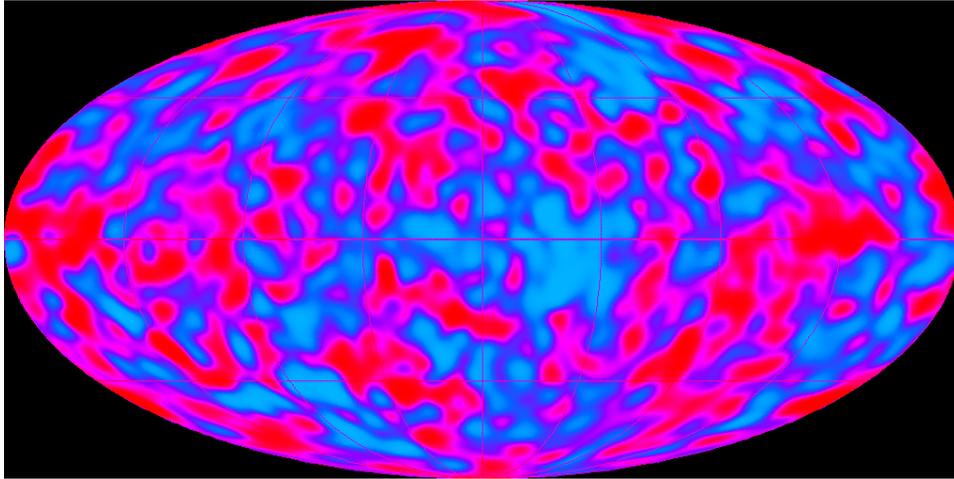


Figure 1: Temperature fluctuations in the CMBR observed by COBE DMR. Lighter areas correspond to low-temperature regions. The typical fluctuation amplitude is of order 10^{-5} [2].

generated? Beyond these basic questions there are others for which more precise observational answers are eagerly anticipated at the millennium, for they involve parameters for the models, in particular: how fast the universe is expanding, the rate at which the expansion is being slowed by gravity, and the age of the universe.

The first section provides an overview of the standard cosmology and the observational evidence in its favor. Subsequent sections focus on scenarios for the early universe, their predictions for the generation of structure, and how they may be tested by observations in the near future. Most of the topics introduced here will be treated in greater depth in subsequent chapters. The Appendix introduces the standard notation and units used.

§2 The Standard Cosmology

2.1 Homogeneity and Isotropy

The standard hot Big Bang model, based on the homogeneous and isotropic Friedmann-Robertson-Walker (FRW) spacetimes, is a remarkably successful operating hypothesis describing the evolution of the Universe on the largest scales. It provides a framework for such observations as the Hubble law of recession of galaxies, interpreted in terms of the expansion of the universe; the abundances of the light elements, in excellent agreement with the predictions of primordial nucleosynthesis; and the thermal spectrum and angular isotropy of the cosmic microwave background radiation (CMBR), as expected from a hot, dense early phase of thermal equilibrium between matter and radiation. These three pieces of evidence are the fundamental pillars upon which our confidence in the standard cosmology rests.

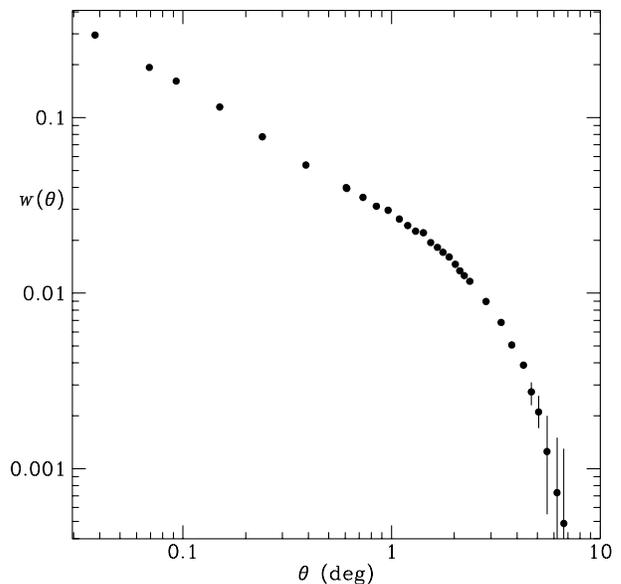


Figure 2: Angular correlation function $w_{gg}(\theta)$ measured in the APM survey, for galaxies in the apparent magnitude range $17 < b_J < 20$ [4].

While homogeneity and isotropy are, strictly speaking, assumptions of the model, they rest on a strong and growing foundation of observational support. The primary evidence for angular isotropy on large scales comes from the smallness of the CMBR large-angle anisotropy detected by the COBE satellite [1, 2] and the FIRS balloon experiment [3]. Fig. 1 shows a map of the CMBR produced from the first two years of COBE DMR data; the rms temperature fluctuation on an angular scale of 10° is $(\Delta T/T) \simeq 10^{-5}$. Additional support comes from the isotropy of radiation backgrounds at other wavelengths, as well as from the isotropy of deep galaxy and radio source counts.

IRAS QDOT Survey ($f_{80} > 0.6 \text{ Jy}$, $|b| > 10^\circ$)
1824 galaxies

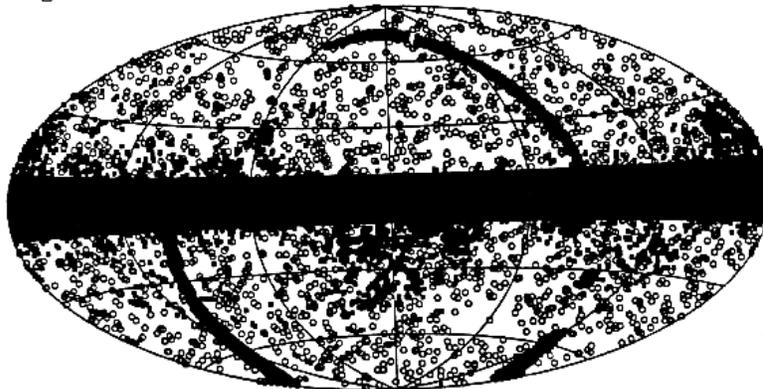


Figure 3: An equal-area projection of the IRAS-QDOT galaxies (open circles) at distances in the range $20 - 500 h^{-1}$ Mpc. Solid squares are regions not surveyed, and the black band is an excluded region encompassing the Galactic disk [9].

These different sources carry information about quite disparate scales and types of matter in the Universe. For example, the APM[4] and EDSGC[5] surveys measured the angular positions of more than one million galaxies over an area covering about 10 % of the southern sky out to an effective depth of roughly $600 h^{-1}$ Mpc. In these surveys, the rms fluctuation in galaxy number density on an angular scale of 5° is of order 10^{-3} (see Fig. 2). In other words, averaged over sufficiently large angular scales, we see roughly the same number of galaxies (brighter than a fixed limit) per steradian in different parts of the sky at this depth. These galaxy surveys give us information about the relatively local distribution of *luminous* matter in the universe. On the other hand, through the Sachs-Wolfe effect, the large-angle CMBR measurements directly probe the gravitational potential over length scales of order $6000 h^{-1}$ Mpc, and are thus sensitive to the mass distribution itself – the distinction is important to keep in mind, since the evidence for *dark matter* leads us to be wary about identifying the distribution of light with that of mass. Of course, the non-zero fluctuations in galaxy density and CMBR temperature contain interesting information about the clustering of galaxies and mass, to which we return below.

Evidence for large-scale homogeneity comes in part from galaxy redshift surveys. However, compared to the angular photometric surveys, which currently give two-dimensional information for several million galaxies, currently complete redshift surveys yield three-dimensional information for typically ten thousand galaxies in a more local neighborhood. In the redshift surveys, one can verify directly that the rms fluctuations in the spatial number density of galaxies become small when averaged over large enough scales. For example, in the full-sky surveys selected from infrared sources in the IRAS catalog (the 1.2

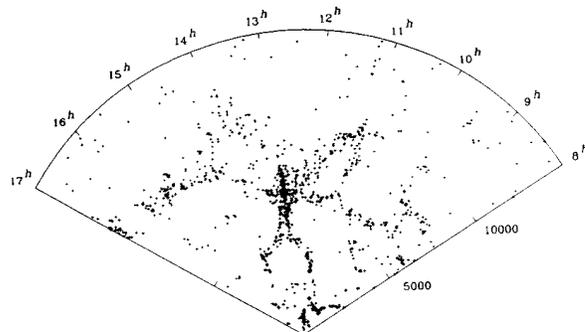


Figure 4: A 6° -thick slice from the CfA redshift survey: the angular coordinate is right ascension (in hours), and the radial coordinate is recession velocity in km/sec [10]. The concentration in the center is the Coma cluster, while the band of galaxies stretching across the figure at $v \sim 7000$ km/sec is part of the Great Wall.

Jy survey [6] and the 1-in-6 QDOT survey [7, 8] shown in Fig. 3), the rms fluctuation in the number of galaxies in cubical volumes of side $L = 60 h^{-1}$ Mpc is of order $\delta N_{gal}/N_{gal} \simeq 0.2$ and decreases with increasing cell volume. This approach to homogeneity is roughly consistent with that seen in the deeper angular surveys. Larger structures such as superclusters, great attractors, great voids, and long filaments do exist, and have received considerable attention. Of particular note in this regard are the Great Wall, extending roughly $170 \times 60 \times 5 h^{-3}$ Mpc³ in the Center for Astrophysics (CfA) survey extension [9, 10] (see Fig. 4), and the peaks separated by ~ 100 Mpc in deep pencil-beam surveys [11]. But in a statistical sense, the net fluctuations in galaxy number become small on the largest scales where they have been reliably counted in large-area redshift surveys. This is consistent with the visual impression from Fig. 3.

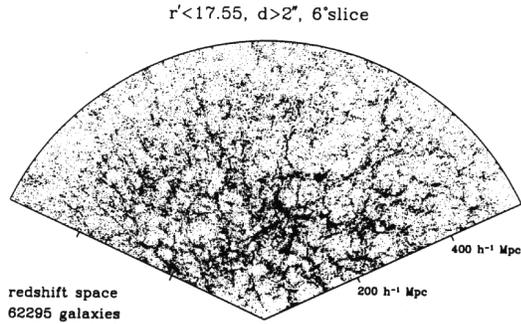


Figure 5: A simulated 6° -thick slice from the Sloan Digital Sky Survey: the radial coordinate is recession velocity in units of the Hubble parameter (redshift) [12]. This slice is roughly four times deeper than that in Fig. 4.

In the near future, the galaxy maps shown here will be substantially extended by new surveys using multi-fiber spectrographs to simultaneously measure many redshifts in the same field, an ingenious technological development which makes large-area redshift surveys possible to greater depths in a finite survey time. Advances in multi-fiber spectroscopy are being exploited by the ongoing Las Campanas survey, the planned AAT 2-degree-field survey, and by the Sloan Digital Sky Survey, which aims to measure one million galaxy redshifts over a contiguous area of π sr in the northern sky (see Fig. 5 [12]). This survey will use a 640-fiber spectrograph to accumulate redshifts at an unprecedented rate. A concurrent photometric survey will measure angular positions for roughly 50 million galaxies.

Below, we will return to discuss the statistical information about large-scale structure—that is, the departures from homogeneity and isotropy—that can be extracted from such surveys, and its use as a probe of models for the early universe.

2.2 Hubble's Law

The second piece of evidence for homogeneous and isotropic expansion is Hubble's law: if the universe remains homogeneous and isotropic over time, the distance $l(t)$ between any two fundamental comoving observers (each of whom defines the local cosmic rest frame) satisfies $l(t) = l_0 a(t)$, where l_0 is the initial separation and $a(t)$ is the universal scale factor. Then the relative speed of one observer with respect to the other is given by

$$v(t) = dl/dt = \dot{a}l_0 = (\dot{a}/a)l(t) \equiv H(t)l(t) \quad , \quad (2.1)$$

where a dot denotes derivative with respect to time t . For observers sufficiently nearby that the light travel time l/c is small compared to the time over which $H(t)$ changes appreciably, we can replace H by its present

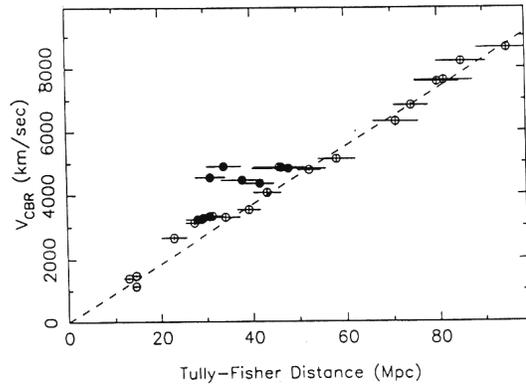


Figure 6: Hubble's law: galaxy recession speed is shown against distance inferred from the Tully-Fisher relation [13].

value, $H(t_0) \equiv H_0$, where subscript 0 denotes the present epoch. Relation (2.1) then reproduces Hubble's observation in 1929 that the recession speed of a galaxy is proportional to its distance from us. The Hubble parameter, $H_0 = (\dot{a}/a)_0$, measures the expansion rate; it has been estimated by a variety of methods to be in the range 40–100 km/sec/Mpc (it is conventional to write this as $H_0 = 100h$ km/sec/Mpc, with $0.4 < h < 1$).

