CHAPTER 7

COSMOLOGICAL CHALLENGES FOR THE $21^{\rm st}$ CENTURY

PAUL J. STEINHARDT Department of Physics and Astronomy University of Pennsylvania, Philadelphia, PA

[7.1 INTRODUCTION](#page--1-0)

The 21st century shows every promise of being an age of historic discovery in cosmology due to an extraordinary influx of new technologies and new ideas. As the century approaches, the field is replete with controversial issues great and small, as demonstrated at the Princeton 250th Anniversary cosmology counterpart to this meeting, entitled "Critical Dialogues in Cosmology." In this paper, the focus will be on a few pivotal issues likely to dominate the 21^{st} century, shaping the future of astrophysics and cosmology and influencing our understanding of fundamental physics.

To gain a perspective on the challenges of the $21st$ century, it is instructive to recall the successes of the past century, many of which have been pioneered by theorists and experimentalists at Princeton. Cosmology in the 20th century has undergone a remarkable metamorphosis from a field of pure speculation to a field of hard science due to a series of technological advances that have made it possible to probe the distant universe and to test our ideas through observations and experiments. We have seen the first definitive observational evidence that the sun is not the center of the universe; the first evidence that most nebulae are galaxies of stars far away rather than clouds of gas close by; and the first evidence that the universe has been expanding and cooling, explaining the motion of galaxies and the origin of the elements. These discoveries have forced us away from the strongly preferred notion that we live in a static, time-invariant universe. We have also discovered that the early universe had a slightly inhomogeneous distribution of matter and energy that may account for the large-scale structure seen today.

All of these observational breakthroughs have shaped our modern view of

the universe, leading us towards a standard model known as the *hot big bang picture.* According to this model, the universe began as a infinitesimal patch of space filled with gas of nearly infinite temperature and nearly infinite density. Suddenly, for reasons not understood, the universe began to expand and cool. The universe observed today is simply the result of fifteen billion years of expanding and cooling and the consequent evolution of matter subject to known physical laws.

Although the hot big bang model is consistent with all observations to date, it leaves important aspects of the universe unexplained. For example, the hot big bang model does not enable us to understand how much matter and energy there is in the universe, or what forms it takes, or from where it originated. It does not explain the geometry of the universe. It does not explain why the universe is homogeneous and isotropic on large scales and, yet, highly inhomogeneous on smaller scales. Finally, the hot big bang model fails to explain the ultimate question: why is the universe the way it is?

In order to address these questions, the universe must be observed to much farther distances and much further back in time. What is so exciting about the coming century is that a whole host of new technologies are providing these very capabilities. Mention of just a few of these new windows on the universe can give some impression of the remarkable potential of the new century:

Over the next ten years, there will be red shift surveys of millions of galaxies that will produce a full, three-dimensional map of the arrangement of galaxies in the universe. Gravitational lensing, the bending of light of distance sources by foreground galaxies and clusters of galaxies, will be used to measure the amount and distribution of matter in the universe, both ordinary baryonic matter and dark matter of unknown kinds. It will become possible to measure with precision the x-ray luminosity of gas in large clusters of galaxies. This makes it possible to measure the amount of baryonic matter which lies in the gas between galaxies. When x-ray and gravitational lensing measurements are combined, it becomes possible to dissect clustered matter into its baryonic and its dark matter components. Observations of the absorption of quasar radiation by Lyman- α clouds, foreground clouds of primordial gas in intergalactic regions where there has been little or no star formation, are another promising source of new information. The absorption of the background quasar radiation may shed light on the first stages of structure formation and on the abundance of elements in the primordial universe. Studies of the Sunyaev-Zel'dovich effect, the rescattering of cosmic microwave background radiation from hot gas and clusters, and the systematic searches for supernova at cosmic distances are providing new standard candles for accurately measuring distance in the universe and determining the expansion, deceleration, and equation-of-state of the universe. And, perhaps the most powerful observational tool of all is the cosmic microwave background anisotropy. The all-sky map of the anisotropy anticipated in the next decade will be a snapshot of the distribution of matter and energy in the universe when it was

only 100,000 years old, a decisive test of cosmological models and a new method to constrain key cosmic parameters.

With the wealth of new data pouring in, one should anticipate that the 21st century will bring some startling, unanticipated surprises in cosmology. Hence, the following speculations on the key challenges of the 21^{st} century are offered with some trepidation. The reader is warned: these views may be made obsolete in only a few short years due to a discovery which will dramatically change the the theoretical view and preoccupy cosmologists for at least a century to come.

[7.2 Is THE UNIVERSE FLAT?](#page--1-0)

The common view today is that determining the spatial curvature of the universe is important because it is one of the key tests of inflationary cosmology. Indeed, the inflationary model of the universe does explain how the universe became flat due to extraordinary expansion during the first instants after the big bang and does predict that the universe should remain flat today [1]. However, the issue of the flatness of the universe goes beyond inflation; even if inflation does not survive as an explanation of the observable universe, flatness is a critical issue for the future science of cosmology because it cuts to the heart of a basic assumption of nearly all models, the cosmological principle.

The cosmological principle is the notion that the universe observed from our limited vantage point is representative of the universe entire. That is to say, there should exist a length scale such that a coarse-grain average produces a nearly homogeneous picture. It is important that this length scale be well within an observationally accessible range in order for us to be able to determine the average properties of the universe.

When the cosmological principle was introduced, the dominant idea was that the universe is static and time-invariant. This would mean that there has been a semi-infinite amount of time for distant regions of the universe to interact and reach an equilibrium. Also, there is no limit to the range of vision since there is sufficient time for light to have propagated an unbounded distance. In this scenario, the notion of a coarse-grain length scale and an approach to homogeneity seems plausible.

However, our concept of the universe has changed due to the success of the hot big bang picture. Apparently, our patch of universe has only existed for fifteen billion years; consequently, the maximum distance that can be observed is about fifteen billion light-years, referred to as the "horizon distance." Most likely, the space within our horizon is only a tiny, infinitesimal corner of a much larger universe. This is an essential but seldom emphasized aspect of the big bang picture. Furthermore, regions separated by fifteen billion light-years or more have not had a chance to interact since the beginning of time. Given these conditions, the plausibility of the cosmological principle seems less certain. Is it really possible to to understand the universe entire when we are constrained by causality from observing most of it? The only hope is that there exists a coarse-grain length scale which is smaller than the horizon distance such that the properties of the entire universe can be determined by coarse-grain averaging observations within our limiting horizon. Whether this is so is an issue that must be tested, rather than accepted as assumption.

