

AMiBA 2001: High-z Clusters, Missing Baryons, and CMB Polarization
ASP Conference Series, Vol. 999, 2002
L-W Chen, C-P Ma, K-W Ng and U-L Pen, eds

Studies of Structure Formation and Cosmology with Galaxy Cluster Surveys

Joseph J. Mohr

Departments of Astronomy and Physics; University of Illinois, 1002 W. Green St.; Urbana, IL 61801 USA

Abstract.

Surveys of galaxy clusters provide a promising method of testing models of structure formation in the universe. Within the context of our standard structure formation scenario, surveys provide measurements of the geometry of the universe and the nature of the dark energy and dark matter. Cluster catalogues will be constructed using some combination of X-ray, optical/near-IR, and mm or cm-wave observations. These catalogues will be used to study the cluster redshift and mass distributions along with the correlations of the cluster spatial distribution. These measurements probe the volume-redshift relation, the power spectrum of density fluctuations and the evolution of galaxy cluster abundance. All are sensitive to the amount of dark matter Ω_M , the amount of dark energy Ω_E , the equation of state of the dark energy $w(z)$ and any other parameter, which affects the expansion history of the universe.

1. Introduction

Over the last few years, cosmological constraints from Type Ia SNe (Schmidt et al. 1998; Perlmutter et al. 1999), cluster baryon fractions (White et al. 1993a; David et al. 1995; White & Fabian 1995; Burles & Tytler 1998; Mohr et al. 1999; Arnaud & Evrard 1999), the cosmic microwave background (CMB) anisotropy (Hanany et al. 2000; Jaffe et al 2000; Lange et al. 2001) and other complementary measures (Bahcall et al. 1999, and references therein) have pointed toward a dark energy dominated universe ($\Omega_\Lambda \sim \frac{2}{3}$), with a significant dark matter component ($\Omega_m \sim \frac{1}{3}$) and a trace of baryonic matter. The recent detections of the 2nd and 3rd acoustic peaks in the CMB anisotropy (Halverson et al. 2001; Netterfield et al. 2001; Pryke et al. 2001) lend additional support to these conclusions and bring several important questions into sharp focus. At the dawn of this new era of precision cosmology, the important questions concern the very nature of the dark matter (collisionless or self-interacting) and the characteristics of the dark energy (which we can parametrize by the equation of state parameter w , where the pressure $p = w\rho$).

Recent theoretical and experimental developments make future cosmological studies that utilize galaxy clusters extremely promising. One particularly promising approach is the use of galaxy cluster surveys, which enable one to measure the cluster redshift distribution and the correlations in the cluster spatial

distribution. Surveys are now being carried out using cluster X-ray emission, the near-IR/optical light from cluster galaxies, the distorted morphologies and alignment of background galaxies, and the effect that hot electrons within clusters have on the cosmic microwave background (the so-called Sunyaev-Zel'dovich effect or SZE; Sunyaev & Zel'dovich 1972). To use these surveys to full effect in cosmology studies, we must first test the standard model of structure formation. In addition, we must sharpen our understanding of the nature and evolution of galaxy cluster internal structure and the relationships between cluster observables (i.e. SZE decrement, X-ray emission, galaxy light) and the cluster halo mass.

In these proceedings we describe a fundamental test of the hierarchical structure formation model, and then we examine in some detail the cosmological dependences of the cluster redshift distribution. We end by highlighting some of the challenges that currently exist in using cluster surveys to precisely constrain cosmological quantities like the equation of state of the dark energy.

2. Structure Formation Constraints from High- z Cluster Surveys

Because of the nature of the power spectrum of density fluctuations, we expect that structure formation proceeded hierarchically from small to ever larger scales (i.e. Peebles 1993 and references therein). Low mass galaxy clusters ($\sim 10^{14} M_{\odot}$) are expected to first emerge at redshifts of $z = 2$ to $z = 3$ within the currently favored model. Higher mass clusters ($\sim 10^{15} M_{\odot}$) appear later at lower redshifts. An appealingly powerful test of structure formation would be to probe the cluster population with sufficient sensitivity to detect the first emerging low mass systems. High sensitivity SZE surveys are particularly well suited for studies of the high redshift galaxy cluster population, because of the redshift independence of the decrement ΔT :