A recent example of this is shown in Fig. 6, from [13]; each circle indicates a cluster of galaxies, with distance estimated using the Tully-Fisher relation for spiral galaxies and recession velocity inferred from the redshift; the horizontal error bars indicate the $1 - \sigma$ spread in inferred distances for a number of galaxies per cluster. (The closed circles around 40 Mpc indicate clusters in the vicinity of the Great Attractor region.)

Photon wavelengths are stretched and their frequencies redshifted by the expansion, $\lambda(t) \sim \nu^{-1} \sim a(t)$. This leads to the definition of redshift z :

$$1 + z(t_e) = \frac{\lambda_{obs}}{\lambda_{em}} = \frac{a(t_0)}{a(t_e)} \quad (2.2)$$

is the ratio of the photon wavelength observed at the present (t_0) to its wavelength at emission (t_e). The redshift z thus plays several roles: through the Doppler shift, it is a measure of recession velocity, $v \simeq cz$ for $z \ll 1$; through the Hubble law, it is a measure of distance, $v = cz = H_0 l$; and through eqn.(2.2), the redshift $z(t)$ can be used to characterize a cosmological epoch t . The most distant objects directly observed are luminous quasars, which have been seen out to redshifts approaching $z = 5$. By comparison, the photons in the cosmic microwave background radiation were probably last scattered at a redshift $z \simeq 1000$.

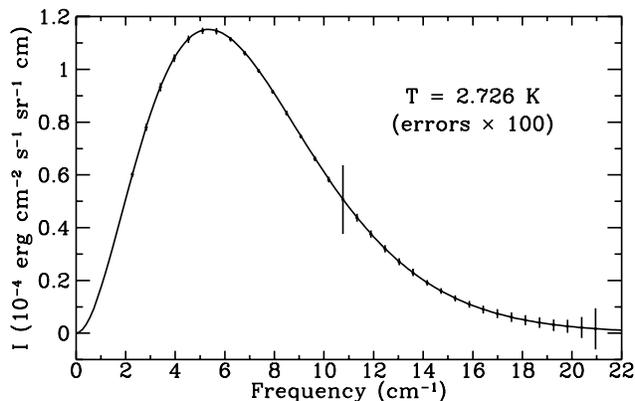


Figure 7: CMBR brightness vs. frequency measured by COBE FIRAS [14]. The error bars on the points have been multiplied by a factor of 100 to make them visible.

2.3 The Hot Big Bang: the CMBR and Primordial Nucleosynthesis

The expanding universe picture naturally traces back to a denser, hotter past—the Big Bang. Two pieces of evidence strongly support this model: the thermal spectrum of the CMBR and the light element abundances in concordance with the predictions from big bang nucleosynthesis. Spectral measurements of the CMBR cover the range of photon wavelengths from roughly 0.05 - 50 cm. The most impressive verification of the thermal Planck spectrum comes from the COBE FIRAS observations in the range 0.05 - 1 cm, shown in Fig. 7, in excellent agreement with a blackbody temperature of $T = 2.726 \pm 0.010$ K [14]. Since the recent universe is apparently transparent to CMBR photons (radio galaxies are seen as point sources out to redshifts of order $z \sim 4$), the CMBR could not have been thermalized at recent epochs and is plausibly the remnant of a much earlier ionized plasma phase when the matter and radiation were in thermal equilibrium.

When the temperature of the universe was comparable to nuclear binding energies, $T \sim 1$ MeV, protons and neutrons fused into nuclei. The predicted abundances of the light elements ^4He , ^3He , D, and ^7Li depend predominantly on the baryon-to-photon ratio at the time of nucleosynthesis. The predictions are in good agreement with the primordial abundances inferred from observations if the fraction of the critical density in baryons, $\Omega_b = \rho_b/\rho_c$, lies in the range $0.01 \lesssim \Omega_b h^2 \lesssim 0.02$ [15].

Taken together, the Hubble expansion, CMBR spectrum and isotropy, and light element abundances indicate that the homogeneous and isotropic Big Bang provides a reliable framework for the evolution of the universe over the approximate temperature range 10 MeV to 2.726 K, corresponding to times from roughly 10^{-2} sec to 13 - 15 Gyr. In constructing early universe mod-

els, one extrapolates from this observational foundation up to temperatures close to the Planck scale, $T \sim 10^{19}$ GeV, corresponding to times $t \sim 10^{-43}$ sec. This would clearly be foolhardy were it not for the fact that these models leave signatures which can be probed by observations of the present Universe. However, these signatures from the early universe are filtered through subsequent evolution that depends upon the global cosmological parameters and especially upon the (dark) matter content of the universe. The current uncertainty in these parameters allows for considerable latitude in the range of allowed models. This provides strong motivation for attempts to substantially improve the observational determination of the cosmological parameters, and there is reason to hope for such improvements as the millennium draws near. The next section provides a brief overview of the current status of the cosmological parameters and dark matter, and some anticipated developments in the near future; again, these topics will be addressed in greater depth in subsequent chapters.

§3 Cosmological Parameters: Expansion Rate and Age of the Universe

The principal observable cosmological parameters of the FRW models are the Hubble parameter, H_0 , the age of the Universe, t_0 , the ratio of the present (non-relativistic) mass density to the critical density (see the Appendix), Ω_0 , the deceleration parameter, $q_0 = -(1 + \dot{H}_0/H_0^2)$, which measures the rate at which the gravitational attraction of the matter is slowing down the expansion, the cosmological constant Λ , and the present CMBR temperature T_0 . Of these, only the CMBR temperature is well determined. The age of the universe is related to the other parameters through an expression of the form $H_0 t_0 = 1.02h(t_0/10^{10}\text{yr}) = f(\Omega_0, \Omega_\Lambda)$, where f is a function of order unity and h parametrizes the uncertainty in the value of the Hubble constant. Thus the *Hubble time* $H_0^{-1} = 9.8 \times 10^9 h^{-1}$ yr sets the timescale for the age of the universe, while the *Hubble length* $cH_0^{-1} = 3000h^{-1}$ Mpc sets the lengthscale for the present observable universe. For a matter-dominated universe with $\Lambda = 0$, f falls monotonically with increasing Ω_0 , and two useful limits are $f(0, 0) = 1$ ($a \sim t$ for $\Omega_0 = 0$) and $f(1, 0) = 2/3$ (since $a \sim t^{2/3}$ for $\Omega_0 = 1$). More generally, over the range $0 < \Omega_0 \leq 1$, $k \leq 0$, $\Omega_0 - 3\Omega_\Lambda/7 \leq 1$, an excellent approximation is [16]

$$H_0 t_0 \simeq \frac{2}{3} \frac{\sinh^{-1}(\sqrt{(1 - \Omega_a)/\Omega_a})}{\sqrt{1 - \Omega_a}} \quad (3.1)$$

where $\Omega_a = \Omega_0 - 0.3(\Omega_0 + \Omega_\Lambda) + 0.3$.

Some examples of the evolution of the scale factor $a(t)$ are shown in Fig.8, including three cases with $\Lambda = 0$:

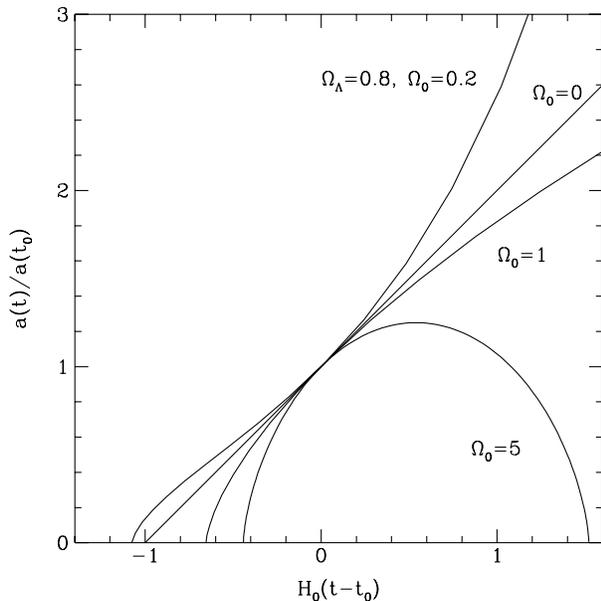


Figure 8: Evolution of the cosmic scale factor in four FRW models: 3 models with $\Lambda = 0$ ($\Omega_0 = 0, 1, 5$) and one with $\Omega_\Lambda = 0.81 - \Omega_0$.

empty ($\Omega_0 = 0$), flat ($\Omega_0 = 1$), and closed ($\Omega_0 = 5$), and one example of a flat ($k = 0$) model dominated by a cosmological constant ($\Omega_\Lambda = 0.8 = 1 - \Omega_0$). This figure displays the decrease in the expansion age $H_0 t_0$ from unity as Ω_0 is increased from zero. It also demonstrates that a significant cosmological constant term can make $H_0 t_0$ large. The solution for the spatially flat ($k = 0$) model with non-zero Λ has the form

$$\frac{a(t)}{a(t_0)} = \left(\frac{\Omega_0}{\Omega_\Lambda} \right)^{1/3} \sinh^{2/3} \left[\frac{3\Omega_\Lambda^{1/2} H_0 t}{2} \right], \quad (3.2)$$

where $\Omega_\Lambda + \Omega_0 = 1$. At late times, it approaches the exponential de Sitter solution, $a(t) \sim e^{\sqrt{\frac{\Lambda}{3}} t}$.

Historically, significant effort was spent trying to measure or constrain the parameters q_0 and Ω_Λ through the classical ‘cosmological tests’, such as the Hubble diagram, angular size as a function of redshift, and galaxy counts as a function of redshift and apparent brightness. For example, to construct the Hubble diagram, one measures the apparent brightness of a well-defined sample of objects (say, the brightest galaxies in clusters) as a function of the object’s redshift; for galaxies of fixed intrinsic luminosity, the scaling of apparent magnitude with redshift is a function of the cosmological parameters. Unfortunately, galaxies at large distances ($z \sim 1$), where the distinction between model parameters becomes observationally significant, are seen when they were much younger than their nearby counterparts, so a model for galaxy luminosity evolution must be used to interpret the results. So far,

these tests have not placed stringent universally accepted constraints on the cosmological parameters. However, significant progress has recently been made in understanding galaxy evolution, and there is hope that the effects of evolution and cosmology might be disentangled in coming years, particularly with the advent of deep spectroscopic surveys planned for large telescopes (such as the DEEP project on the Keck 10m telescope). Recently, it has been pointed out [17] that the probability that a quasar at a given redshift is gravitationally lensed by a foreground galaxy is a sensitive test for the cosmological constant: in models with $\Omega_\Lambda > 0$, one generally expects a higher lensing probability. Based on current surveys for lensed quasars, such as the HST Snapshot Survey, the bound $\Omega_\Lambda \lesssim 0.6 - 0.8$ has been inferred in the case of a spatially flat ($\Omega_\Lambda + \Omega_0 = 1$) universe[18, 19].

3.1 H_0 and the Distance Scale

The Hubble parameter relates the observed recession velocity v_r or redshift z of a galaxy to its distance d : for $v_r \ll c$, the recession velocity is $v_r = cz = H_0 d + v_p$, where v_p is the peculiar radial velocity of the galaxy with respect to the Hubble flow, usually assumed to arise from gravitational clustering. Galaxy redshifts can be measured quite accurately, so all the difficulty in determining H_0 resides in finding reliable distance indicators for extragalactic objects at distances large enough that the Hubble term dominates over the peculiar motion. Observed peculiar velocities are typically of order several hundred km/sec, so that distance measurements well beyond 50 Mpc or more are required for reasonable accuracy.

A wide variety of techniques has been used to establish an extragalactic distance scale[20], and this is reflected in the spread of results for H_0 , roughly 40 – 100 km/sec/Mpc. Distance estimates made using methods such as the Tully-Fisher relation between 21-cm rotation speed and infrared luminosity for spiral galaxies, calibrated by observations of Cepheid variable stars in several nearby galaxies, have yielded high values for the expansion rate, roughly $H_0 = 80 \pm 10$ km/sec/Mpc. Two newer methods, planetary nebula luminosity functions[21] and galaxy surface brightness fluctuations[22] yield values for H_0 in this range as well, and are being further developed. Methods using Type Ia supernovae as standard candles have yielded low values, $H_0 \simeq 50 \pm 10$ km/sec/Mpc, although a recent variation of this method using light curve shapes finds[23, 24] $h \simeq 0.65$. SNe Ia are thought to be the explosions of white dwarfs which accrete matter from binary companions until they reach the Chandrasekhar mass, and there is some evidence that they form a homogeneous class; they also have the advantage that they can be observed to great distances.