How might we test the cosmological principle? One test is to measure the distribution of galaxies and determine if there exists a length scale over which we can coarse-grain average and approach homogeneity. Before this year, cosmologists would have to confess that there is no evidence of this coarse-grain length scale. Red shift surveys, such as the Center for Astrophysics Survey [1], show an inhomogeneous distribution of large filamentary and wall-like concentrations of galaxies separated by large voids. In a survey that stretches out several hundred megaparsecs (several hundreds of millions of light-years), the structures themselves stretch several hundreds of megaparsecs, and there is not any indication of approaching homogeneity. In recent months, though, the situation has begun to change as red shift surveys probing deeper distances have begun to report. For example, the Las Campanas Survey of Schechtman et al. [2] has produced maps which reach a distance almost an order of magnitude farther than the Center for Astrophysics study. What that Las Campanas Survey shows is that hundredmetaparsec structures similar to those found in the Center for Astrophysics Survey are found throughout the Las Campanas map, too, but there do not seem to be yet larger structures. If further deep surveys support this conclusion, it will be historic. At the beginning of this century, the predominant view was that the universe consist of a uniform distribution of stars; instead, a hierarchy of increasingly large structures has been found: galaxies, galaxy clusters, superclusters, filaments, walls, and voids. The Las Campanas Survey suggests that the end of this hierarchy has finally been reached and that there truly is a coarse-grain length scale for the galaxy distribution of some few hundred megaparsecs.

Another kind of evidence can be obtained by measuring various electromagnetic radiation backgrounds, such as the x-ray background or the cosmic microwave background. The cosmic microwave background can be interpreted as a snapshot of the distribution of matter and radiation at a time when the universe was only a few hundred thousand years old. As the radiation decoupled from matter and began to stream towards us, the radiation was red-shifted or blue-shifted depending on the gravitational potential in the region from which it last scattered (the "last scattering surface"). Inhomogeneities in the distribution of the radiation can, therefore, be used to determine the homogeneity of matter and energy at this early epoch. The Cosmic Background Explorer (COBE) satellite [4] has measured the root-mean-square inhomogeneity to be a few parts in $10⁵$, where the microwave antenna produce a coarse-grain average over 10°. Subsequent ground and balloon based experiments suggest similar inhomogeneities at smaller angular scales ranging down to 0.5 degrees. Hence, the cosmic microwave background is strong evidence that the distribution of matter and energy in the large-scale universe was highly homogeneous at early epochs.

For more subtle reasons, the cosmological principle is also tied to the flatness of the universe. Suppose that future empirical evidence were to point to finite spatial curvature based on observations made within our horizon. One possible interpretation is that the entire universe is homogeneously curved (globally). Most discussions of cosmology suggest this interpretation. However, this explanation requires that we live at a very special epoch: although the horizon distance and the curvature vary at different rates as a function of time, it would have to be that we live at the particular epoch when their magnitudes are comparable. (According to the big bang picture, the horizon distance would have to have been negligibly small compared to the radius of curvature at earlier times, and the universe would appear flat; at later times, the curvature would become much smaller than the horizon distance, dominating the expansion of the universe and causing large optical distortions.) It is a coincidence sometimes referred to as the "flatness problem," so-called because it would be an important feature of the universe which had no natural explanation. An embarrassing coincidence is not the only possibility, though.

If future experiments point to non-zero spatial curvature, a second interpretation is that we have stumbled upon a spectrum of large scale spatial distortions spanning scales from subhorizon to superhorizon. By measuring the curvature within our horizon, we have sampled the distribution over some random, horizonsized patch. A similar patch in a different region of space would have a different spatial curvature.

One explanation, the homogeneously curved universe, is consistent with the cosmological principle. But, the second explanation means that conditions within our horizon are not the same as conditions elsewhere in the universe. The problem is that no observation can distinguish the two possibilities. Consequently, if nonzero curvature becomes established observationally, the cosmological principle and any attempt to explain the universe beyond our horizon must be regarded suspiciously.

Given the significance of the issue, it would be good to report that the curvature can be reliably measured in the near future. Many recent papers have suggested that it is likely in the near future. However, an important warning is due: a careful analysis shows that methods cited as testing the flatness of the universe do not measure the curvature directly, but only in combination with uncertain, model-dependent assumptions. In particular, assumptions must be made about the matter-energy content of the universe, or the spectrum of primordial fluctuations, or other properties of the universe. Consequently, a discussion of measuring the flatness of the universe will be omitted in this section. It will appear in the subsequent section where it is tied to the testing of cosmological models, as is logically appropriate.

[7.3 Do WE LIVE IN AN INFLATIONARY UNIVERSE?](#page--1-0)

Determining the validity of the inflationary model of the universe [1, 5] is one of the key challenges for the 21^{st} century. Inflation is our best hope for answering many of the questions left open by the hot big bang picture [6, 7]. In a single stroke, inflation explains the amount of matter and energy in the universe; how the matter and energy in the universe originated; how the universe became spatially flat; why the universe is homogeneous on large scales; and how the inhomogeneities arose on smaller scales which eventually gave rise to the formation of galaxies and larger-scale structures.

It is noteworthy that the inflationary model creates dynamically the remarkable conditions suggested by the cosmological principle: namely, a universe in which there exists a coarse-grain length scale within the observable horizon over which the universe appears homogeneous and isotropic. Averaged properties within our universe are representative of the greater universe. The cosmological principle is derived rather than assumed.

How will inflation be tested? Two of its generic predictions are a spatially flat universe and a scale invariant spectrum of gaussian, adiabatic energy density fluctuations [8]. The first, spatial flatness, is equivalent to the prediction that the total energy density of the universe is equal to the critical density; that is, Ω_{total} is equal to one. Throughout the remainder of the paper, Ω_i is used to represent the ratio of the energy density of type *i* to the critical density needed to close the universe.

Note that I have emphasized here the term "total energy" referring to the sum of all forms of energy including matter energy, radiation, and any other contributions. In particular, inflation does not predict that $\Omega_{\text{matter}} = 1$ necessarily, but only that the total energy density is equal to the critical density. $\Omega_{\text{matter}} = 1$, as assumed in the standard cold dark matter model of structure formation, is only a special, simple case. Hence, recent evidence suggesting that $\Omega_{\text{matter}} < 1$ is inconsistent with the standard cold dark matter model, but it is not inconsistent with inflation.