$$\frac{\Delta T}{T_{cmb}} = -2 \frac{\sigma_T}{m_e c^2} \int d \ln_e k_B T_e, \quad (1)$$

where T_{cmb} is the cosmic microwave background (CMB) temperature, σ_T is the Thomson cross section, m_e is the electron rest mass, c is the speed of light, k_B is the Boltzmann constant and n_e and T_e are the electron number density and temperature. In other words, if one has a galaxy cluster described by a particular distribution of n_e and T_e , the magnitude of the SZE distortion of the CMB along a line of sight passing through the cluster would be independent of the cluster redshift. This together with our expectation for how cluster structure evolves with redshift, makes SZE instruments capable of detecting clusters of a particular mass no matter what that cluster's redshift (Holder et al. 2000). This is a particularly power approach for determining the redshifts when galaxy clusters first emerged.

In contrast to this SZE behavior, cluster X-ray emission (along with any emission) suffers from $(1+z)^4$ cosmological dimming. The X-ray surface brightness I_x is

$$I_x = \frac{1}{4\pi(1+z)^4} \frac{\mu_e}{\mu_H} \int d \ln_e^2 \Lambda(T_e), \quad (2)$$

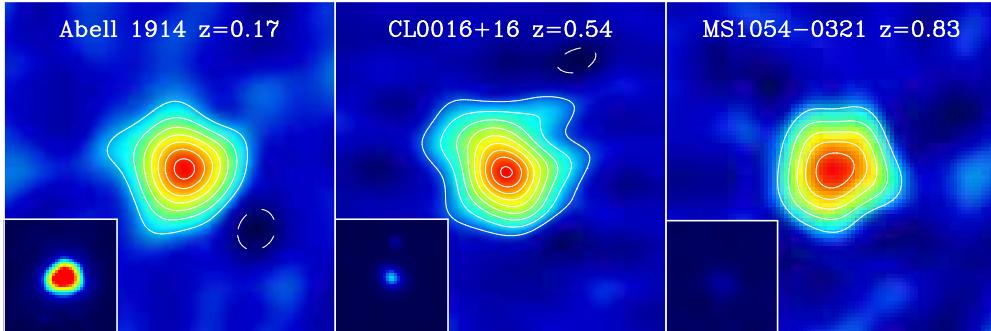


Figure 1. Interferometric SZE and ROSAT X-ray (inset) images of three galaxy clusters of comparable mass at redshifts $z = 0.2, 0.5$ and 0.8 . SZE contours are $75 \mu\text{K}$, and the X-ray color scale is the same in each cluster. Note that while the SZE signal remains comparable at increasing redshift, the X-ray surface brightness suffers cosmological dimming. This characteristic makes SZE observations particularly well suited to studies of high redshift clusters. (figure courtesy J.E. Carlstrom and J.J. Mohr)

where $n_e m_p \mu_e \equiv \rho$, m_p is the proton rest mass, ρ is the intracluster medium mass density, and Λ is the temperature dependent X-ray emission coefficient describing bremsstrahlung and line emission. This strikingly different behavior of X-ray emission and the SZE is qualitatively illustrated in Figure 1 by the panel of interferometric SZE observations (with X-ray image insets) of three clusters of comparable mass at redshifts $z = 0.2, 0.5$ and 0.8 . The SZE contours and X-ray color scales are the same for all three clusters. Although the cluster SZE signal is similar at all redshifts, the X-ray emission dims rapidly, as expected.

High sensitivity interferometric SZE surveys carried out with a new generation of SZE optimized interferometers will soon carry out the fundamental test of hierarchical structure formation described above. Three such instruments, the SZ-Array, AMiBA and AMI (all described elsewhere in this volume), are all funded and currently in various stages of construction.