There are also a variety of methods being employed

to measure the distances of extragalactic objects directly, bypassing the extragalactic distance ladder built up from Cepheids. The expanding photosphere method[25] has been used to determine the distances to a number of type II supernovae at large distances, generally yielding good agreement with the Tully-Fisher distances for these galaxies. Other ‘direct’ methods which hold future promise include measurement of the Sunyaev-Zel’dovich effect, due to the Compton upscattering of CMBR photons by hot gas in rich clusters[26, 27], and the differential time delay between images in gravitationally lensed quasars[28].

Recently, observations at the CFHT telescope[29] and with the Hubble Space Telescope (HST)[30] have found respectively 3 and 20 Cepheids in two galaxies in the Virgo cluster. The inferred distance to Virgo (assuming these galaxies are close to the center of the cluster) are in accord with the high values of the Hubble constant, $h \simeq 0.8 \pm 0.17$. In the future, HST observations of Cepheids in other nearby galaxies (as well as other Virgo cluster galaxies) which can be used as secondary distance calibrators should reduce this uncertainty by a factor of order two.

3.2 The Age of the Universe, t_0

Three methods have traditionally been used to infer the age of the Universe, t_0 . Nuclear cosmochronology is based on radioactive dating of r -process elements, that is, heavy elements formed by rapid neutron capture, most probably in supernovae. The element ratios Re/Os and Ur/Th generally indicate $t_0 = 10 - 20$ Gyr[31], with a large uncertainty due to the unknown element formation history (e.g., the star formation rate over time). A second method involves the cooling of white dwarfs: when low-mass stars exhaust their nuclear fuel, they become degenerate white dwarfs, gradually cooling and becoming fainter. The number of white dwarfs as a function of luminosity drops dramatically for $L_{wd} < 3 \times 10^{-5} L_{\odot}$, suggesting that there has not been sufficient time for them to fade below this value (assuming the drop is not a selection effect). Coupled with models of white dwarf cooling, this implies that the age of the galactic disk is about $t_0 \simeq 9 - 10 \pm 2$ Gyr[32, 33].

The most extensively studied technique for constraining t_0 is the determination of the ages of the oldest globular clusters in the galaxy. When stars finish burning hydrogen, they turn off the main sequence, characteristically reddening and brightening. By observing the color-magnitude (color vs. apparent brightness) diagram for a cluster, one can determine the apparent brightness of stars in the cluster that are now leaving the main sequence. Knowing the distance to the cluster then gives the absolute luminosity of stars at the turn-off. On the other hand, stellar evolution theory relates stellar lumi-

nosity to the time a star spends on the Hydrogen-burning main sequence. Thus, the measurement of the main sequence turn-off luminosity gives an estimate of the age of the cluster. The turn-off luminosity in the oldest globular clusters in the galaxy is slightly below that of the sun, yielding the age estimate $t_{gc} = (13 - 15) \pm 3$ Gyr[34]. In practice, cluster ages are determined by comparing the luminosity ratio of horizontal branch stars to those at the main-sequence turn-off. Theoretical uncertainties include the degree of mass loss, the role of convection, and the initial Helium abundance, but it appears very difficult to bring the oldest globular ages down below 11-12 Gyr [35]. A significant observational source of error for the method is the uncertainty in the distances to the globular clusters, determined by RR Lyrae stars. Observations with the corrected Hubble Space Telescope mirror may reduce the observational uncertainty in t_{gc} to as little 10%.

3.3 The ‘Age Crisis’

It is useful to place these numbers for the cosmological parameters in theoretical context. As Fig. 8 shows, cosmological evolution depends on the product

$$H_0 t_0 = 1.06 \left(\frac{H_0}{80 \text{ km/sec/Mpc}} \right) \left(\frac{t_0}{13 \text{ Gyr}} \right) . \quad (3.3)$$

If an extended period of inflation took place in the early universe, then the spatial geometry should now be observationally indistinguishable from the spatially flat Einstein-de Sitter model. If the cosmological constant vanishes, spatial flatness then implies the critical matter density, $\Omega_0 = 1$ (with the concomitant requirement of non-baryonic dark matter—see below), and $H_0 t_0 = 2/3$. Comparing with (3.3), this yields an uncomfortably low age for globular clusters unless $H_0 \lesssim 50$ ($t_0 \gtrsim 13$ Gyr) and is definitely problematic unless $H_0 < 65$ ($t_0 > 10$ Gyr), still on the low side of the majority (but not all) of the H_0 observations.

As shown in Fig. 8, the problem is alleviated if the cosmological constant is non-zero, which opens up the possibility that $H_0 t_0 > 1$. In the spatially flat model, $\Omega_0 + \Omega_{\Lambda} = 1$, the most stringent constraint on the cosmological constant appears to be that from gravitational lens statistics, $\Omega_{\Lambda} \lesssim 0.7$. For $\Omega_{\Lambda} = 1 - \Omega_0 < 0.7$, eqn.(3.2) implies $H_0 t_0 < 0.96$. Thus, from (3.3), with a cosmological constant saturating this bound, the globular cluster age $t_0 \geq 13$ Gyr ($t_0 > 10$ Gyr) implies $H_0 < 72$ km/sec/Mpc ($H_0 < 94$ km/sec/Mpc). The implied value of $\Omega_0 \gtrsim 0.3$ is consistent with dynamical mass estimates from galaxy clusters (within the errors) and with most of the dynamical estimates from large scale flows. However, the fact that we would be observing the universe just at the epoch when Ω_0 is comparable to Ω_{Λ} might seem to beg for explanation.

A third possibility is that inflation did not occur (or last sufficiently long) and that we live in an open, low-density (perhaps purely baryonic) universe with negligible Λ . As Fig. 8 shows, this also helps alleviate the ‘age problem’ somewhat compared to the Einstein-de Sitter model. Dynamical mass estimates suggest that $\Omega_0 \gtrsim 0.1$, which implies $H_0 t_0 \lesssim 0.9$ for the open model. Taking $t_0 > 13$ Gyr as before, this implies $H_0 < 68$ km/sec/Mpc. The challenge in this case is to form large-scale structure without violating CMBR anisotropy constraints and to account for the large-scale peculiar velocities. In this case, it also would be a ‘coincidence’ that we live just at the epoch when the curvature term in (A.5) is becoming appreciable compared to the matter term. Clearly the existence or severity of the age problem must be weighted by the present *uncertainty* in the cosmological parameters. As we approach the millennium, we face the exciting prospect that astronomical observations may be able to pin down these parameters and thus determine which of the above (or none of the above!) scenarios corresponds to our Universe.

§4 Dark Matter and Ω

The evolution of large-scale structure, and therefore the connecting thread between the early universe and the observed clustering of galaxies, depends critically on the matter content of the universe. The density of luminous matter, that is, of matter associated with typical stellar populations in galaxies, is inferred to be $\Omega_{lum} \sim 0.007$. However, it is well known that the luminous parts of galaxies are not the whole story: there is strong evidence from flat spiral galaxy rotation curves, from a variety of observations of galaxy clusters, and from large-scale peculiar motions that there is substantially more mass in these systems than can be attributed to the luminous matter: the bulk of the matter in the universe is *dark*.

In a typical spiral galaxy, the luminous components (the bulge and the disk) generally cannot account for the observed extended flat rotation curve traced by stars and gas beyond several kpc from the galactic center. To reproduce the velocity observations, one posits a quasi-spherical dark halo truncated at some large radius. If these systems are typical of the universe, this provides a lower bound of roughly $\Omega_0 \gtrsim 0.02 h^{-1}$ for the matter associated with galaxy halos.

In clusters of galaxies, there are now a variety of methods for estimating cluster masses. The traditional dynamical method, first used by Zwicky in the 1930’s, applies the virial theorem to the measured cluster galaxy velocity dispersion, and typically yields an equivalent $\Omega_0 \sim 0.1 - 0.2$. However, this method assumes that galaxies trace the underlying cluster mass and that the galaxy velocity dis-

tribution is isotropic. Independent information on the dark matter distribution in clusters comes from the giant luminous arcs and arclets, high-redshift galaxies gravitationally lensed by rich foreground clusters into extended banana-shaped images [36, 37]. These estimates are generally in reasonable agreement with the cluster masses inferred from the virial theorem in the cores of rich clusters. More recently, this method has been extended into the ‘weak lensing’ regime, in which one studies the shear distortion pattern induced by a cluster statistically by measuring the image shapes and orientations of a large number of faint background galaxies [38]. This allows one to probe the cluster mass distribution over larger scales and in clusters too weak to produce arcs. In some cases, the weak lensing masses are consistent with the virial masses, while in others the lensing-inferred masses appear to be several times larger. This relatively new method holds considerable promise for the future, as large-area imaging cameras are installed on 4-m class telescopes at sites with excellent seeing. Finally, X-ray satellites have begun to map the density and temperature profiles of the hot gas which permeates many clusters. If the gas is in hydrostatic equilibrium, it can be used to directly trace the cluster mass distribution. Cluster masses inferred by this method are generally comparable to the virial estimates. The X-ray observations also indicate that clusters are surprisingly baryon-rich: the gas constitutes typically $(5 - 10)h^{-3/2}\%$ of the inferred binding mass within approximately $1h^{-1}$ Mpc of the center of a rich cluster like Coma. If this ratio is representative of the baryon mass fraction of the universe, then the nucleosynthesis bound on Ω_B would imply [39] the *upper* limit $\Omega_0 \lesssim (0.2 - 0.4)h^{-1/2}$, well below closure density. As the study of clusters matures, it will be interesting to see whether this implies a sub-critical universe or that clusters are more complex than previously thought.

On still larger scales, peculiar velocities (i.e., deviations from the Hubble flow) have been used to infer the cosmic density [40]. This idea has been applied in a number of ways. One method is to compare samples of the density perturbation field and the divergence of the peculiar velocity field covering the same volume; assuming they arise gravitationally, the proportionality between them depends on the rate of growth of the density fluctuations, which in turn depends on Ω_0 . On very large scales, linear perturbation theory is a reasonable first approximation, and one finds the relation

$$\nabla \cdot \vec{v}_p = -H_0 \Omega_0^{0.6} \delta \quad , \quad (4.1)$$

where the density field $\delta(\vec{x}) = (\rho(\vec{x}) - \bar{\rho})/\bar{\rho}$, and the perturbation growth rate enters through $d \ln \delta / d \ln a \simeq \Omega_0^{0.6}$. If one expresses distances d in terms of their equivalent Hubble velocities, $v = H_0 d$, then H_0 drops out of Eqn.(4.1),

so the uncertainty in the Hubble parameter does not undermine this method.

The topographic correlation between the observed galaxy field (from redshift surveys) and the density field inferred from galaxy peculiar velocity surveys suggests that galaxies do broadly trace the mass distribution on large scales, in the sense that the local galaxy density increases approximately linearly with the local velocity divergence. Nevertheless, the galaxy distribution may be ‘biased’ with respect to the mass. In the simplest linear bias model, the smoothed galaxy and mass density fields are assumed to be proportional, $\delta_{gal}(\vec{x}) = b\delta(\vec{x})$, where b is the bias factor, taken to be constant for a given class of galaxies. Since δ_{gal} is what is observed, the comparison between the galaxy and velocity fields actually constrains the combination $\Omega_0^{0.6}/b$, where the bias factor refers, e.g., to galaxies selected from the IRAS catalog. Recent determinations have found $\Omega_0^{0.6}/b \sim 0.5 - 1$; for a bias factor of order unity, this is consistent with $\Omega_0 \sim 0.3 - 1$. It is well to keep in mind, however, that biasing is presumably a complex process associated with the non-linear stages of galaxy formation, so that the proportionality of the galaxy and density fields may be non-linear and/or scale-dependent. In addition, even with substantial smoothing of the density field, the perturbation amplitude in many regions is not small, and corrections to the linear theory must be taken into account.

An alternative incarnation of this method does not use the peculiar velocity field directly (which requires independent distance measurements for the galaxies) but makes use of the fact that peculiar velocities distort the shape of the galaxy power spectrum in redshift surveys. Measuring the anisotropy of galaxy clustering in redshift surveys then yields an estimate of $\Omega_0^{0.6}/b$; results from, e.g., the IRAS 1.2 Jy survey have recently come in at the low end of the above range,[41] but the sampling errors due to the relatively small surveyed volume are still large. The accuracy of this method to determine Ω_0 should improve substantially with the next round of large-scale redshift surveys. In addition, measurement of the Sunyaev-Zel’dovich effect in a number of clusters at moderate redshift will probe the peculiar velocity field over very large scales.

The observational picture on Ω_0 is currently unclear: the determinations on small scales (clusters and below) consistently yield $\Omega_0 \sim 0.1 - 0.2$ (with the possible exception of weak lensing), while dynamical determinations on large scales show large scatter, $\Omega_0 \sim 0.3 - 1$. Either way, there is a strong dynamical case that over 90 % of the mass in the universe is dark. As the dynamical evidence for the existence of dark matter continues to mount, and as surveys begin to map the distribution of dark matter in different systems and over various scales, it becomes increasingly urgent to determine its composition. In short,

we know dark matter exists, we are making progress in understanding *where* it is and how much of it there is, but we now need to know *what* it is. In the complex lexicon of particle astrophysics, the possibilities are divided into two categories: baryons and non-baryonic matter.