The second prediction of inflation is a scale invariant spectrum of fluctuations that should have left a mark on the cosmic microwave background anisotropy and may have been responsible for the formation of large-scale structure in the universe [8]. The spectrum is generated by quantum fluctuations which distort the distribution of energy when the universe is microscopically small during the first stages of inflation. Inflation freezes the fluctuation amplitude and stretches the wavelength to cosmic scales. The details of this process can be computed from first principles. The resulting spectrum is scale invariant in the sense that, if one expresses the density as a sum of Fourier modes, the amplitudes are nearly the same from wavelength to wavelength when averaged over the entire universe. Of course, our observations are restricted to a finite horizon, and so measurements in this restricted range of space will produce some deviations from the cosmic

average. The spectrum is Gaussian in the sense that the deviations from the cosmic mean are Gaussian-distributed. The spectrum is adiabatic in the sense that all forms of matter and radiation undergo the same fluctuations.

Although there are many types of inflationary scenario in the literature, the predictions described above rely on their common features and, hence, are generic. It is a strong feature of inflationary cosmology that its predictions do not depend sensitively on the details. (As in theoretical models of nearly any kind, it is possible to push on parameters and add contrivances to violate one or more generic predictions in any type of inflationary scenario. However, the conditions are so extreme that the models have little or no predictive power and are physically implausible.)

Even if inflation should pass these initial tests, some would not find them to be a convincing proof of inflation, arguing that the predictions that the universe must be spatially flat [9] and that the primordial perturbations spectrum must be scale invariant (Harrison-Zel'dovich-Peebles) [10] predate the invention of the inflationary model by several years. The argument is questionable. For, while it was discussed that flatness and scale invariant spectra are attractive for a viable cosmological model, there was no dynamic explanation of how they came to be prior to inflation. Symmetry arguments are not compelling. Flatness, while symmetrical, is highly unstable since curvature grows rapidly in big bang cosmology. Scale invariance, while symmetrical, is hard to explain without inflation since it requires primordial fluctuations and scales that exceed the causal horizon. Nevertheless, given that some find the aforementioned inflationary tests unpersuasive, it is important to emphasize more refined predictions of inflation that were not anticipated.

The first refinement is that the predicted perturbation spectrum is not perfectly scale invariant; instead, there is a small but measurable "tilt" [14, 15]. The spectrum can be characterized by a spectral index, n , where the energy perturbation amplitude is defined as $\delta \rho / \rho \propto \lambda^{(1-n)/2}$. Then, $n = 1$ corresponds to perfect scale-invariance. Inflation predicts that the spectral index will deviate from $n = 1$ by a few percent to several tens of a percent, depending on the details of the model. Inflation would predict a perfectly scale invariant $(n = 1)$ spectrum if the expansion rate were uniform. However, the expansion rate cannot remain uniform since eventually inflation must slow down to return the universe to the standard big bang expansion rate. In many models, the expansion rate is changing slowly throughout inflation. The tilt is limited by the fact that, if the expansion rate changes too rapidly, inflation ends prematurely, before there is sufficient inflation to solve the cosmological horizon and flatness problems. The allowed range is $n \sim 1.0 \pm 0.3$ [16, 17].

A second refinement is that the perturbations predicted by inflation are a combination of fluctuations in the energy density, which can be sources for the formation of structure, plus fluctuations in the space-time metric which will evolve into gravitational waves [11, 12, 13]. Furthermore, inflation predicts a relationship between the ratio of gravitational waves to energy density fluctuations and the tilt. These more refined predictions may require more than ten years to test, but it is rather likely that they will be tested well before end of the next century. The discovery of these effects should certainly be convincing to the last skeptics, since these are new predictions stemming specifically from analysis of inflationary models.

The critical test of these predictions will come from measurements of the cosmic microwave background anisotropy, which provides a detailed fingerprint of conditions in the early universe. The cosmic fingerprint [16] is obtained from a temperature difference map (Fig. 7.1) which displays the fractional fluctuation in the cosmic background temperature, $\Delta T(\mathbf{x})/T$, as a function of sky direction x. The map represents the deviations in temperature from the mean value, *T =* 2.726±.010, after foreground sources of radiation are subtracted. Testing inflation and other cosmological models entails comparing statistical properties of this map to the theoretical predictions. The simplest and most decisive statistical test is the two-point or temperature autocorrelation function. See Fig. 7.1. The temperature autocorrelation function, $C(\theta)$, compares the temperature at points in the sky separated by angle *6:*

$$
C(\theta) = \left\langle \frac{\Delta T}{T}(\mathbf{x}) \frac{\Delta T}{T}(\mathbf{x}') \right\rangle
$$

=
$$
\frac{1}{4\pi} \sum_{\ell} (2\ell + 1) C_{\ell} P_{\ell}(\cos \theta),
$$
 (7.1)

where $\langle \rangle$ represents an average over the sky and $\mathbf{x} \cdot \mathbf{x}' = \cos \theta$. The coefficients, C_f , are the *multipole moments* (for example, C_2 is the quadrupole, C_3 is the octopole, *etc.*). Roughly speaking, the value of C_{ℓ} is determined by fluctuations on angular scales $\theta \sim \pi/\ell$. A plot of $\ell(\ell + 1)C_{\ell}$ is referred to as the *cosmic microwave background (CMB) power spectrum.*

There is valuable information in the cosmic microwave background anisotropy in addition to the CMB power spectrum that will be extracted some day. Higherpoint temperature correlation functions (entailing three or more factors of $\Delta T/T$) could be obtained from the temperature difference map and be used to test if the fluctuation spectrum is Gaussian, as predicted by inflation. However, the fact that statistical and systematic errors in $\Delta T/T$ compound for higher-point correlations makes precise measurements very challenging. Polarization of the microwave background by the last scattering of photons from the anisotropic electron distribution is another sky signal that provides quantitative data that can be used to test models. However, for known models, the predicted polarization requires more than two orders of magnitude better accuracy than anisotropy measurements alone in order to discriminate models [18]. Although forthcoming satellite experiments will attempt to detect polarization or measure higher-order correlation functions, the most reliable information in the near future will be the CMB power spectrum. Fortunately, the CMB power spectrum is packed with information that can be used, by itself, to discriminate inflation from alternative models.