3. Cosmological Constraints from the Cluster Redshift Distribution

Within the context of the standard structure formation scenario, it is possible to use cluster surveys to measure cosmological parameters. The abundance of galaxy clusters and its redshift evolution have been recognized as sensitive probes of the normalization of the power spectrum and the mean matter density in the nearby universe (White et al. 1993b; Viana & Liddle 1999). The parameter degeneracy between σ_8 , the *rms* amplitude of mass fluctuations in the universe filtered on an $8h^{-1}$ Mpc scale, and the matter density parameter Ω_m can be broken by extending cluster surveys to higher redshift (Bahcall et al. 1997). This particular probe is highly complementary to the CMB anisotropy, because it probes the era of structure formation—when dark energy becomes dominant—as opposed to the era of recombination.

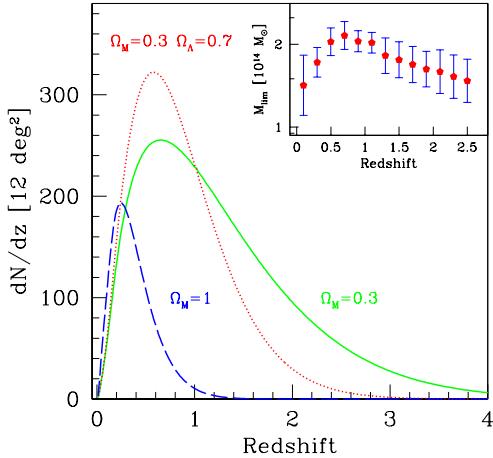


Figure 2. The cluster redshift distribution of an interferometric SZE survey similar to the one planned with the SZA. All curves are normalized to produce the observed local abundance of massive clusters. The redshift distribution is sensitive to cosmological parameters. Note the inset, which shows the limiting mass of clusters, which result in 5σ detections in mock SZE observations.

The observed cluster redshift distribution in a survey (see Figure 2) is the comoving volume per unit redshift and solid angle $dV/dz d\Omega$ times the comoving density of clusters n_{com} with masses above the survey detection limit M_{lim} : written as

$$\frac{dN}{dz d\Omega} = \frac{dV_{com}}{dz d\Omega} n_{com} = \frac{c}{H(z)} d_A^2(z) (1+z)^2 \int_{M_{lim}(z)}^{\infty} dM \frac{dn}{dM}, \quad (3)$$

where dn/dM is the cluster mass function, $H(z)$ is the Hubble parameter as a function of redshift and d_A is the angular diameter distance. The cosmological sensitivity comes from the three basic elements:

- **Volume:** the volume per unit solid angle and redshift depends sensitively on cosmological parameters (i.e. higher Ω_Λ or lower Ω_m increases the volume per solid angle). Figure 3 (left) is a plot of the comoving volume element ($dV/dz/d\Omega$) versus redshift for three cosmological models. Note the rapid increase in the volume element at modest redshift, which is responsible for the rapid rise in the cluster redshift distribution in Figure 2. At higher redshift the comoving volume element flattens out and eventually turns over.

The cosmological sensitivity of the distance-redshift and volume-redshift relation derives essentially from the expansion history of the universe $E(z)$, where $H(z) = H_0 E(z)$, where H_0 is the Hubble parameter and the parameter $E(z)$ describes its evolution. Within our cosmological framework, the expansion history of the universe simply depends on the nature and amount of the constituents that make up the universe. That is, $E^2(z) = \Omega_M (1+z)^3 + (1 - \Omega_M - \Omega_E) (1+z)^2 + \Omega_E (1+z)^{3(1+w)}$.

- **Abundance:** the number density of clusters at a given redshift depends sensitively on the growth rate of density perturbations. This growth rate is highly sensitive to cosmology (i.e. higher Ω_m speeds the growth of density perturbations so that clusters “disappear” more quickly as we probe to higher redshift). Figure 3 (right) is a plot of the comoving abundance of

clusters above a fixed mass, where the abundance is normalized to reproduce the observed local abundance of massive clusters. Note that abundance differences increase dramatically with redshift and are responsible for the high redshift ($z > 1$) behavior of the cluster redshift distribution (Figure 2).

As stated above, the cosmological sensitivity of the abundance evolution appears to derive from the growth rate of density perturbations. Within the linear regime, the differential equation that describes growth depends, again, on the expansion history of the universe $E(z)$. The rapid evolution of the abundance is due to an exponential dependence of abundance upon the amplitude of density fluctuations on the galaxy cluster scale (Press & Schechter 1974; Jenkins et al 2001).