4.1 Baryonic Dark Matter

The density of baryons in the universe is an important ingredient in structure formation models, particularly in regard to theoretical predictions for small-angle CMBR anisotropies. Concordance of the observed light element abundances with the predictions of primordial nucleosynthesis restricts the baryon density to the range $0.01 < \Omega_B h^2 < 0.02$ [15]. This range may shift or be narrowed as deuterium measurements from Lyman- α clouds accumulate. The corresponding baryon to photon ratio, $n_B/n_\gamma = 2.7 \times 10^{-8} \Omega_B h^2 \simeq 3 \times 10^{-10}$, should ultimately be a predictable number from models for the baryon asymmetry. Recent theoretical work has concentrated on the possibility of generating this asymmetry in a (weakly) first-order electroweak phase transition. Comparison of the nucleosynthesis range for Ω_B with the density of luminous matter Ω_{lum} suggests that some or all of the baryons are dark or in underluminous populations. On the other hand, comparison with the density inferred from galaxy rotation curves, $\Omega_0 \sim 0.02h^{-1}$, shows that baryons could constitute some or all of the dark matter in galaxy halos.

One possibility for dark baryons in halos are degenerate brown dwarfs, substellar ($M < 0.08M_\odot$) objects which did not reach sufficiently high temperature to burn hydrogen. Currently three independent groups are searching for galactic brown dwarfs (which have been dubbed MACHOS, for massive compact halo objects); the signature is a gravitational microlensing event, in which a background star, say in the LMC or the bulge of the Milky Way, symmetrically brightens and fades as a MACHO passes near its line of sight.[42] Although they are distinguishable from ordinary variable and flare stars by the time symmetry and achromaticity of the light curve, such microlensing events are intrinsically rare—the probability that a given background star in the LMC is lensed by a halo-filling MACHO population is $\tau \sim 5 \times 10^{-7}$, so a large number of stars must be accurately monitored. The American-Australian MACHO project, the French EROS collaboration, and the Polish-American OGLE collaboration first announced the discovery of microlensing events in the fall of 1993 [43, 44, 45]. As of early 1995, they have a combined total of more than 50 microlensing events; the vast majority are in the direction of the galactic bulge, and only several are toward the LMC. [46, 47] The preliminary indication from the low rate toward the LMC is that the Milky Way halo is not predominantly composed of MACHOs in the mass range to which the experiments

are sensitive.

4.2 Non-baryonic Dark Matter

If $\Omega_0 > \Omega_B$, as the dynamical observations on scales larger than clusters suggest, some form of non-baryonic matter must be invoked to make up the balance. Particle physics theories provide a host of weakly interacting particles as dark matter candidates. It is convenient to distinguish two broad classes of non-baryonic dark matter, hot and cold, on the basis of their clustering properties. The prototypical hot dark matter candidate is a light neutrino with mass $m_\nu \simeq 20$ eV (more generally, the mass density of light neutrinos is $\Omega_\nu h^2 = (m_\nu/92\text{eV})$). Since they are relativistic until relatively recent epochs, light neutrinos would free-stream out of and damp out density perturbations up to the scale of galaxy clusters, unless there is some mechanism (such as topological defects) which continually seeds the formation of perturbations on small scales. Galaxies would form after clusters via fragmentation ('top down'). Due to phase-space constraints[48], light neutrinos would not cluster significantly on the scale of galaxies: baryons would constitute the predominant dark matter in galaxy halos, while neutrinos would dominate in clusters. Cold dark matter, on the other hand, is non-relativistic at all epochs of interest for structure formation, so it has negligible free-streaming length—it clusters on all scales. In cold dark matter models, structure generally forms hierarchically, with smaller clumps merging to form larger ones ('bottom up'). It is also worth noting the recently surging popularity of a mix 'n match scenario: a combination of 70-80% cold and 20-30% hot dark matter (say, with $m_\nu \simeq 5$ eV) produces a favorable spectrum of large-scale density perturbations in the context of inflation, with some advantages over pure cold dark matter.

The theoretically favorite candidates for cold dark matter are weakly interacting massive particles (WIMPs), with masses generally in the range 20 – 150 GeV, and the axion, an ultra-light scalar with a mass of order 10^{-5} eV. The most attractive WIMP candidate is the neutralino, the lightest supersymmetric fermionic partner of the standard model bosons; its weak annihilation rate in the early universe naturally leaves it with an abundance comparable to that corresponding to the present critical density. The axion is the pseudo-Nambu-Goldstone boson associated with spontaneous breakdown of a global $U(1)$ symmetry (the Peccei-Quinn symmetry) introduced to explain why the strong interactions conserve CP. The global symmetry is spontaneously broken at some large mass scale f_{PQ} , through the vacuum expectation value of a complex scalar field, $\langle \Phi \rangle = f_{PQ} \exp(ia/f_{PQ})/\sqrt{2}$. At energies below the scale f_{PQ} , the only relevant degree of freedom is the massless axion field a , the angular mode around the

bottom of the Φ potential. At a much lower energy scale, $\Lambda_{QCD} \sim 100$ MeV, the symmetry is *explicitly* broken when QCD becomes strong, and the axion obtains a periodic potential of height $\sim \Lambda_{QCD}^4$. In 'invisible' axion models with Peccei-Quinn scale $f_{PQ} \sim 10^{12}$ GeV, the resulting axion mass is $m_a \sim \Lambda_{QCD}^2/f_{PQ} \sim 10^{-5}$ eV and $\Omega_a \sim 1$. Although light, invisible axions interact so weakly, with cross-section $\sigma \sim 1/f_{PQ}^2$, that they were never in thermal equilibrium: they form as a cold Bose condensate.

Accelerator searches for neutrino mixing and beta-decay experiments on neutrino mass should provide useful constraints on the possibility of neutrino dark matter. Active experimental efforts are also underway to detect both WIMPs and axions. Direct WIMP detection looks for the signals produced when a halo WIMP collides with a nucleus in a kg-size cryogenic crystal, depositing of order 10 keV in ionization and phonons (detector schemes based on scintillation and excitations of superfluids and superconductors are also being developed). Indirect WIMP detectors search for high energy neutrinos produced when WIMPs annihilate in the Sun and the Earth; large underground or underwater detectors currently in place or under development with sensitivity to WIMP annihilations include SuperKamiokande, MACRO, AMANDA, and DUMAND. Accelerator searches for supersymmetry also will constrain the neutralino parameter space. A large-scale axion search based at Livermore will search for resonant conversion of halo axions to microwave photons in a cavity with a strong magnetic field. A scaled-up version of an idea originally proposed by Sikivie, this detector should approach the cosmologically interesting region of axion couplings for the first time[49].

Finally, given the paucity of direct evidence for dark matter, it is probably healthy to keep an open mind to alternatives. While it is natural to ascribe flat galaxy rotation curves and large cluster velocity dispersions to unseen matter, Milgrom[50] and others have argued that they may instead signal a breakdown of Newton's law of inertia at very low acceleration. The extent to which Milgrom's modified Newtonian dynamics (MOND) accounts for all the phenomena normally imputed to dark matter is controversial, and a full theory with which one could explore cosmology has been lacking.

§5 The Early Universe

The thermal spectrum of the CMBR and the concordance of primordial nucleosynthesis predictions with light element abundances strongly suggest that the early universe can be accurately described as a dilute, adiabatically expanding gas of relativistic particles in local thermal equilibrium, with instantaneous temperature $T(t) \simeq 1$ MeV $(t/\text{sec})^{-1/2}$. While this may reflect conditions over

$T(\text{GeV})$	$t(\text{sec})$	Event
10^{19}	10^{-43}	Planck era: quantum gravity
10^{15}	10^{-35}	Grand unification: inflation, topological defects
10^3	10^{-12}	Supersymmetry, technicolor
10^2	10^{-10}	Electroweak transition: $SU(2) \times U(1) \rightarrow U(1)_{em}$
0.2	$10^{-4.5}$	Quark-hadron transition; chiral symmetry breaking
10^{-3}	1	Neutrino decoupling
5×10^{-4}	4	e^+e^- annihilation
7×10^{-5}	200	Big Bang nucleosynthesis
10^{-9}	5×10^{11}	Matter-radiation equality
3×10^{-10}	10^{13}	H recombination; photon decoupling
6×10^{-13}	10^{17}	Non-linear structures form

Table 1: Thermal history of the universe.

much of the early history of the universe, we know it is not the full story, because a gas in thermal equilibrium is featureless and structureless. Although thermal equilibrium is a good starting point, the interesting epochs in cosmic history—the ones which leave potentially observable signatures or relics, such as the light elements, the baryon asymmetry, and particle dark matter—are those in which a particle species i goes out of thermal equilibrium, because the rate of interactions keeping it in equilibrium, Γ_i , falls below the expansion rate $H(t)$. This happens at the freeze-out temperature T_F , defined implicitly by $(\Gamma/H)_{T_F} \simeq 1$. It is worth noting that there may also be particle relics, such as magnetic monopoles and axions, which were never in thermal equilibrium.

A timeline for the early universe with some important epochs delineated is shown in Table 1. The CMBR probes conditions back to the time of photon decoupling, while the light element abundances probe conditions at temperatures comparable to nuclear binding energies, $T \sim \text{MeV}$. Going backward in time, QCD predicts that chiral symmetry should be broken at a temperature of order 100 MeV. At about that time, quarks should also become confined inside hadrons, but it is not clear if this is a smooth (second-order) transition or a first-order transition involving the release of latent heat and the nucleation of hadron bubbles. If the quark-hadron transition is first-order, the resulting inhomogeneous baryon distribution might leave observable signatures in the light element abundances.

Going to earlier times, the electroweak symmetry should be restored at a temperature of order 100 GeV. The dynamics of this transition is currently an area of active investigation: if it is a first-order transition, it is possible that the resulting non-equilibrium conditions were ripe for the generation of the baryon asymmetry. If baryons and anti-baryons had been present in exactly equal numbers in the early universe, their annihilation would eventually have driven the baryon to photon ra-

tio down to a value many orders of magnitude smaller than the observed value of $n_B/n_\gamma \simeq 3 \times 10^{-10}$. Thus, when the baryons were relativistic, there must have been a small asymmetry between the density of baryons and anti-baryons, $\eta = n_B/n_b = (n_b - n_{\bar{b}})/n_b \sim \eta$. As first pointed out by Sakharov, the generation of such a baryon asymmetry requires baryon-number and CP-violating interactions as well as a departure from thermal equilibrium (since $n_b = n_{\bar{b}}$ in equilibrium). It is currently thought that baryogenesis takes place at the GUT or electroweak eras.

Going back earlier than 10^{-10} sec, we must invoke physics beyond the standard electroweak model, and the events become increasingly speculative. Particle physics models suggest there may be new physics lurking at the TeV scale—perhaps supersymmetry, technicolor, or various extensions of the standard model. In the simplest grand unified theories, the strong and electroweak interactions are unified (the symmetry between them restored) at an energy scale of order 10^{15} GeV. A cosmological phase transition at that epoch might lead to inflation or to the generation of topological defects such as monopoles, cosmic strings, or textures; the resulting density and gravitational wave perturbations produced could provide the seeds for large-scale structure and leave a signature in the CMBR anisotropy. Classical cosmology runs into a wall at the Planck era, $t \sim 10^{-43}$ sec: at that epoch, quantum fluctuations in the spacetime metric are expected to be large, and a quantum theory of gravity is required. If superstrings provide the fundamental description of nature, inherently stringy effects would become important around that scale.

One of the major developments in theoretical cosmology over the last fifteen years is the discovery of scenarios for the generation of primordial density fluctuations. Currently there are two early universe paradigms for the origin of perturbations: inflation and topological (or non-

topological) defects. Both scenarios predict large-angle anisotropies in the CMBR and produce large-scale structure, but their predictions differ in details that are under active study.

§6 The Inflationary Universe

The inflationary scenario originally arose out of the attempt to solve several puzzles encountered when one extrapolates the standard cosmology back to the very early universe, in particular the horizon and flatness problems and the overabundance of magnetic monopoles in grand unified theories [51]. The horizon problem arises from the observation that the CMBR is isotropic over large angular scales (see Fig. 1): in the standard cosmology, regions corresponding to separations larger than about one degree were not yet in causal contact when the photons last scattered (at $t \sim 10^{13}$ sec – see Table 1). Yet the microwave sky is remarkably isotropic over all scales larger than a few arcminutes, to of order one part in 10^5 . The flatness problem is the puzzle of why the spatial curvature term in (A.5) does not presently dominate over the matter-density term by a large margin: if $\Lambda = 0$, we can rewrite (A.5) suggestively as $|\Omega(t) - 1|^{-1} = a^2(t)H^2(t) = \dot{a}^2$, which is a decreasing function of time if the universe is dominated by a fluid with non-negative pressure (or, more precisely, from (A.3) by a fluid with $p > -\rho/3$). The natural timescale for the universe to become curvature-dominated is the Planck time, $t_{Pl} \sim 10^{-43}$ sec, yet our universe is still not strongly curvature-dominated at $t_0 \sim 10^{17}$ sec $\sim 10^{60}t_{Pl}$.