Figure 7.1: The temperature autocorrelation function, $C(\theta)$, is obtained from a map of the sky (here represented by the oval) displaying the difference in the microwave background temperature from the average value, $\Delta T/T$. $C(\theta)$ is computed by taking the map-average of the product of $\Delta T/T$ measured from any two points in the sky separated by angle θ . If $C(\theta)$ is expanded in Legendre polynomials, $P_{\ell}(\cos \theta)$, the coefficients C_{ℓ} are the *multipole moments*.

[Figure 7.2 d](#page-10-0)isplays a representative CMB power spectrum for an inflationary model [19]. For this example, the spectral index is $n = 0.85$, with equal contributions of energy density and gravitational wave fluctuations to the largeangular scale anisotropy. To the left of $\ell \approx 100$ are multipoles dominated by fluctuations over distances much larger than the size of the Hubble horizon at the time of last scattering, corresponding to angles $> 1^{\circ} - 2^{\circ}$ on the sky. According to the inflationary model, these wavelengths did not have a chance to evolve before last scattering and the beginning of the photon trek towards our detectors. Hence, these fluctuations preserve the imprint of whatever fluctuations were set by inflation. If the fluctuations are remnants of a nearly scale-invariant inflationary spectrum, then the low- ℓ part of the CMB power spectrum should be featureless, just as shown in the figure. If the spectrum is tilted, as in this case, there should be a small downward (or upward) slope over the low multipoles, as shown.

The spectrum includes, in general, both energy density and gravitational wave contributions, as indicated in the example of Fig. 7.2. For inflationary models, the two contributions are predicted to be statistically independent and simply sum to

give the total power. The fluctuations in $\Delta T/T$ are also predicted to be Gaussiandistributed for inflationary models. Hence, the C_{ℓ} 's, which are an average over $2\ell + 1$ Gaussian-distributed variables, have a χ^2 -distribution.

To the right of $\ell \approx 100$ are multipoles dominated by fluctuations with wavelengths smaller than the horizon at last scattering [16, 17]. The right side of the power spectrum figure appears different from the left because inhomogeneities spanning scales smaller than the Hubble horizon have time to evolve causally. Gravitational waves inside the horizon red-shift away. For energy density fluctuations, the baryon and photon begin to collapse and oscillate acoustically about the centers of high and low energy density, adding to the net microwave background perturbation. Each wavelength laid down by inflation initiates its acoustic oscillation shortly after entering the Hubble horizon. Hence, there is a well-defined phase-relation between the acoustic oscillations on different wavelengths. Waves just entering the horizon and smaller-wavelength waves which have completed a half-integral number of oscillations by last scattering will be at maximum amplitude. Wavelengths in between are mid-phase and will have smaller amplitudes. In a plot of C_{ℓ} 's, increasing ℓ corresponds to multipoles dominated by decreasing wavelengths. The variations of the oscillation phase with wavelength results in a sequence of peaks as a function of ℓ . These peaks are sometimes referred to as Doppler peaks or acoustic peaks.

The position of the first Doppler (or acoustic) peak is of particular interest. Its position along the ℓ -axis, left or right, is most sensitive to the value of Ω_{total} : the peak moves to the right in proportion to $1/\sqrt{\Omega_{\text{total}}}$, for large Ω_{total} [20]. There is only weak dependence on the Hubble constant and other cosmological parameters [21, 22]. Decreasing Ω_{total} to 0.1, say, causes the first Doppler peak to shift to $\ell \approx 600$ instead of $\ell \approx 200$, a dramatic and decisive difference. As a test of Ω_{total} , the first Doppler peak has the advantage that it is relatively insensitive to the form of the energy density, whether it be radiation, matter, or cosmological constant, and it is relatively difficult to mimic using other physical effects.

In sum, Fig. 7.2 illustrates how all three key features of inflation can be tested by the microwave background power spectrum. Large-angular scale fluctuations are consequences of approximate scale-invariance and the combination of energy density and gravitational wave perturbations. The presence of a combination of energy density and gravitational wave fluctuations can be detected from more subtle features, such as the ratios of the Doppler peaks to the plateau at small ℓ . A gently sloped CMB power spectrum at small ℓ is the signature of being slightly tilted from scale-invariance. Small-angular scale fluctuations, especially the position of the first Doppler peak, are consequences of Ω_{total} being unity. A sequence of subsequent Doppler peaks is a check that the perturbations are adiabatic [23].

The inflationary prediction for the CMB power spectrum is not unique, since there are undetermined, free parameters having to do both with inflation and with basic cosmic parameters. Figure 7.3 is a representative band of predicted curves

for different values of the spectral index. Each example lies within the parameter space achievable in inflationary models. Although the band is wide, there are common features among the curves which can be used as the critical tests of inflation. All have a plateau at small ℓ , a large first Doppler peak at $\ell \approx 200$, and then smaller Doppler hills at larger *L*

Figure 7.2: The CMB Power Spectrum: A plot of $\ell(\ell+1)C_{\ell}$ vs. multipole moment number ℓ is the cosmic microwave background power spectrum. For a given ℓ , C_{ℓ} is dominated by fluctuations on angular scale $\theta \sim \pi/\ell$. In inflation, the power spectrum is the sum of two independent, scalar and tensor contributions.

At this point, the best available CMB anisotropy data is in rough agreement with inflationary predictions [16], but is rather imprecise. In the next ten years, however, there will be dramatic improvements due to a combination of land, air, and space-based experiments, including the MAP (Microwave Anisotropy Probe) satellite to be launched by NASA and the COBRAS/SAMBA satellite to be launched by ESA. These experiments will have an uncertainty comparable to the line thickness at small angular scales, detecting every bump and wiggle.

If inflation or another recognized model (such as cosmic textures) is verified

by these measurements, then it will also be possible to determine the flatness, as well, to very high precision. However, it is also important to appreciate that if the measured spectrum does not conform to one of the known patterns, then the value of Ω_{total} will be unconstrained, remaining a key, unsettled issue in the field.

Figure 7.3: A band of microwave background power spectra allowed by inflation. The uppermost curve is a pure-scalar, scale-invariant spectrum, and the lower curves have tilt $(n < 1)$ and gravitational waves. Inflationary models with spectra somewhat higher than the uppermost curve are also possible. The common features among these curves—the prime targets for experimental tests of inflation—are a plateau at large angular scales, a prominent first Doppier (or acoustic) peak, and subsequent acoustic oscillation peaks at small angular scales.