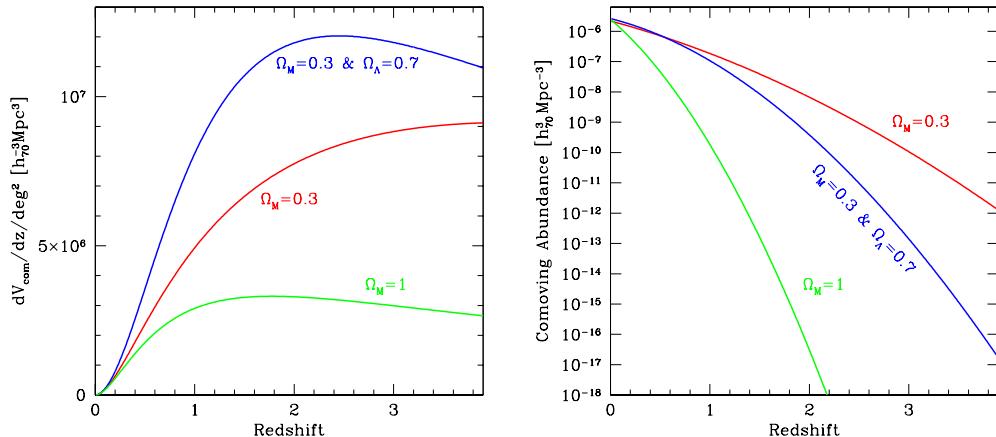


Figure 3. The comoving volume element (left) and cluster abundance above a fixed mass (right) in three different cosmological models. The abundances are normalized to produce the observed local abundance of massive clusters. Differences in the cluster redshift distribution are dominated by volume at low redshift and by abundance at high redshift.

- **Mass limit:** the mass of a cluster, which is just luminous enough to appear above the detection threshold, typically depends on the luminosity or angular diameter distance as well as the evolution of cluster structure— both are sensitive to cosmological parameters. The survey yield and redshift distributions are both sensitive to the limiting mass, as indicated in Figure 4. Figure 4 shows the cluster redshift distribution in a fiducial cosmology for a limiting mass of $M = 2 \times 10^{14} M_\odot$, and for limiting masses 10% above and below this value.

The challenging aspects of using galaxy clusters to constrain cosmology, aside from building the instruments to carry out the surveys, include an understanding of how cluster abundance evolves within a variety of cosmologies and how to relate cluster observables like X-ray emission, SZE distortion, galaxy light and weak lensing shear to halo mass. These relations are required for all redshifts. Theoretical studies of structure formation suggest that the mass function

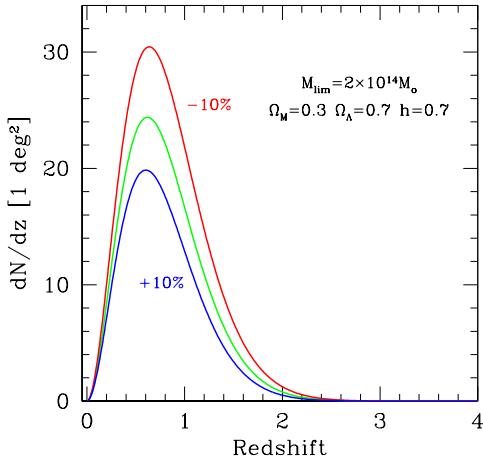


Figure 4. The cluster redshift distribution within a fiducial cosmological model for a mass limit of $M = 2 \times 10^{14} M_{\odot}$ and for mass limits 10% above and below this value. The mass sensitivity of the survey yields and redshift distributions means that accurate cosmological constraints require unbiased estimators of cluster halo mass.