It must be stressed that these puzzles are not inconsistencies of the standard cosmological model. Rather they point to features of the observed universe which the standard model does not explain and which moreover appear highly unlikely when the standard cosmology is embedded in a somewhat larger class of cosmological models: the present nearly homogeneous and spatially flat state of the Universe appears to be very sensitive to the initial conditions. Inflation was designed to reduce this sensitivity by widening the class of initial conditions which evolve to a nearly homogeneous and spatially flat state within the observable universe. This is the sense in which inflation is said to ‘solve’ the horizon and flatness problems.

Inflation is a hypothetical early epoch in cosmic history during which the scale factor of the universe accelerates, $\ddot{a} > 0$. This rapid expansion drives the spatial curvature toward zero, $\Omega(t) \rightarrow 1$. To solve the horizon and flatness problems, this epoch must last long enough for the scale factor to grow by a factor of e^{N_e} , where the number of inflationary e-folds of growth satisfies $N_e > N_{min} \simeq 60 \pm 10$ (N_{min} depends slightly upon initial conditions and logarithmically upon the reheat temperature after inflation). In most inflation models, the growth of the scale factor

during inflation is quasi-exponential, $a \sim e^{tH(t)}$, where $H(t)$ decreases on a timescale long compared to the expansion time, $|\dot{H}| \ll H^2$. Because of the quasi-exponential behavior, unless N_e is almost exactly equal to N_{min} , a sufficiently long period of inflation implies that the present universe should be observationally indistinguishable from the spatially flat model, i.e., $\Omega_0 + \Omega_\Lambda = 1 + \epsilon$, where $\epsilon = \mathcal{O}(e^{-2(N_e - N_{min})})$ is exponentially small. (Recently, there has been some interest in ‘open’ inflation models, where one arranges $\Omega < 1$ by having $N_e = N_{min} \pm$ a few, but this requires fine tuning either of initial conditions or of scalar field parameters [52, 53, 54]. This fine tuning has nothing to do with the small coupling problem (see below), although the latter, which is a feature of *all* inflation models, is often inaccurately stated to be a fine tuning problem.)

Acceleration of the scale factor requires that the energy density be dominated by a fluid with equation of state $p < -\rho/3$; this is most simply achieved with a scalar field ϕ , the *inflaton*, undergoing a classical ‘slow-rollover’ phase [55, 56], where its stress-energy-momentum is dominated by its potential, $V(\phi)$. For a spatially homogeneous scalar field $\phi(t)$, the energy density and pressure are given by

$$\begin{aligned} \rho &= \frac{\dot{\phi}^2}{2} + V(\phi) \\ p &= \frac{\dot{\phi}^2}{2} - V(\phi) \end{aligned} \tag{6.1}$$

Thus, accelerated expansion takes place if ϕ dominates the stress-energy of the universe and if $V(\phi) > \dot{\phi}^2$; if this inequality is strongly satisfied, $V(\phi) \gg \dot{\phi}^2$, then $p_\phi \simeq -\rho_\phi = -V(\phi)$, and the expansion is approximately exponential.

6.1 Fluctuations from Inflation

Soon after its invention, it was realized that inflation could in principle provide one of the holy grails of cosmology: a causal mechanism for the origin of density fluctuations that later grow to form large-scale structure. Curvature perturbations cannot be causally generated on scales larger than the instantaneous Hubble radius H^{-1} , because of local energy conservation. Perturbations on galaxy scales today were larger than the Hubble radius at redshifts larger than $z \sim 10^6$ ($T \gtrsim 100$ eV). If there is no new physics at atomic scales to produce fluctuations at late times after they enter the Hubble radius, then in the standard cosmology (without inflation) they must be postulated as initial conditions. (Note that this argument applies only to curvature perturbations: an important loophole is the early generation of isocurvature perturbations, the mechanism of defect-mediated structure formation.)

An early epoch of inflation alters this situation: during a phase of accelerated expansion, the Hubble radius H^{-1} grows more slowly than the scale factor. Thus a perturbation of comoving wavelength λ can be created causally when its physical scale $\lambda_{phys} = a(t)\lambda$ is less than H^{-1} during inflation, expand outside the Hubble radius at some time $t_A(\lambda)$ during inflation, and then eventually ‘re-enter’ the Hubble radius in the recent radiation- or matter-dominated era at a time $t_B(\lambda)$. Scales λ corresponding to the range from the galaxy scale (\sim Mpc) to the current Hubble radius ($3000 h^{-1}$ Mpc) crossed outside the Hubble radius in the interval 50–60 e-folds before the end of inflation, i.e., when $a(t) = a(t_{end})e^{-(50-60)}$.

Quantum fluctuations of the slowly rolling inflaton field provide the dynamical mechanism for generating density perturbations in inflation [57, 58, 59, 60]. Mathematically, one quantizes the spatial fluctuations of the inflaton field (and the corresponding metric perturbation) about its homogeneous classical expectation value. In its vacuum state, such a quantum field has a spectrum of zero-point fluctuations by the Heisenberg uncertainty principle, with an rms amplitude on scale λ at Hubble-radius crossing (t_A) given by $(\Delta\phi)_{\lambda^{-1}=aH} = H/2\pi$. These spatial variations in ϕ correspond to fluctuations in the inflaton energy density, which are stretched by the accelerated expansion and converted into adiabatic density fluctuations in all species during the reheating phase, when ϕ decays at the end of inflation. The resulting perturbation amplitude when a comoving wavelength λ re-enters the Hubble radius is given approximately by

$$\left(\frac{\delta\rho}{\rho}\right)_{t_B(\lambda)} \simeq \frac{(\Delta\phi)_{\lambda^{-1}=aH} V'(\phi)}{\dot{\phi}^2} = \left(\frac{3H^2}{2\pi\dot{\phi}}\right)_{t_A(\lambda)} \quad (6.2)$$

Letting ϕ denote the value of the inflaton when a given scale λ crosses outside the Hubble radius and defining $(\delta\rho/\rho)_{t_B(\lambda)} \equiv A_S(\phi)$, this can be written as

$$A_S(\phi) = \sqrt{384\pi} \left(\frac{V^{3/2}(\phi)}{m_{Pl}^3 V'(\phi)}\right)_{t_A(\lambda)} \simeq \left(\frac{H^2(\phi)}{m_{Pl}^2 |H'(\phi)|}\right)_{t_A(\lambda)} \quad (6.3)$$

where a prime denotes differentiation with respect to ϕ .

Since the period of ten e-folds of the scale factor over which structure on astronomical scales is generated is brief, and the Hubble expansion rate, which sets the scale of the fluctuation amplitude, is slowly varying during inflation ($\dot{H} \ll H^2$), the density perturbation amplitude at Hubble-radius crossing $A_S(\phi)$ is nearly independent of scale λ . The density perturbation amplitude at Hubble crossing ($\lambda_{phys} = H^{-1}$) is essentially the gravitational potential perturbation on that scale, $\delta\Phi_\lambda \sim G\delta M/\lambda \sim (\delta\rho/\rho)_\lambda (\lambda_{phys}/H^{-1})^2$, and it sets the scale for the large-angle CMBR anisotropy through the Sachs-Wolfe effect,

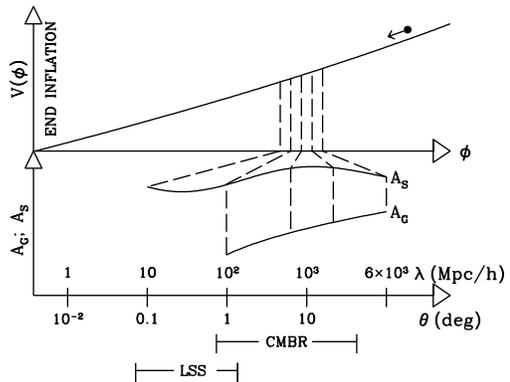


Figure 9: Sketch of an inflaton potential $V(\phi)$ and the scalar $A_S(\phi)$ and tensor $A_G(\phi)$ fluctuation spectra produced by quantum fluctuations. ϕ is the value of the classical field when the comoving scale λ crossed outside the Hubble radius during inflation. Also shown is the corresponding angular scale on the surface of last scattering and the rough angular and lengthscale intervals probed by CMBR anisotropy experiments and galaxy surveys (LSS).

$\delta T/T \sim \delta\Phi$. Thus, inflation predicts that the gravitational potential fluctuations are nearly independent of scale. Such a scale-invariant spectrum was first proposed on different grounds by Harrison, Zel’dovich, and Peebles and Yu ten years before inflation. Considerable excitement was generated when analysis of maps of the CMBR sky from the first year of COBE DMR data yielded a large-angle spectrum consistent with the scale-invariant spectrum [1].

In addition to the scalar metric fluctuations arising from quantum fluctuations in the inflaton field, there are also gravitational wave perturbations associated with quantum fluctuations in the transverse traceless (tensor) part of the metric field [61]. The perturbation equations for the graviton amplitudes $h_{+,x}$ are identical to those for a massless, minimally coupled scalar field with amplitude $\delta\phi = h/\sqrt{16\pi G}$. The resulting amplitude at Hubble-radius crossing is

$$h_{\lambda^{-1}=aH} \equiv A_G(\phi) \simeq \frac{2H(\phi)}{\sqrt{\pi}m_{Pl}} \simeq \frac{V^{1/2}(\phi)}{m_{Pl}} \quad (6.4)$$

and the associated large-angle CMBR anisotropy is of order $\delta T/T \sim A_G(\phi)$. Since $H(\phi)$ does not vary significantly over the observationally accessible range of length-scales, the gravitational wave perturbations are also approximately scale-invariant. A sketch of the correspondence between scalar field evolution and the imprint of large-scale density and gravity wave fluctuations is shown in Fig. 9 (from [62]).

While the nearly scale-invariant spectra of scalar and tensor perturbations are generic features of inflation, the amplitude of the perturbations, from Eq.(6.3) and (6.4), are clearly model-dependent, varying with the scale and

shape of the inflaton potential. Since these amplitudes are now tightly constrained by CMBR anisotropy measurements on large scales, we already have significant constraints on the form of phenomenologically acceptable inflation models (see below).

Although the perturbations from inflation are *nearly* scale-invariant, the (generally small) deviations from scale-invariance are also model-dependent, and it is useful to characterize them in a phenomenological way. Over the range of scales of interest, the scalar and tensor spectra can be approximated by power laws in wavelength,

$$A_S(\phi) \sim \lambda^{(1-n_S)/2} \quad ; \quad A_G(\phi) \sim \lambda^{-n_T/2} \quad (6.5)$$

where the spectral indices n_S and n_T characterize the spectra; for exact scale-invariance, $n_S - 1 = n_T = 0$. The scalar gravitational potential spectrum corresponds to a density fluctuation spectrum (at fixed time), $P_\rho(k) \sim k^{n_S}$. The spectral indices can be calculated as an expansion in the ‘slow-roll’ parameters, defined by

$$\begin{aligned} \varepsilon &\equiv \frac{m_{Pl}^2}{16\pi} \left(\frac{V'(\phi)}{V(\phi)} \right)^2 \\ \eta &\equiv \frac{m_{Pl}^2}{8\pi} \frac{V''(\phi)}{V(\phi)} \end{aligned} \quad (6.6)$$

The conditions $\varepsilon, \eta \lesssim 1$ are sufficient for the scalar field energy-momentum to be dominated by the potential, $V(\phi) \gtrsim \dot{\phi}^2$ and thus for inflation to take place. In terms of the slow roll parameters, the spectral indices are

$$\begin{aligned} n_S &\simeq 1 + 2\eta - 6\varepsilon \\ n_T &\simeq -2\varepsilon \end{aligned} \quad (6.7)$$

Since ε is positive definite, $n_T \leq 0$.

The inflation perturbations can thus be characterized by four numbers: an amplitude and spectral index for the scalar and tensor modes. One way to characterize the amplitudes is by their contribution to the (mean square) quadrupole anisotropy through the Sachs-Wolfe effect, denoted by C_2^S and C_2^T for the scalar and tensor modes respectively. The COBE DMR measurements essentially fix the sum $C_2^S + C_2^T$, leaving three undetermined variables: n_S , n_T , and the relative tensor to scalar contribution to the large-angle anisotropy, $r = C_2^T/C_2^S$. From Eqns.(6.3-6), this ratio can be expressed as

$$r = \frac{C_2^T}{C_2^S} = 14\varepsilon = -7n_T \quad (6.8)$$

so there are only two independent quantities to be determined. Note that n_T cannot be arbitrarily large, for then

the COBE-normalized density fluctuation amplitude is too small to make non-linear structure on galaxy scales. A likelihood analysis of power law spectra using the CMBR anisotropy detected in two years of COBE DMR data finds $n \simeq 1.17 \pm 0.31 (0.96 \pm 0.36)$ including (excluding) the measured quadrupole in the analysis [63], consistent with a nearly scale-invariant spectrum.