7.4 DOES $\Omega_{\text{matter}} = 1$?

In order to speculate further, we shall suppose that during the next ten years inflation passes the tests described in the previous section during the next ten years and that the universe is proven to be flat. The next critical issue for cosmology will be whether or not $\Omega_{\text{matter}} = 1$. There are two reasons why this is a critical issue. If $\Omega_{\text{matter}} \geq 0.2$, then we know that some form of non-baryonic dark matter exists. On the other hand, if $\Omega_{\text{matter}} < 1$ and the universe has been proven to be flat *(e.g.,* according to the position of the first acoustic peak in the CMB spectrum), then there must exist some other form of energy in the universe besides matter which is a significant, and perhaps, even a dominant contributor to the total energy density of the universe.

The claim that $\Omega_{\text{matter}} \geq 0.2$ implies non-baryonic matter stems from the constraint on the baryon density derived from primordial nucleosynthesis. A few seconds after the big bang (or inflation), the universe was sufficiently hot to enable fusion of protons and neutrons into light nuclei, and yet cool enough that the nuclei were not destroyed by subsequent collisions. From the knowledge of the expansion rate and temperature history, along with details of nuclear interactions, the relative abundances of the light elements produced in this primordial epoch can be reliably predicted. The predictions depend on one unknown, $\Omega_b h^2$, where Ω_b is the ratio of the baryon density to the critical density and $h =$ $H_0/(100 \text{ km/s} - \text{Mpc})$ is a standard, dimensionless re-expression of the Hubble constant, *Ho.* Comparison of observations of primordial helium, deuterium, and lithium with the theoretical predictions have been used to constrain $\Omega_b h^2$ to a range of small values, between 0.01 and 0.02 [24]. Since *h* lies somewhere between 0.5 and 1, according to current observations, Ω_b is less than 10%. Consequently, if Ω_{matter} is found to be greater than 0.2, then non-baryonic dark matter must exist.

An important breakthrough is the attempt to constrain the primordial abundance of the light nuclear elements by measuring the deuterium abundance in Lyman- α clouds by measuring absorption of background quasar radiation [25]. These clouds occur in regions far from galaxies where little or no stellar nucleosynthesis has taken place. It is reasonable to suppose that the clouds are representative of matter that evolved very little since primordial nucleosynthesis. Hence, measuring the deuterium abundance compared to the hydrogen abundance in the clouds produces a new observational limit. $\Omega_b h^2$ from this technique have yielded a value slightly higher than the previous bound, $\Omega_b h^2 = 0.024 \pm 1.00$ 0.03 [25]. As more quasar/Lyman- α systems are measured, it will be possible to test the consistency of the method and perhaps improve the constraint further, making it the most precise method for constraining primordial nucleosynthesis and measuring the $\Omega_b h^2$.

Some may worry about the fact that these measurements are in disagreement with what was before the preferred range. More specifically, the measurements are marginally inconsistent with limits on primordial helium abundance. Some have even spoken about there being a potential crisis in big bang cosmology [26]. However, the the apparent, marginal contradiction is more likely due to underestimated systematic errors. In terms of Ω_b , the higher value of $\Omega_b h^2$ obtained from quasar absorption would permit Ω_b as high as 10%, still well below the critical density. It is possible that further improvements will lead to a higher value

yet for $\Omega_b h^2$, but it is hard to imagine that it will exceed 0.035, a range ruled out by additional constraints from solar system measurements. Even allowing a value of $\Omega_b h^2$ near this uppermost bound, Ω_b is constrained to be less than 15%. Hence, even the crudest limits justify our remark that $\Omega_{\text{matter}} \geq 0.20$ is convincing evidence of significant non-baryonic matter in the universe.

The upper bound on Ω_{matter} is also extraordinarily important. In a recent study, J. Ostriker and I mapped all of the constraints on $\Omega_{\rm matter}$ and the Hubble expansion rate coming from observations [27]. Figure 7.4 is an updated version using the new constraints on $\Omega_b h^2$ obtained from the Lyman- α cloud measurements [25]. The plot illustrates the constraints coming from measurements of the Hubble constant directly, measures of the age of the universe, measurements of large scale structure, measurements of primordial nucleosynthesis, x-ray luminosity in gas clusters, and bounds on the cosmological constant.

The striking feature is that there is a substantial range of the Ω_b -*h* plane which is in agreement with all known astrophysical constraints. Furthermore, the range of "cosmic concordance" is well above $\Omega_m = 0.2$ but also well below $\Omega_m = 1$. If the state of affairs does not change, then two striking conclusions emerge: significant amounts of non-baryonic dark matter exist in the universe, *and* there is some additional "missing energy" density in the universe accounting for the difference between Ω_m and unity. At this point, the measurements of the cosmic microwave background anisotropy are also consistent with this trend, but the experimental uncertainty is too large to reach any firm conclusions.

Although the case for missing energy is not conclusive, it is important to appreciate that the current indications come from a variety of observations, and that a combination of changes in disparate measurements is needed to change the qualitative conclusions. The importance of non-baryonic matter has been discussed above and is well-known in the physics community. The notion of missing energy is much less well-known, and, given the observational justification, is an idea worth exploring.

7.5 IF $\Omega_{\text{matter}} < 1$ AND $\Omega_{\text{total}} = 1$, WHAT ELSE IS THERE?

If future observations reveal that the universe is flat but also that $\Omega_{\rm matter}$ is less than unity, then then there must exist another contribution to the total energy density of the universe. This additional energy, which we have called "missing energy," must have a different equation-of-state from ordinary matter and nonbaryonic dark matter. While missing energy is consistent with inflation and other cosmologies, it is not predicted by any model. The existence of missing energy would surely be one of the most surprising discoveries in the history of cosmology and would immediately produce a puzzle: what is the nature of the missing energy? This question emerges as one of the most important cosmological issues of the 21st century.

The first candidate for missing energy likely to be considered is vacuum density or cosmological constant. However, this choice is based on historical familiarity; it is not uniquely dictated by the observational evidence. In fact, all the data that has been described in this paper would only tell us the amount of missing energy and that this energy has some equation-of-state different from that of ordinary matter. It would not tell what the equation-of-state is.