$dn/dM(z)$ is well behaved, and may be described by a “universal” form when suitably parametrized (Jenkins et al 2001; White 2001). Further study is clearly required. Observational studies of galaxy cluster scaling relations suggest regularity in the cluster population similar to the regularity in the elliptical galaxy population (Mohr & Evrard 1997; Mohr, Mathiesen & Evrard 1999; Horner, Mushotzky & Scharf 1999). Hydrodynamical simulations of cluster formation suggest that scaling relations between cluster observables and halo mass evolve in a simple way (Evrard et al. 1996; Bryan & Norman 1998), even in the presence of some early preheating (Bialek, Evrard & Mohr 2001). Much more study using higher resolution simulations that incorporate additional physics is clearly required here to enable more accurate, unbiased estimators of cluster mass.

A recent study by Diego et al (2001) suggests that a joint analysis of the cluster redshift distribution and the observed scaling relations (all available from the same survey data) can allow one to solve for the evolving scaling relation and cosmological parameters simultaneously (see also Verde, Haiman & Spergel 2001). More complete studies of the degeneracies between the evolution of cluster scaling relations and cosmological parameters in the analysis of cluster surveys is ongoing.

4. Precision Cosmology with Galaxy Cluster Surveys?

Recently, Haiman, Mohr & Holder (2001) emphasized that large cluster surveys extending to high redshift can in principle provide precision measurements of any cosmological parameter, which affects the expansion history of the universe (i.e. Ω_m , Ω_Λ , and w). This is emphasized in Figure 5, which shows the 1σ , 2σ and 3σ joint constraints on Ω_m and the equation of state parameter w of the dark energy for an X-ray cluster survey which yields $\sim 10^3$ clusters with measured emission weighted mean temperatures (and therefore virial mass estimates). Only flat models ($\Omega_m + \Omega_\Lambda = 1$) are considered and the fiducial model $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, constant $w = -1$ and $h = 0.65$ is adopted. Note the $\Omega_m - w$ degeneracy. Also shown (dashed line) is the $\Omega_m - w$ degeneracy for CMB anisotropy and SNe Ia distance measurements. The CMB degeneracy assumes that the angular scale of the first peak (at fixed $h = 0.65$) is known to an accuracy of 1%,

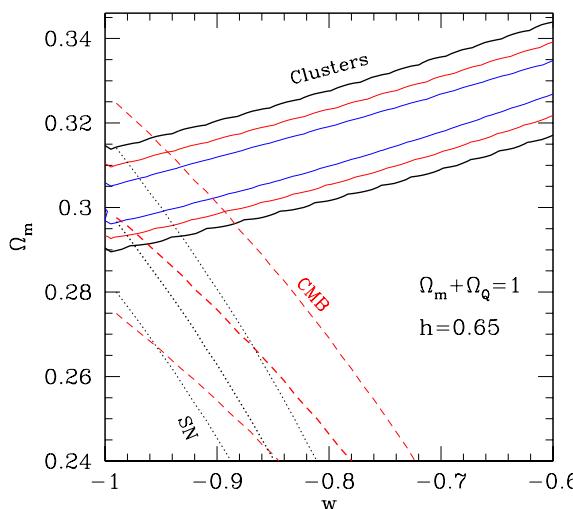


Figure 5. We plot estimated confidence regions for joint constraints on Ω_m and the equation of state parameter w of the dark energy from an X-ray survey proposed in a NASA Small Explorer competition. All cosmological models are flat, and a fiducial model of $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ and $w = -1$ is assumed. Also shown are the confidence regions which correspond to a 1% measurement of the angular scale of the first acoustic peak in the CMB anisotropy and a 1% measurement of the luminosity distance to $z = 1$. This figure illustrates the power of a cluster survey, which yields $\sim 10^3$ clusters with measured temperatures. In addition, it shows that the parameter degeneracy in the cluster constraints is roughly orthogonal to the parameter degeneracy from the CMB or SNe Ia measurements. (figure from Haiman, Mohr & Holder 2001)

whereas the SNe Ia degeneracy assumes that the luminosity distance to $z = 1$ is known to 1%. This figure indicates that an X-ray survey yielding 10^3 clusters has comparable constraining power to 1% CMB or SNe Ia measurements. In addition, the roughly orthogonal degeneracy between the cluster constraints and those from the CMB and SNe emphasizes the complementarity of these independent constraints on cosmological parameters.