6.2 Naturalness and the Small Coupling Problem

The result (6.3) for the density perturbation amplitude provides a severe constraint on the form of the inflaton potential. As an example, consider a chaotic inflation model [64] with a simple quartic potential, $V(\phi) = \lambda\phi^4/4!$. The inflationary slow-roll conditions $\varepsilon, \eta < 1$ are satisfied for $\phi > m_{Pl}\sqrt{3/2\pi}$ and the number of e-folds of accelerated expansion is larger than N_{min} if the initial value of the field satisfies $\phi_i \gtrsim \phi_{min} = 3m_{Pl}$. The density perturbation amplitude on the scale of the present Hubble radius is $(\delta\rho/\rho)_{H_0^{-1}} = (\pi\lambda)^{1/2}(\phi_{min}/m_{Pl})^3 \simeq 50\lambda^{1/2}$. The COBE DMR observation indicates $(\delta\rho/\rho)_{H_0^{-1}} \simeq 10^{-4}$, which implies that the inflaton self-coupling must satisfy $\lambda \lesssim 4 \times 10^{-12}$.

This example illustrates a general conclusion that applies to all inflation models: the inflaton field must be very weakly self-coupled. In different models, this constraint appears in different guises, but it is always present in one form or another: the potential must contain a very small dimensionless number of order 10^{-12} . Attitudes concerning this problem vary widely among theorists. It is often said that $\lambda_\phi \sim 10^{-12}$ constitutes unacceptable ‘fine tuning’, but that is a misapplication of the fine tuning concept in this context: we are not interested in restricting small changes to λ but in ensuring that λ itself is small. To others, it is not an issue of great concern, because we know there exist other small numbers in physics, such as lepton and quark Yukawa couplings $g_Y \sim 10^{-5}$ and the weak hierarchy ratio $M_{weak}/M_{Pl} \sim 10^{-17}$.

Nevertheless, in the context of a particular model, one can ask whether such a small value for λ_ϕ is in principle unnatural. A small parameter λ is said to be ‘technically natural’ if it is protected against large radiative corrections by a symmetry, *i.e.*, if setting $\lambda \rightarrow 0$ increases the symmetry of the system. For example, in this way, low energy supersymmetry might protect the small ratio M_{weak}/M_{Pl} by cancelling boson and fermion loops. Essentially all inflation models that have been proposed in the last decade satisfy this criterion of technical naturalness. However, in models that are only technically natural, the small coupling λ_ϕ , while stable against radiative corrections, is itself unexplained. Technical naturalness is a useful concept for low energy effective Lagrangians,

such as the electroweak theory and its supersymmetric extensions, but it points to a more fundamental level of theory for its origin. For example, the underlying origin of the mass hierarchy M_{weak}/M_{Pl} in supersymmetric theories is thought to be associated with hidden sector physics in supergravity/superstring theories, which start out with no small dimensionless couplings at the Planck scale. Since inflation takes place *relatively* close to the Planck scale, it would be preferable to find the inflaton in such particle physics models which are “strongly natural”, that is, which have no small numbers in the fundamental Lagrangian. In this case, one would expect the inflaton coupling to arise from renormalization group factors, e.g., $\lambda_\phi \sim e^{-n/\alpha}$, where α is a gauge coupling constant and n is a number of order a few. A convincing model with these properties that is also desirable from the viewpoint of particle physics phenomenology has not yet been found. The task of constructing such a theory remains a paramount goal for inflation theorists.

§7 Topological Defects

Spontaneous symmetry breaking, the idea that there are underlying symmetries of nature that are not manifest in the the vacuum state, plays a central role in particle physics models that unify the fundamental interactions. In unified gauge theories—including the standard electroweak model—the underlying symmetry of the Lagrangian is larger than $SU(3)_C \otimes U(1)_{EM}$, the symmetry of the vacuum. However, when the temperature of the universe is higher than the characteristic unification scale, spontaneously broken symmetries are restored. Thus, in the course of cosmic evolution, we expect that there was a series of phase transitions associated with the spontaneous breakdown of gauge (and perhaps global) symmetries (see Table 1).

Depending on the topology of the vacuum manifold, spontaneously broken gauge theories can contain stable, quasi-classical configurations of gauge and Higgs fields such as domain walls, cosmic strings, and magnetic monopoles. For example, when a symmetry group G is broken to a subgroup H , strings form if the fundamental group $\pi_1(G/H)$ (loosely, the group of mappings of the manifold of degenerate vacua onto the circle S^1) is non-trivial, i.e., if the coset space G/H is multiply connected. Topological defects are quasi-localized regions where the gauge and scalar fields depart from the vacuum, and the symmetry is effectively restored in the defect core. Kibble first pointed out that, for particle physics models containing defect solutions, defects would inevitably form in symmetry-breaking phase transitions, since the scale over which the fields can be correlated is limited by the particle horizon at the time of the transition [65]. Due to the

effective symmetry restoration in their cores, defects carry substantial stress-energy-momentum and can have important gravitational effects. They are thus of significant cosmological interest: the strings formed in the breakdown of a class of grand unified theories may act as seeds for the formation of large-scale structure.

Defects such as strings and monopoles can also arise in the spontaneous breakdown of global symmetries. For global defects, however, the gradient energy density of the Higgs field outside the defect core is not compensated by the gauge field, so that the energy density of global defects is not as strongly localized as that of their local defect counterparts. Moreover, global symmetry breaking can give rise to new structures that are not topologically stable but nevertheless may survive for cosmologically long times. For example, in the spontaneous breaking of the global symmetry $O(4) \rightarrow O(3)$, textures form—knots in the Higgs field that arise when the field winds around the three-sphere. These knots are generally formed by misalignment of the field on scales larger than the horizon at the symmetry-breaking phase transition, because of the Kibble mechanism. For the breakdown of larger global symmetries, e.g., $O(N) \rightarrow O(N-1)$ with $N > 4$, there is no topological structure; nevertheless, large-scale spatial gradients in the Higgs field will be formed by the Kibble mechanism, and they can also give rise to density perturbations when a given gradient scale enters the horizon.

In potentially viable defect models, such as gauge cosmic strings and global $O(N)$ models with $N \geq 2$, the population of defects or field gradients approaches a scaling solution, in which the energy density of the fields scales as $\rho \propto t^{-2}$. The resulting spectrum of density perturbations seeded by defects is approximately scale-invariant, and the perturbation amplitude when a given lengthscale crosses the Hubble radius is set by the scale of symmetry breaking M , $(\delta\rho/\rho)_{\lambda=H^{-1}} \sim (M/m_{Pl})^2$. To match the amplitude of observed galaxy clustering, the symmetry breaking scale should be comparable to that expected in grand unified theories, $M \sim 10^{16}$ GeV. A pleasing feature of defect models is that the density fluctuations are independent of dimensionless coupling constants. On the other hand, the cosmological evolution of defects and the manner in which they seed density perturbations is much more complex than in inflation models, and the detailed predictions of these models are still being worked out. For example, in addition to scalar density perturbations, defects generate a spectrum of gravitational waves; however, unlike the case of inflation, these tensor perturbations are correlated with the scalar modes. Also, due to the coherence of the defect fields, the density perturbations they generate are generally non-Gaussian. By contrast, in most inflation models (those involving a single, weakly coupled inflaton field) the primordial fluctuations are Gaussian distributed, because they arise from an es-

entially free quantum field in its vacuum state.

§8 Testing theories of the early Universe

8.1 CMBR anisotropies

The COBE DMR measurements have found anisotropies on large angular scales (the angular resolution of the DMR instruments is several degrees) at approximately the level theoretically expected and with a spectrum roughly consistent with scale-invariant primordial fluctuations. This is an important first test for the primordial gravitational instability picture for structure formation. While it places important constraints on models (e.g., inflation models with excessive tilt, $n_s \lesssim 0.7$, are excluded), it does not yet discriminate strongly between competing models. For example, Figure 10 shows a sample microwave sky generated in the cosmic string scenario; on these large angular scales, the spectrum is broadly consistent with the COBE DMR detection [66]. On smaller scales, the spectrum of inflationary and defect models is expected to differ significantly; in addition, in most defect models, one would expect non-Gaussian features to arise in CMBR maps at arc minute to degree scales.

A number of ground-based and balloon-borne experiments are now probing CMBR anisotropies on smaller angular scales. Although there are still wide variations between different experimental groups, there does seem to be a general trend of a larger anisotropy amplitude on scales of order 1/2 degree than on larger scales. Again, this is expected in most models. In the future, one looks forward to full-sky maps with sub-degree resolution, most probably requiring a satellite mission, in order to probe the anisotropy power spectrum in sufficient detail for model discrimination and to test for non-Gaussian features.

The CMBR temperature autocorrelation function $C(\theta)$ is defined operationally by measuring the temperature fluctuation at some direction (angle) α , $\delta T(\alpha) = T(\alpha) - \langle T \rangle$, multiplying by the temperature at angle $\alpha + \theta$, and averaging over positions α , $C(\theta) = \langle \delta T(\alpha) \delta T(\alpha + \theta) \rangle_\alpha$. It is useful to expand $C(\theta)$ in Legendre polynomials: $C(\theta) = (4\pi)^{-1} \sum_l (2l + 1) C_l P_l(\cos \theta)$. The coefficients C_l are the multipole moments, which specify the angular power spectrum of the temperature fluctuations.

There are some important features to look for in the spectrum of temperature fluctuations. If inflation is correct and the primordial spectrum is approximately scale-invariant, $n_s \simeq 1$, then at large angular scale (multipole $l \ll 100$) the spectrum $l(l + 1)C_l$ should be roughly flat. These scales were outside the Hubble radius at last scattering, and the signal is dominated by the gravitational redshift of photons emerging from gravitational poten-

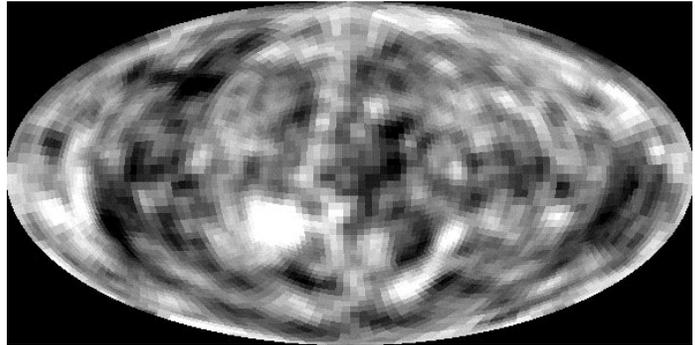


Figure 10: Map of fluctuations in the CMBR predicted in the cosmic string model. The angular resolution is several degrees and the peak amplitude is $\delta T/T = 50 \times 10^{-6}$.

tial wells at the time of last scattering (Sachs-Wolfe effect). The multipole $l \sim 200$ corresponds approximately to the angular scale of 1° , roughly the angle subtended by the Hubble radius on the last scattering surface. On scales comparable to and smaller than this ($l \gtrsim 200$), the anisotropy signal is dominated by acoustic oscillations of the photon-baryon fluid. For primordial adiabatic fluctuations, this effect leads to the characteristic ‘Doppler peaks’ at multipoles $l \gtrsim 200$. At very small angular scales ($l \gtrsim 1000$), these oscillations, and the resulting anisotropy amplitudes, are depressed by Silk damping (photon diffusion) and the finite thickness of the last scattering surface. Since the Hubble radius at the time of last scattering depends on Ω , the position of the first Doppler peak is sensitive to the mean density, $l_{peak} \simeq 200/\sqrt{\Omega}$ —it should be shifted to higher l in an open universe [67]. In inflation models, the amplitudes of the peaks depend on the inflation parameters n_s , n_T , and r , as well as on cosmological parameters such as h , Ω_B , and Ω_Λ . (They also depend on whether the universe was reionized at high redshift.) At higher l , the anisotropy spectrum is also sensitive to assumptions about the dark matter. The anisotropy power spectrum from an inflation model with $n_s = 1$ and $r = 0$ is shown in figure 11 for standard cold dark matter and mixed (cold plus hot) dark matter [68]. Note the flat region at small l and the Doppler peaks at $l \gtrsim 200$. In defect models, the Doppler peaks may be less pronounced and shifted in l .

The detection of CMBR fluctuations so far does not strongly discriminate between theoretical models. In the future, however, detailed observational study of the anisotropy power spectrum can provide a significant test for inflation. The first step will be a clearer indication of a Doppler peak in the spectrum and the determination of its amplitude and position. Ultimately, we would like to determine the angular power spectrum to the cosmic variance limit (the accuracy of an all-sky map) for multipoles up to $l = 1000$ or beyond. This would help

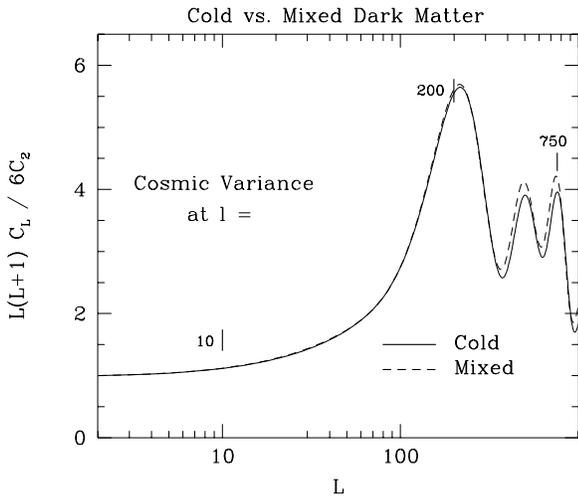


Figure 11: The spectrum of temperature fluctuations predicted in an inflationary model containing no tensor perturbations and a primordial scalar perturbation spectrum of the Harrison-Zel'dovich form, $n_S = 1$, normalized to the quadrupole C_2 , for both cold and mixed dark matter. Note the flat region at small l and the Doppler peaks at $l \geq 200$. The vertical marks show the expected cosmic variance for a full-sky map.

reduce the degeneracy ('cosmic confusion') between variations in model and cosmological parameters. Full-sky maps with sub-degree scale resolution are also required to test whether the fluctuations in the temperature are Gaussian distributed, as predicted by inflation; non-Gaussian CMBR features in topological defect models would be expected to show up on these scales.