The equation-of-state is defined as the ratio of the pressure to the energy density, α . Vacuum density or cosmological constant corresponds to $\alpha = -1$ and matter density has $\alpha = 0$. The missing energy could have a value in between, and perhaps the equation-of-state is time-dependent. (For the purposes of this discussion, we do not consider $\alpha > 0$ since this would lead to a universe with a shorter lifetime than a flat model with $\Omega_m = 1$, which is already marginally

Figure 7.4: A plot in the the Ω_m -h plane showing the range of parameters in concordance with the known astronomical observations. The figure differs from the plot shown in Ostriker and Steinhardt, Ref. [27], in that a higher value of Ω_b is assumed, in accordance with recent limits on deuterium abundance from measurements of Lyman- α absorption of quasar emission [25]. The shaded region (stripes) is the concordance domain for flat CDM + Λ models.

in conflict with lower bounds on the lifetime of the universe.) A simple example of missing energy different from matter or vacuum energy is the field energy associated with a field rolling down an exponential potential, $\exp(-\beta \phi)$, which can have an equation-of-state which lies between $\alpha = 0$ and -1 , depending upon the coefficient, β .

The effect of missing energy is to change the expansion rate of the universe at recent epochs. If the missing energy has equation-of-state $\alpha = p/\rho$ =constant, then it will contribute an energy density which scales $1/a^m$, where a is the scale factor which describes the stretching of the universe and *m* can be simply related to the equation-of-state, $m = 3(1 + \alpha)$. For the cosmological constant, $\alpha = -1$ and ρ = constant. For more general forms of missing energy with α < 0, the energy density decreases more slowly than matter or radiation energy. In order for the missing energy to be a substantial fraction of the critical density today, it must have been an insignificant fraction of the critical density in the early universe.

If identifying the missing energy emerges as one of the critical questions of the 21st century, the first challenge will be to determine observationally its equationof-state, α . This may be one of the most difficult tasks in observational cosmology. For example, consider the most powerful tool for measuring cosmological parameters, the cosmic microwave anisotropy. Figure 7.5 illustrates a sequence of curves indicating the predicted microwave background anisotropy. The lowest curve corresponds to the prediction for an inflationary model with $\Omega_{\text{matter}} =$ 1. The upper curves correspond to an inflationary model (flat) with a matter density which is 35% of the critical density. The sheath about them indicates the cosmic variance, or theoretical uncertainty in the inflationary prediction. A close look reveals that there is a whole sequence of curves which lie within the cosmic variance sheath and, hence, cannot be distinguished observationally. These correspond to $\Omega_{\alpha} = 0.65$, with the equation-of-state given values between -0.5 and -1 . Hence, the CMB power spectrum does not provide a precise way of determining α . (N.B. The figure also shows that, for $\alpha > -0.5$, the spectrum undergoes more dramatic changes in shape which make it possible to distinguish from α < -0.5.) In fact, I have looked for several ways of distinguishing the equation-of-state, including measurements of the luminosity distance-red shift relation for cosmic supernovae and other standard candles in the universe. Thus far, it appears that one cannot determine α precisely by any known method. Should observational evidence establish that there exists missing energy, finding a method for precisely measuring its equation-of-state will emerge as a grand challenge for observational cosmology in the 21st century.

Not only is there the challenge of explaining what the missing energy is, but, in addition, there remains a puzzling "cosmic coincidence." Figure 7.6 shows a plot of the energy density of the universe versus the scale factor. The matter energy density, decreasing as $1/a³$ and the missing energy density, decreasing as $1/a^m$ with $m < 3$, are indicated. The two lines cross at a time which must correspond to the present epoch, since the matter and missing energy density are

Figure 7.5: A plot of the cosmic microwave background anisotropy power spectrum multipoles vs. multipole moment illustrating the dependence on equation-of-state (α) . The middle dashed curve represents standard cold dark matter model with $\Omega_m =$ 1. This is easily distinguished from the upper family of (solid) curves corresponding to cosmologically flat models with $\Omega_m = 0.35$ and missing energy $\Omega_{\text{unknown}} =$ 0.65. The only difference among the upper curves is the equation-of-state, α , which varies between -1 (vacuum energy density) and -1/2. There is negligible change among the upper curves compared to the cosmic variance (sheath surrounding the curves). If α is decreased further, then the difference in the equation-of-state causes a distinguishable spectrum, e.g., see $\alpha = -1/3$ dot-dashed curve. Some recent data from COBE, Saskatoon, and CAT experiments [16] are superimposed [21].

comparable. Why we should happen to live at this special crossroads? If the missing energy turns out to be the cosmological constant or vacuum density then the missing energy density is strictly constant, a horizontal line in the plot. It is hard to imagine a physical explanation for the coincidence. On the other hand, if the missing energy turns out to have some other equation-of-state, $\alpha > -1$, then it may be due to some field or fields, and one could at least imagine that there are interactions between the missing energy and ordinary matter that might explain naturally why the two energy densities should be nearly equal to one another. Hence, distinguishing whether the missing energy is vacuum energy or not will be essential to theoretical understanding of the apparent cosmic coincidence.

Figure 7.6: A schematic log-log plot of energy density ρ versus scale factor α illustrating the behavior of the matter density ρ_m and putative missing energy density ρ . The missing energy density decreases more slowly than the matter density. Hence, the missing energy was insignificant in the early universe. The circle represents the cosmic coincidence in which the the missing energy and the matter energy are comparable today.

[7.6 ULTIMATE CHALLENGE](#page--1-0)

Should cosmologists meet all of the challenges described in the previous sections without major changes in the current paradigm, then the field will probably proceed along two different paths. One direction is a deeper understanding of largescale structure formation and evolution. Large-scale structure already provides some of the most important cosmological tests, and will become an even more powerful constraint as forthcoming red shift surveys are completed. Here, the strategy is to use statistical properties of large-scale structure over the observable universe to constrain cosmological models and parameters. However, the details and variations in large-scale structure formation are interesting as well. What role, for example, does the ionization of the history of the universe have on the formation of the first structures in the universe? In spite of all the progress anticipated in the previous sections, a realm that will remain difficult to probe is below red shift $z = 1000$, where the cosmic microwave background anisotropy provides a snapshot, and above $z = 10$, beyond the reach of optical, x-ray, and infrared measurements. Theorists and observers will struggle to understand how the universe evolved from the tiny ripples seen in the cosmic microwave background to the first stages of large-scale structure formation.