The requirements for such precision are (1) a large cluster sample extending to intermediate or high redshift and (2) cluster mass estimators that are unbiased at the $\sim 5\%$ level. Assuming these two requirements can be satisfied, cluster surveys have as much potential to reveal the nature and amount of the dark energy in our universe as either high redshift type Ia supernovae observations or observations of the anisotropy of the cosmic microwave background.

In fact, with accurate mass estimators, the cluster redshift distribution is far more cosmologically informative than simple distance measurements. Figure 6 contains a plot, which compares the cosmological sensitivity of SNe Ia distance estimates to cluster survey constraints, assuming distances and cluster masses are both accurately estimated from the data. The figure plots the ratio of the cluster redshift distribution (solid lines) and the luminosity distance (dashed lines) as a function of redshift for a few cosmological models. The denominator in each ratio is the quantity from the fiducial model, taken to be $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$ in this example. The degree to which the ratio deviates from 1.0 provides an indication of the sensitivity to differences in the two models. Except for a narrow window around $z = 1$, cluster redshift distribution contains more cosmological information than do luminosity distances. Given the discussion

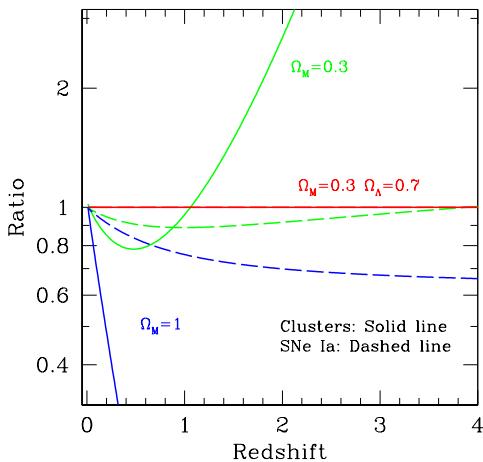


Figure 6. The ratio of the cluster redshift distribution (solid line) and the luminosity distance (dashed line) as a function of redshift for three cosmological models. The fiducial model is $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$, and the redshift distribution and distances from this model appears in the denominator of all ratios. Aside from a narrow window around $z = 1$, cluster redshift distributions are far more sensitive to cosmological parameters than are luminosity distances.

in Section 3 above, this is easy to understand. Surveys probe the volume–redshift relation, which scales as the square of the distance. Abundance evolution depends exponentially on the growth rate of density perturbations. In addition, limiting masses depend on luminosity or angular diameter distances. Naturally, these different dependencies can interfere constructively or destructively.

Of course, cluster surveys will only achieve high precision if cluster masses can be accurately estimated, on the average, from observables. That is, precision requires that systematic biases in mass estimators be small. This is similar to the case with type Ia supernovae, in that SN Ia distances are only accurate to the extent that the SNe themselves are standard candles. The high potential of both approaches has led people to invest significant effort in better understanding possible sources of systematics. Currently, our theoretical understanding of the formation and evolution of clusters is less developed than our understanding of the dynamics of density perturbations (well within the linear regime) at and before the epoch of recombination; however, it seems to me that our understanding of structure formation has progressed well beyond our understanding of why SNe Ia’s form a one parameter family of standard candles that have not evolved since before the universe was one quarter its present age. This theoretical heritage in structure formation is an important resource as we move toward interpreting ongoing and planned surveys.

Precision cosmology with clusters requires large cluster ensembles extending over large (10^2 - 10^3 deg 2) solid angles. X-ray and SZE surveys of this sort will not necessarily have the optical/near-IR data available on every system to estimate a sufficiently accurate redshift. Nevertheless, X-ray and SZE surveys are attractive, because high signal to noise detections are possible, projection effects are minimized, and observations indicate that tight scaling relations involving X-ray observables exist. The price of not having redshifts can be severe, as shown in Figure 7. This figure shows the confidence regions in Ω_M and w corresponding to a 4000 deg 2 SZE observations indicate that tight scaling relations involving X-ray observables exist. survey carried out from the South Pole. Confidence regions include marginalization over σ_8 , and only flat models are

considered. The contours correspond to constraints using only the total number of detected clusters, whereas the solid region denotes the constraints in the case that cluster redshifts are available. As was made clear in Section 3, the cluster redshift distribution is cosmologically rich.