8.2 Large-scale Structure

The clustering of matter in the universe provides another probe of early universe models for structure formation. Defining the density fluctuation field, $\delta(\mathbf{x}, t) = (\rho(\mathbf{x}, t) - \langle \rho \rangle) / \langle \rho \rangle$, it is convenient to consider the Fourier transform,

$$\delta(\mathbf{x}, t) = \int \frac{d^3k}{(2\pi)^3} \delta_{\mathbf{k}}(t) e^{i\mathbf{k} \cdot \mathbf{x}}, \quad (8.1)$$

in terms of which the density power spectrum $P_\rho(k)$ is defined as

$$\langle \delta_{\mathbf{k}} \delta_{\mathbf{k}'}^* \rangle = P_\rho(k) \delta_D(\mathbf{k} + \mathbf{k}') \quad (8.2)$$

As noted above, the primordial density spectrum from inflation is generally a power law in wavenumber, $P_\rho(k, t_i) = A k^{n_s}$, with $n_S \sim 1$. The spectrum subsequently subsequently evolves to produce the density spectrum at late times, $P_\rho(k, t_0) = T^2(k) P_\rho(k, t_i)$, where the transfer function $T(k)$ encodes the scale-dependence of the linear ($\delta \ll 1$) gravitational evolution of the perturbation

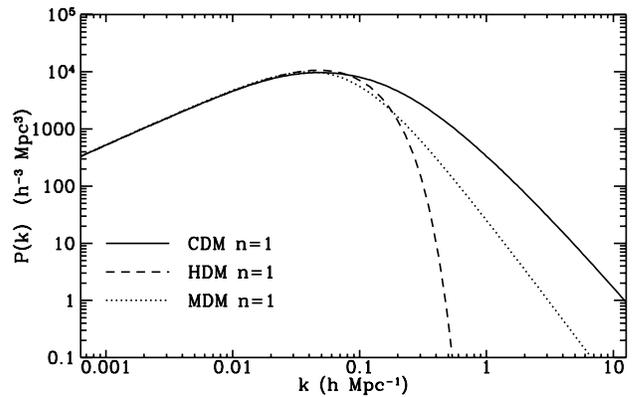


Figure 12: The processed linear power spectrum for a primordial spectrum of the Harrison-Zel'dovich form, $n_S = 1$, for cold, hot, and mixed dark matter.

modes. The transfer function depends on the nature of the dark matter (hot, warm, cold, or a mixture), its density (Ω_{DM}), the baryon density Ω_B , the Hubble parameter h , and the nature of the processes generating the fluctuations (adiabatic, isocurvature, or seeded defects). Power spectra for scale-invariant ($n_S = 1$) models with adiabatic fluctuations and three choices of dark matter are shown in Fig. 12. On small scales, there is an additional correction for non-linear evolution. Finally, the spectrum of luminous objects such as galaxies is related to the density spectrum by a bias prescription: for the simplest linear bias model, $P_{gal}(k, t_0) = b_{gal}^2 P_\rho(k, t_0) = b_{gal}^2 T^2(k) P_\rho(k, t_i)$ (not including the non-linear power correction). However, as noted above, there is little theoretical reason to expect the relation between mass and light to be that simple. Ultimately, the complexities of non-linear evolution and the inclusion of non-gravitational (hydrodynamical) processes in the formation of galaxies must be addressed by numerical simulations.

There are now a number of catalogs which have been used to determine the galaxy power spectrum $P_{gal}(k)$ down to wavenumbers $k \sim 0.01 h \text{ Mpc}^{-1}$. These suggest that the standard cold dark matter (CDM) spectrum (with $n_S = 1$, $\Omega_{cdm} h = 0.5$, and linear bias) does not have exactly the right shape to match the observed galaxy spectrum: it has relatively too little (too much) power on large (small) scales, as shown in Fig. 13 (from [69]). Currently popular variants which improve the spectral shape include a mixture of hot and cold dark matter, with $\Omega_{hot} \simeq 1 - \Omega_{cold} \simeq 0.2$, or low-density CDM ($\Omega_{cold} \simeq 0.3$, with or without a cosmological constant to retain spatial flatness). The spectral shape is also improved by increasing Ω_B above the range normally allowed by nucleosynthesis, reducing h to its lower limit of 0.4 or below, or tilting the primordial spectrum to $n_S < 1$. By

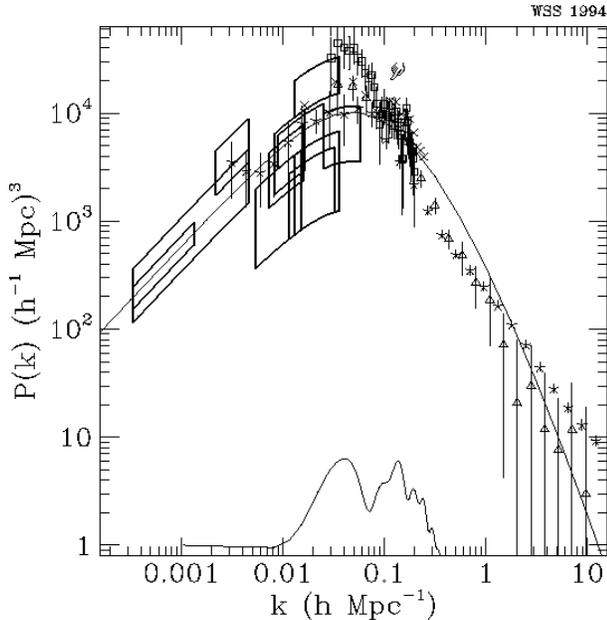


Figure 13: The power spectrum for standard CDM (solid line) is compared with the results from several galaxy surveys (points) and with CMBR anisotropy measurements (the error boxes, assuming $n_S = 1$). Due to the different methods used and corrections applied to the data, this figure should not be overinterpreted.

the turn of the millennium, larger galaxy redshift surveys will extend the measurements of $P_{gal}(k)$ down to smaller wavenumbers, overlapping the scales probed by COBE DMR. This should allow more direct constraints on the bias between galaxies and mass. In addition, we will have other methods that more directly probe the *mass* power spectrum over large scales: large-scale peculiar velocity measurements (e.g., from the Sunyaev-Zel'dovich effect in clusters) and weak gravitational lensing measurements of the shear field over degree scales.

§9 Conclusions

The robustness of the evidence for the standard cosmology has allowed cosmologists to use it as a jumping off point for theoretical exploration of the early universe. Now, cosmology is entering a new era: theoretical models for the evolution of the early universe that have been developed over the last two decades are becoming subject to a variety of observational tests. These observational probes fall roughly into four categories: determination of the cosmological parameters (e.g., H_0 , t_0 , and Λ); determination of the nature, quantity, and spatial distribution of the dark matter (including dark baryons); the microwave background anisotropy; and the large-scale structure of the (dark and luminous) matter. New measurements planned or underway in each of these areas will allow us to test

models for the early universe and discriminate between them.

In each of these areas, we can identify critical advances that are anticipated over the next decade. Improved measurements of the cosmological parameters are necessary to pin down the transfer function for perturbation growth and for CMBR anisotropies. In particular, additional Cepheid distances from HST should significantly advance the calibration of secondary distance indicators, while the alternative direct distance estimators (e.g., the Sunyaev-Zel'dovich effect) should continue to mature. The settling of the H_0 debate will have critical impact on cosmology—e.g., if a consensus emerges on the high value $h = 0.8$, the simple, elegant Einstein-de Sitter model with $\Omega_0 = 1$ will have to be abandoned. If the ‘expansion age’ $H_0 t_0$ turns out to be 1 or larger and gravitational lens constraints on Λ are not flawed, then the standard cosmology itself would be in crisis. Deep spectroscopic surveys and the discovery of significant numbers of high-redshift supernovae should ultimately provide important information on the deceleration parameter. In the area of dark matter, one looks forward to further developments in WIMP and axion detection, and to neutrino mass experiments. As of this writing, the LSND experiment at Los Alamos has reported preliminary evidence for neutrino oscillations consistent with a light neutrino in the mass range of several eV. In the microwave background, long-duration balloon flights and perhaps a post-COBE satellite will produce CMBR sky maps at sub-degree-scale resolution. These should go a significant way towards testing inflation and defect models as well as provide additional constraints on the cosmological parameters. In the area of large-scale structure, new galaxy surveys will uncover the statistics of galaxy clustering out to scales that overlap those probed by COBE. There will also be much larger samples of quasars and absorption line systems than heretofore, to probe clustering at early times. In addition, the mass power spectrum will be probed by peculiar velocity and weak lensing studies.

Cosmology thus stands hopefully but rather uncertainly on the threshold of the new millennium: we are soon to learn whether the speculative physical ideas developed in the last two decades are consistent with the universe around us. While the initial tests have been passed, the crucial tests of ‘cosmology beyond the standard model’ are yet to come.

§Acknowledgements

I would like to thank the participants and organizers of Snowmass '94 for an enjoyable, stimulating conference. This research was supported by the DOE and by NASA grant NAG5-2788 at Fermilab.

§Appendix: Friedmann-Robertson-Walker Models

First, a brief word on notation and units. The fundamental unit of distance in extragalactic astronomy is the parsec (pc): $1 \text{ pc} = 3.26 \text{ light-years} = 3.09 \times 10^{18} \text{ cm}$. The typical galactic scale is measured in kiloparsec (kpc): the solar system is approximately $R_0 = 8 - 8.5 \text{ kpc}$ from the center of the Milky Way. The nearest large galaxy to our own is Andromeda, at a distance of roughly 700 kpc , and the typical distance between neighboring large galaxies (outside clusters) is a few Megaparsec (Mpc). Absolute extragalactic distances are uncertain by a factor of order 2, and are usually given in terms of $h^{-1} \text{ Mpc}$, where the Hubble parameter $H_0 = 100h \text{ km/sec/Mpc}$, and observations indicate $0.4 < h < 1$. Cosmological distances are often estimated using redshifts via the Hubble law, $v = cz = H_0 d$, and so are sometimes expressed in terms of recession velocity: $10,000 \text{ km/sec} = 100 h^{-1} \text{ Mpc}$ and corresponds to a redshift $z = 0.033$. Astronomical masses and luminosities are measured in solar units, $M_\odot = 1.989 \times 10^{33} \text{ gm}$, $L_\odot = 3.86 \times 10^{33} \text{ erg/sec}$. Large galaxies commonly have a luminosity of order $10^{10} - 10^{11} L_\odot$ and mass of order $10^{11} - 10^{12} M_\odot$.

In the context of general relativity, homogeneity and isotropy severely restrict the form of the spatial geometry: a uniform stress-energy momentum tensor implies that the constant-time 3-surfaces have uniform spatial curvature. The spacetime metric thus takes the Friedmann-Robertson-Walker (FRW) form

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

where $a(t)$ is the global scale factor, r, θ, ϕ are the fixed coordinates carried by fundamental comoving observers, t is proper time measured on the clocks they carry, and $k = 0, 1, -1$ is the sign of the spatial 3-curvature. The case $k = 0$ corresponds to flat, Euclidean 3-space (R^3) written in spherical coordinates, $k = 1$ corresponds to the geometry of the three-sphere (S^3), and $k = -1$ to the 3-hyperboloid (H^3), the three-dimensional analogue of a hyperbolic saddle. Thus, models with $k \leq 0$ are spatially infinite (open), while those with $k = 1$ are spatially finite (closed).

The FRW models are characterized by the global scale factor $a(t)$, whose dynamics is determined by the matter content of the universe through Einstein's equations,

$$H^2 \equiv \left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho - \frac{k}{a^2} + \frac{\Lambda}{3} \quad (\text{A.2})$$

and

$$\frac{\ddot{a}}{a} = -\frac{4\pi G(\rho + 3p)}{3} + \frac{\Lambda}{3}. \quad (\text{A.3})$$

Here ρ is the mean energy density of matter, p is its pressure, and Λ is the cosmological constant, *i.e.*, the effective contribution to the energy-momentum from the vacuum state.