The other path for cosmology will be towards answering its ultimate challenge: why is the universe the way it is? How did the universe we see naturally spring from the fundamental laws which govern the universe? The answers to these questions will not come from observation cosmology so much as from fundamental physics. The issues that fundamental physics must ultimately explain include:

- Nature of dark matter and missing energy
- Initial conditions of the universe
- Value of the cosmological constant
- Mechanism that drives inflationary cosmology

I have listed these issues as topics for future research, but in fact, they are so irresistibly tempting that a significant number of papers have been written on these issues already. There are a large number of candidates proposed for non-baryonic dark matter, fewer proposals for candidates for missing energy. There is the intriguing work of Hartle and Hawking on the wave function of the universe as a proposal for explaining initial conditions [28]. There is also Linde's chaotic, selfreproducing universe picture [29]. Many brands of inflationary models have been proposed which rely on different detailed mechanisms for initializing inflation and bringing inflation to a halt.

However, my own intuition is that some of these speculations may be premature. My special concern is for ideas that depend heavily on processes that take place uncomfortably close to the Planck scale. Many of these ideas rely on field theoretic ideas that may not be valid. For example, recent progress in string theory or M -theory strongly suggests a sequence of compactification scales; this means that field theory may be a bad approximation until the energy density in the universe reduces to 10^{16} GeV or less. In that case, the Hartle-Hawking wavefunction and Linde's chaotic, self-reproducing universe may have to be discarded as explanations of the initial conditions since they are based on quantum field theory extrapolated to near the Planck scale. Topological defect models of structure formation are also in jeopardy because they require the Kibble-mechanism [30], a semiclassical field-theoretic notion of bubble nucleation and collision, to produce cosmic strings or textures at masses of order 10^{17} GeV, as required to explain structure formation. Similarly, the monopole problem [31], one of the original motivations for inflationary cosmology, may be obviated since it, too, depends on a Kibble mechanism operating at near Planckian scales. (In supersymmetric models, the monopole mass is above 10^16 GeV.)

Inflationary cosmology is not necessarily affected directly by physics near the Planck scale. It is certainly possible to construct models in which inflation occurs at energies well below 10^{16} GeV. However, my own intuition is that inflation is enmeshed in strings and Planck-scale physics as well. At present we describe inflation in terms of a scalar field or a scalar condensate of fermions evolving along an effective potential. The effective potential has various features, such as a flat region where the energy density remains nearly constant, a steep region for reheating, and a graceful end to inflation. (The exceptions, chaotic inflation models, rely on near Planck-scale physics.) All of the field theoretic scaffolding is there for one simple purpose: to create a change in the equation-of-state from an epoch in which $\alpha < -1/3$ and the universe inflates, to an epoch where $\alpha \geq 0$ and the universe returns to a decelerating, hot big bang universe. It is conceivable that the same can be achieved without invoking any scalar fields or fermions. For example, some of us are looking at recent progress in understanding stringy properties in the vicinity of black hole horizons to provide insights about stringy properties in a different setting, de Sitter (inflationary) horizons. In the end, it may simply be the behavior of superstrings at high temperatures and high densities that automatically produces this change in the equation-of-state in a manner that cannot be understood from the point-of-view of field theory.

The comments in this section are highly speculative since they are based on highly uncertain guesses about fundamental physics and cosmology. The principal point is that, as each proceeds, fundamental physics and cosmology will have to be reconciled. I have argued that this may not be the time for reconciliation since both are too unsettled. But, based on the current rate of progress, the right time to begin may lie within the $21st$ -century window. It promises to be one of the most exciting and profound steps in the history of science, drawing together all of our knowledge of space, time, energy, and matter. And, most likely, bringing together detailed knowledge of cosmology and fundamental physics will produce numerous new puzzles and debates that will carry beyond the 21^{st} century. Of all of the predictions I have made, the surest is the last one: that cosmology will once again be a hot topic of discussion one hundred years from now when our great grandchildren meet to celebrate at the Princeton 350th Anniversary Celebration.

ACKNOWLEDGEMENTS

I would like to thank my many collaborators whose work and discussions have shaped the views expressed in the paper.This research was supported by the DOE at Penn (DOE-EY-76-C-02-3071).