Possibilities for redshift followup include large solid angle, multiband optical or near-IR photometric surveys and direct spectroscopic followup of member galaxies. It may even be possible to obtain rough redshift estimates directly from the SZE or X-ray data, but further work is required to explore the feasibility of this approach (Diego et al, in preparation). Photometric redshift estimates of multiple galaxies within each cluster should be sufficiently accurate to allow precision cosmology. The required redshift precision is set less by the scale of change in the theoretical cluster redshift distributions (which are smoothly varying- see Figure 2) than by the fact that accurately inferring cluster masses from measured fluxes requires redshifts.

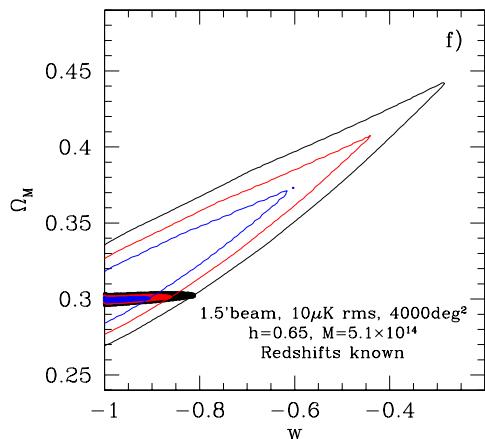


Figure 7. Confidence regions on Ω_M and mean w from a bolometric SZE survey carried out over 4000 deg^2 with a proposed South Pole 8 m telescope. This survey yields $\sim 20,000$ clusters extending to redshifts $z = 2$. This plot underscores the importance of having redshifts for detected clusters; the solid confidence regions correspond to constraints from the cluster redshift distribution, whereas the contours correspond to confidence regions from the total number of detected clusters.

The Sloan Digital Sky Survey is one such multiband, photometric dataset, and it will be extremely useful in estimating redshifts for clusters at $z < 0.6$ (i.e. relatively low redshift samples like the high mass *Planck* Surveyor cluster sample). However, for reasonably deep X-ray surveys and high sensitivity SZE surveys, only a small fraction of the sample will lie at these low redshifts, and so deeper, multiband followup will be required. The bad news is that the effort required to carry out deep, multiband surveys in the optical and near-IR is comparable to the effort required to execute the initial SZE or X-ray survey. The good news is that there are several projects being designed independently of planned and proposed SZE and X-ray surveys that will provide the required data. These survey projects include PRIME—a NASA Small Explorer Class proposal in Phase A study that will survey one quarter of the sky in the $1\text{-}3\mu\text{m}$ range, VISTA—a 4 m class telescope with a wide field near-IR and (eventually) optical camera that will carry out surveys in the southern hemisphere, the VST—an SDSS-like survey telescope with a large optical camera operating in the southern hemisphere, and the Large Synoptic Survey Telescope—a 6–8 m class telescope with wide field of view to carry out frequent, repeated imaging of large portions of the sky. In addition, there are extremely useful large field of view CCD

cameras available at KPNO/CTIO and on the CFHT. With these projects and others together with the exciting science possible with cluster surveys, it is only a question of time until very large cluster catalogs can be derived from large solid angle, multifrequency surveys.

5. Discussion

This contribution contains a description of two ways of using cluster surveys to learn about structure formation and cosmology: (1) the context free test of hierarchical structure formation using SZE cluster surveys, which are sufficiently sensitive to detect low mass clusters no matter what their redshift, and (2) the use of cluster redshift distributions within the context of our standard model for structure formation to determine the quantity and nature of dark matter and dark energy in the universe. It's important to emphasize that there is additional information that comes with a cluster survey. This information allows one to study the cluster mass function dn/dM as a function of redshift, likely improving the constraints derived from integrals over the mass function (i.e. equation 3). In addition, surveys (perhaps with some targeted followup) enable one to study cluster scaling relations such as the X-ray, optical or SZE luminosity–temperature or luminosity–mass relations; a combined study of scaling relations and the redshift distribution may well allow one to solve for the scaling relation evolution and cosmological parameters simultaneously (Diego et al 2001).