Observations suggest that the fluid energy density of the universe is currently dominated by non-relativistic matter (m), while the early universe was dominated by ultrarelativistic particles, or radiation (r). For non-relativistic matter, the energy density scales as $\rho_m \sim a^{-3}$, while for radiation, $\rho_r \sim a^{-4}$. The transition from radiation to matter domination ($\rho_m = \rho_r$) thus occurred at the epoch

$$1 + z_{eq} = \frac{a_0}{a_{eq}} = \frac{\rho_m^0}{\rho_r^0} = 2.3 \times 10^4 \Omega_0 h^2, \quad (\text{A.4})$$

where the present matter density $\rho_m^0 = \Omega_0 \rho_{crit} = 1.88 \Omega_0 h^2 \times 10^{-29} \text{ g cm}^{-3}$, $\rho_{crit} = 3H_0^2/8\pi G$ is the present density of the spatially flat ($k = 0$) model, and the present radiation energy density in the CMBR and an assumed three species of massless neutrinos is $\rho_r^0 = (\pi^2/30) g_* T_0^4 = 8.1 \times 10^{-34} \text{ g cm}^{-3}$.

Two features of the solutions to the Friedmann equation (A.2) with $\Lambda = 0$ are worth noting. In this case, there is a one-to-one correspondence between the spatial geometry and the fate of the universe: open models ($k \leq 0$) expand forever, while closed models ($k > 0$) eventually recollapse, because the energy density for matter and radiation fall off faster than a^{-2} . Conversely, in the early universe, $a \ll a_0$, the matter and radiation terms dominate over the spatial curvature, and the dynamics of the model is well approximated by setting $k = 0$. In this limit, for a matter-dominated universe, $a(t) \sim t^{2/3}$, $\rho_m = 1/6\pi G t^2$, while for radiation-domination, $a(t) \sim t^{1/2}$ and $\rho_r = 3/32\pi G t^2$.

The principal observable cosmological parameters of the FRW models are the Hubble parameter, $H_0 = (\dot{a}/a)_0$, the age of the Universe, t_0 , the present (non-relativistic) mass density relative to the 'critical' density of the spatially flat, Einstein-de Sitter ($k = \Lambda = 0$) model, $\Omega_0 = \rho_0/\rho_{crit} = 8\pi G \rho_0/3H_0^2$, the deceleration parameter, $q_0 = -(\ddot{a}/\dot{a}^2)_0$, which measures the rate at which the gravitational attraction of the matter is slowing down the expansion, and the contribution of the cosmological constant to the present expansion rate, $\Omega_\Lambda = \Lambda/3H_0^2$. From the Friedmann equations, these parameters are related by

$$1 = \Omega_0 + \Omega_\Lambda - \frac{k}{a_0^2 H_0^2} \quad (\text{A.5})$$

$$q_0 = \frac{\Omega_0}{2} - \Omega_\Lambda. \quad (\text{A.6})$$

For vanishing cosmological constant, $\Omega_0 - 1$ determines the sign of the spatial curvature: $\Omega_0 = 1$ for the spatially flat model ($k = 0$), and it is less than one for open models.

§References

- [1] G. F. Smoot, C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess, E. S. Cheng, G. De Amici, S. Gulkis, M. G. Hauser, G. Hinshaw, C. Lineweaver, K. Loewenstein, P. D. Jackson, M. Jansen, E. Kaita, T. Kelsall, P. Keegstra, P. Lubin, J. Mather, S. S. Meyer, S. H. Moseley, T. Murdock, L. Tokke, R. F. Silverberg, L. Tenorio, R. Weiss, and D. T. Wilkinson. *Astrophysical Journal Letters*, 396:L1, 1992.
- [2] C. L. Bennett, A. Kogut, G. Hinshaw, A. Banday, E. L. Wright, K. Gorski, D. T. Wilkinson, R. Weiss, G. F. Smoot, S. S. Meyer, J. C. Mather, P. Lubin, K. Loewenstein, C. Lineweaver, P. Keegstra, E. Kaita, P. D. Jackson, and E. S. Cheng. *Astrophysical Journal*, 436:423, 1994.
- [3] K. Ganga, E. Cheng, S. Meyer, and L. Page. *Astrophysical Journal Letters*, 410:L57, 1993.
- [4] S. J. Maddox, G. Efstathiou, W. J. Sutherland, and J. Loveday. *Monthly Notices of the Royal Astronomical Society*, 242:43p, 1990.
- [5] C. A. Collins, R. C. Nichol, and S. L. Lumsden. *Monthly Notices of the Royal Astronomical Society*, 254:295, 1992.
- [6] K. B. Fisher, M. Davis, M. A. Strauss, A. Yahil, and J. P. Huchra. *Astrophysical Journal*, 402:42, 1993.
- [7] G. Efstathiou, N. Kaiser, W. Saunders, A. Lawrence, M. Rowan-Robinson, R. S. Ellis, and C. S. Frenk. *Monthly Notices of the Royal Astronomical Society*, 247:10p, 1990.
- [8] H. Feldman, N. Kaiser, and J. Peacock. *Astrophysical Journal*, 426:23, 1994.
- [9] V. de Lapparent, M. Geller, and J. P. Huchra. *Astrophysical Journal Letters*, 302:L1, 1986.
- [10] M. Geller and J. P. Huchra. *Science*, 246:897, 1989.
- [11] T. Broadhurst, R. Ellis, D. Koo, and A. Szalay. *Nature*, 343:726, 1990.
- [12] This slice was made by D. Weinberg using N-body simulations of C. Park and J. R. Gott. See J. Gunn and D. Weinberg. in *Wide-Field Spectroscopy and the Distant Universe*, eds. S. Maddox and A. Aragon-Salamanca. World Scientific, Singapore, 1995.
- [13] J. R. Mould, L. Stavelly-Smith, R. Schommer, G. Bothun, P. Hall, M. Han, J. Huchra, J. Roth, W. Walsh, and A. Wright. *Astrophysical Journal*, 383:467, 1991.
- [14] J. C. Mather, E. S. Cheng, D. A. Cottingham, R. E. Eplee, D. J. Fixsen, T. Hewagama, R. B. Isaacman, K. A. Jensen, S. S. Meyer, P. D. Noerdlinger, S. M. Read, L. P. Rosen, R. A. Shafer, E. L. Wright, C. L. Bennett, N. W. Boggess, M. G. Hauser, T. Kelsall, S. H. Moseley, R. F. Silverberg, G. F. Smoot, R. Weiss, and D. T. Wilkinson. *Astrophysical Journal*, 420:439, 1994.
- [15] T. Walker, G. Steigman, D. Schramm, K. Olive, and H. Kang. *Astrophysical Journal*, 376:51, 1991.
- [16] S. Carroll, W. H. Press, and E. Turner. *Annual Review of Astronomy and Astrophysics*, 30:499, 1992.
- [17] M. Fukugita and E. Turner. *Monthly Notices of the Royal Astronomical Society*, 253:99, 1991.
- [18] C. Kochanek. *Astrophysical Journal*, 419:12, 1994.
- [19] D. Maoz and H. Rix. *Astrophysical Journal*, 416:425, 1994.
- [20] G. Jacoby et al. *Publications of the Astronomical Society of the Pacific*, 104:559, 1992.
- [21] G. Jacoby et al. *Astrophysical Journal*, 356:322, 1990.
- [22] J. Tonry. *Astrophysical Journal Letters*, 373:L1, 1991.
- [23] A. G. Riess, W. H. Press, and R. P. Kirshner. *Astrophysical Journal Letters*, 438:L17, 1995.
- [24] M. Hamuy, M. Phillips, J. Maza, N. Suntzeff, R. Schommer, and R. Aviles. *Astronomical Journal*, 109:1, 1995.
- [25] B. P. Schmidt, R. P. Kirshner, and R. G. Eastman. *Astrophysical Journal*, 395:366, 1992.
- [26] M. Birkinshaw and J. Hughes. *Astrophysical Journal*, 420:33, 1991.
- [27] T. Wilbanks, P. Ade, M. Fischer, W. Holzappel, and A. Lange. *Astrophysical Journal Letters*, 427:L75, 1994.
- [28] R. Blandford and R. Narayan. *Annual Review of Astronomy and Astrophysics*, 30:311, 1992.
- [29] M. J. Pierce, D. Welch, R. McClure, S. van den Bergh, R. Racine, and P. Stetson. *Nature*, 371:385, 1994.
- [30] W. Freedman, B. Madore, J. Mould, R. Hill, L. Ferrarese, R. Kennicutt, A. Saha, P. Stetson, J. Graham, H. Ford, J. Hoessel, J. Huchra, S. Hughes, and G. Illingworth. *Nature*, 371:757, 1994.

- [31] J. Cowan, F. Thielemann, and J. Truran. *Annual Review of Astronomy and Astrophysics*, 29:447, 1991.
- [32] D. Winget et al. *Astrophysical Journal Letters*, 315:L77, 1987.
- [33] I. Iben and G. Laughlin. *Astrophysical Journal*, 341:430, 1989.
- [34] A. Renzini in. *Proceedings of the 16th Texas Symposium on Relativistic Astrophysics and 3rd Symposium on Particles, Strings, and Cosmology*, eds. C. Akerlof and M. Srednicki. New York Academy of Sciences, New York, 1992.
- [35] X. Shi. *Astrophysical Journal*, in press, 1995.
- [36] R. Lynds and V. Petrosian. *Bulletin of the American Astronomical Society*, 18:1014, 1986.
- [37] G. Soucail et al. *Astronomy and Astrophysics*, 172:L14, 1987.
- [38] J. Tyson, F. Valdes, and R. Wenk. *Astrophysical Journal Letters*, 349:L19, 1990.
- [39] S. D. M. White, J. Navarro, A. Evrard, and C. Frenk. *Nature*, 366:429, 1993.
- [40] A. Dekel. *Annual Review of Astronomy and Astrophysics*, 32, 1994.
- [41] S. Cole, K. Fisher, and D. Weinberg. *Monthly Notices of the Royal Astronomical Society*, in press, 1995.
- [42] B. Paczynski. *Astrophysical Journal*, 304:1, 1986.
- [43] C. Alcock, R. Allsman, T. Axelrod, D. Bennett, K. Cook, H. Park, S. Marshall, M. Pratt, C. Stubbs, K. Griest, S. Perlmutter, W. Sutherland, K. Freeman, B. Peterson, P. Quinn, A. Rodgers, S. Chan, and C. Akerlof. *Nature*, 365:621, 1993.
- [44] E. Aubourg et al. *Nature*, 365:623, 1993.
- [45] A. Udalski, J. Szymanski, J. Kaluzny, M. Kuźbiak, W. Krzeminski, M. Mateo, G. Preston, and B. Paczynski. *Acta Astronomica*, 43:289, 1993.
- [46] A. Udalski et al. *Astrophysical Journal Letters*, 426:L69, 1994.
- [47] C. Alcock et al. *preprint*, 1994.
- [48] S. Tremaine and J. Gunn. *Physical Review Letters*, 42:407, 1979.
- [49] K. Van Bibber in. *Proceedings of the 7th Meeting of the American Physical Society Division of Particles and Fields*. eds. C. Albright, P. Kasper, R. Raja, and J. Yoh World Scientific, Singapore, 1992.
- [50] M. Milgrom. *Astrophysical Journal*, 333:689, 1988.
- [51] A. Guth. *Physical Review D*, 23:347, 1981.
- [52] D. Lyth and E. Stewart. *Physics Letters B*, 252:336, 1990.
- [53] B. Ratra and P. J. E. Peebles. *Astrophysical Journal*, 432:5, 1994.
- [54] M. Bucher, A. Goldhaber, and N. Turok. *preprint*, 1994.
- [55] A. Linde. *Physics Letters B*, 108:389, 1982.
- [56] A. Albrecht and P. Steinhardt. *Physical Review Letters*, 48:1220, 1982.
- [57] S. Hawking. *Physics Letters B*, 115:295, 1982.
- [58] A. Starobinski. *Physics Letters B*, 117:175, 1982.
- [59] A. Guth and S. Y. Pi. *Physical Review Letters*, 49:1110, 1982.
- [60] J. Bardeen, P. Steinhardt, and M. Turner. *Physical Review D*, 28:679, 1983.
- [61] V. Rubakov, M. Sazhin, and A. Veryaskin. *Physics Letters B*, 115:189, 1982.
- [62] E. Copeland, E. Kolb, A. Liddle, and J. Lidsey. *Physical Review D*, 48:2529, 1993.
- [63] K. Gorski, G. Hinshaw, A. Banday, C. Bennett, E. Wright, A. Kogut, G. Smoot, and P. Lubin. *Astrophysical Journal Letters*, 430:L89, 1994.
- [64] A. Linde. *Physics Letters B*, 129:177, 1983.
- [65] T. Kibble. *Physics Reports*, 67:183, 1980.
- [66] B. Allen, R. Caldwell, P. Shellard, A. Stebbins, and S. Veeraraghavan. *in preparation*, 1995.
- [67] M. Kamionkowski, D. Spergel, and N. Sugiyama. *Astrophysical Journal Letters*, 1994.
- [68] S. Dodelson, E. Gates, and A. Stebbins. *in preparation*, 1995.
- [69] M. White, D. Scott, and J. Silk. *Annual Reviews of Astronomy and Astrophysics*, 32:319, 1994.