REFERENCES

- [1] A.H. Guth, Phys. Rev. D **23** 347, (1981).
- [2] V. de Lapparent, M.J. Geller, and J. Huchra, Ap. J. **302** LI (1986); ibid., **332** 44(1988).
- [3] S.A. Shectman, S.D. Landy, A. Oemler, D. Tucker, R.P. Kirshner, H. Lin, and P. Schechter, Ap. J. **470** 172 (1996).
- [4] G.F. Smoot et al., Ap. J. **396** LI (1992); C.L. Bennett et al., submitted to Ap. J. (1994); K.M. Gorski, et al., Ap. J. **430** L89 (1994).
- [5] A.D. Linde, Phys. Lett. **108B** 389 (1982); A. Albrecht and P.J. Steinhardt, Phys. Rev. Lett. **48** 1220 (1982).
- [6] A.H. Guth and P.J. Steinhardt, "The Inflationary Universe" in *The New Physics,* edited by P. Davies (Cambridge U. Press, Cambridge, 1989), 34-60. Universe, p. 34-60.
- [7] A. D. Linde, in *Particle Physics and Inflationary Cosmology,* (Harwood Academic Publishers, New York, 1990).
- [8] J. Bardeen, P.J. Steinhardt and M.S. Turner, Phys. Rev. D **28** 679 (1983); A.H. Guth and S.-Y. Pi, Phys. Rev. Lett. **49** 1110 (1982); A.A. Starobinskii, Phys. Lett. B **117** 175 (1982); S.W. Hawking, Phys. Lett. B **115** 295 (1982).
- [9] R.H. Dicke and P.J.E. Peebles, in *General Relativity: An Einstein Centenary Survey;* edited by S.W. Hawking and W. Israel (Cambridge: Cambridge U. Press, 1979), 505-517.
- [10] E.R. Harrison, Phys. Rev. D 1 2726 (1970); Ya.B. Zel'dovich, Mon. Not. R. Soc. **160** lp (1972); P.J.E. Peebles and J.T. Yu, Astrophys. J. **162** 815 (1970).
- [11] V.A. Rubakov, M.V. Sazhin, A.V. Veryaskin, Phys. Lett. B **115** 189 (1982).
- [12] A.I. Starobinskii, Sov. Astron. Lett. **11** 133 (1985).
- [13] L.F. Abbott M. Wise, Nucl. Phys. B **244** 541 (1984).
- [14] Tilt in inflationary perturbation spectra was first noted in J. Bardeen, P. J. Steinhardt and M. S. Turner, ibid.
- [15] R.L. Davis, H.M. Hodges, G.F. Smoot, P.J. Steinhardt, and M.S. Turner, Phys. Rev. Lett. **69** 1856 (1992).
- [16] P.J. Steinhardt, in *Proceedings of the Snowmass Workshop on the Future of Particle Astrophysics and Cosmology,* edited by E.W. Kolb and R. Peccei, (World Scientific, 1995); IJMPA **A10** 1091-1124 (1995).
- [17] W. Hu, in *The Universe at High-z, Large Scale Structure and the Cosmic Microwave Background,* edited by E. Martinez-Gonzalez and J.L Sanz (Springer Verlag); W. Hu, N. Sugiyama, and J. Silk, Nature, to appear.
- [18] R. Crittenden, R.L. Davis, and P. Steinhardt, Ap. J. **L13** (1993).
- [19] R. Crittenden, J.R. Bond, R.L. Davis., G. Efstathiou, and P.J. Steinhardt, Phys. Rev. Lett. **71** 324 (1993).
- [20] M. Kamionkowski, D.N. Spergel, N. Sugiyama, Ap. J. Lett. **426** 57 (1994).
- [21] J.R. Bond, R. Crittenden, J.R. Bond, R.L. Davis., G. Efstathiou, and P.J. Steinhardt, Phys. Rev. Lett. **72** 13 (1994).
- [22] G. Jungman, M. Kamionkowski, A. Kosowsky, D.N. Spergel, Phys. Rev. D **54** 1332 (1996).
- [23] W. Hu and M. White, Phys. Rev. Lett. **77** , 1687 (1996).
- [24] C. Copi, D.N. Schramm, and M.S. Turner, Science **267** 192 (1995).
- [25] D. Tytler, X.-M. Fan, and S. Buries, Nature **381** 207 (1996); D. Tytler, S. Buries, and D. Kirkman, astro-ph/9612121
- [26] N. Hata, G. Steigman, S. Bludman, P. Langacker, Phys. Rev. D **55** 540 (1997).
- [27] J.P. Ostriker and P.J. Steinhardt, Nature **377** 600 (1995).
- [28] J. Hartle and S.W. Hawking, Phys. Rev. D **28** 2960 (1983).
- [29] A.D. Linde, Phys. Lett. **129B** 177 (1983); A.D. Linde, Mod. Phys. Lett. Al 81 (1986); Phys. Lett. **175B** 395 (1986).
- [30] T.W.B. Kibble, J. Phys. **A9** 1387 (1976).
- [31] J.P. Presskill, Phys. Rev. Lett. **43** 1365 (1979).

[7.7 DISCUSSION](#page--1-0)

Session Chair: Anthony Hewish Rapporteur: Andrew Millard

NAPPI: You mentioned "typical inflationary models." Can you tell us if there is a model of inflation favored at the moment and which one it is?

STEINHARDT: A variety of mechanisms of initial conditions have been discovered over the past decade which produce an acceptable inflationary scenario. Although individual theorists have their preferences, there are not reliable guiding principles to determine which variant is correct. Perhaps more specific information will come from what we ultimately learn about fundamental physics close to the Planck scale. There are many detailed ways of accomplishing the same general changes in equations of state and solving the cosmological problems. The predictions discussed in the paper, though, are not sensitive to the particular variant of inflation and, hence, serve as generic tests of inflation. The main lesson is that there is more than one way of doing things.

SPERGEL: In addition to the problems you mentioned, determination of the polarization of the microwave background and observation of high red-shift supernovae are two other potentially powerful probes that may help to distinguish between cosmological parameters.

STEINHARDT: I agree, but the interpretation of these measurements is modeldependent. The parameters can be determined only once the underlying model, such as inflation or cosmic defects, is settled first.

MANN: Suppose it might become possible to measure the neutrino cosmic background radiation, just as the photon background has been measured; what additional constraints would this place on cosmological theories?

STEINHARDT: My initial thought is that the results could be reasonably expected to corroborate and refine existing information. The neutrino background won't have formed at a much higher energy than the photon background (only six orders of magnitude), although something unexpected may have happened between the decouplings which would cause a different between the two backgrounds.

BAHCALL: Several observations limit the cosmological constant with, for instance, measurements of lensing and high red-shift supernovae implying a low value. From the microwave background spectrum, you show a strong peak which may suggest a high cosmological constant value. Are those two inconsistent with each other?

STEINHARDT: I don't think so. All these techniques are in early stages of evolution. Lensing statistics as a function of red shift depend on knowing the effects of evolution and dust. Recent work suggests that the dust effect is large enough to erode the constraint on the cosmological constant to a level where it is consistent with the other data shown in the paper. Cosmic supernovae measurements are a newer and highly promising approach for constraining the cosmological constant. Current limits are tight, but there remains a substantial range of parameter space consistent with the supernovae measurements and the other measurements shown in the paper. It is premature to read too much into microwave background data at the present time. The exciting thing is that all of these techniques will be refined over the next few years.

PARTRIDGE: This morning John Hopfield spoke of "the mark of Cain," the desire to explain complex biological systems in terms of simple fundamental laws. You seem to take the same approach, looking to fundamental physics to answer cosmological questions. Could you say a little bit about the possibility that messy astrophysics—with gamma ray bursts, gravitational lensing, and high red-shift yet elderly galaxies—could provide some interesting cosmological answers in the next century?

STEINHARDT: New phenomena provide added details about evolution of cosmic structures that constrain cosmological scenarios, but their relation to microscopic physics appears to be too remote to use them as a precise tool for testing fundamental physics near the Planck scale.

SCULLI: Several times you said that information concerning fundamental physics was needed, especially knowledge about physics near the Planck scale. Where might this come from? Experimental information? String theory?

STEINHARDT: The information I have in mind relies on progress in our understanding of Planck scale physics, such as superstring theory, and its indirect corroboration in low energy experiment, to be discussed in Saturday's talks.