One can also study the spatial correlations among clusters to infer properties of the underlying power spectrum of dark matter density fluctuations. With good halo mass estimates like those required to use the cluster redshift distribution to full effect, it should be possible to use the cluster power spectrum constraints to improve limits on the neutrino mass density. Even in the absence of accurate halo mass estimates, it should be possible to use large surveys (in volume and number) to measure the scale of the break in the transfer function for the evolution of density perturbations. Recently, Cooray et al (2001) have emphasized that the physical scale of the break in the transfer function, which is the horizon scale at matter–radiation equality, depends on the matter density and CMB temperature. The matter density is measured to high precision with CMB anisotropy observations such as those with MAP and Planck. Therefore, the break in the transfer function is a standard rod, whose scale is independent of redshift and is calibrated to high accuracy with CMB data. Thus, measurements of the cluster correlation function within redshift shells returns the angular diameter distance as a function of redshift, much like the SNe Ia but with a strong physical basis for the lack of evolution in the standard rod. This approach is very complementary to the cluster redshift distribution approach, and it hinges less on extracting unbiased estimates of cluster masses from cluster observables like the X-ray or SZE luminosity.

Acknowledgments. I would like to acknowledge Zoltan Haiman for the many cluster survey discussions and calculations we have shared. It is also a pleasure to acknowledge many cluster related conversations with John Carlstrom and Gil Holder.

References

- Arnaud, M. & Evrard, A.E. 1999, MNRAS, 305, 631
 Bahcall, N.A., Fan, X. & Cen, R. 1997, ApJ, 485, L53
 Bialek, J.J., Evrard, A.E. & Mohr, J.J. 2001, ApJ, 555, 597
 Bryan, G.L. & Norman, M.L. 1998, ApJ, 495, 80
 Burles, S. & Tytler, D. 1998, ApJ, 507, 732
 Cooray, A., Hu, W., Huterer, D. & Joffre, M. 2001, ApJ, 557, L7
 Diego, J.M., Martinez-Gonzalez, E., Sanz, J.L., Cayon, L. & Silk, J. 2001, MNRAS, 325, 1533
 Evrard, A.E., Metzler, C.A. & Navarro, J.F. 1996, ApJ, 469, 494
 Haiman, Z., Mohr, J.J. & Holder, G.P. 2001, ApJ, 553, 545
 Halverson, N.W. et al 2001, ApJ, submitted (astro-ph/0104489)
 Hanany, S. et al. 2000, ApJ, 545, L5
 Holder, G.P., Mohr, J.J., Carlstrom, J.E., Evrard, A.E. & Leitch, E.M. 2000, ApJ, 544, 629
 Horner, D.J., Mushotzky, R.F. & Scharf, C.A. 1999, ApJ, 520, 78
 Jaffe, A. et al. 2001, Phys.Rev.Lett, 86, 3475
 Jenkins, A., Frenk, C.S., White, S.D.M., Colberg, J.M., Cole, S., Evrard, A.E., Couchman, H.M.P. & Yoshida, N. 2001, MNRAS, 321, 372
 Lange, A. et al. 2001, Phys.Rev.D, 63, 2001
 Mohr, J.J. & Evrard, A.E. 1997, ApJ, 491, 38
 Mohr, J.J., Mathiesen, B. & Evrard, A.E. 1999, ApJ, 517, 627
 Netterfield, C.B. et al. 2001, ApJ, submitted (astro-ph/0104460)
 Peebles, P.J.E. 1993, *Principles of Physical Cosmology*, Princeton University Press; Princeton
 Perlmutter, S. et al. 1999, ApJ, 517, 565
 Press, W.H. & Schechter, P. 1974, ApJ, 187, 425
 Pryke, C. et al. 2001, ApJ, submitted (astro-ph/0104490)
 Schmidt, B.P. et al. 1998, ApJ, 507, 46
 Sunyaev, R.A. & Zel'dovich, Y.B. 1972, Comments Astrophys. Space Phys., 4, 173
 Verde, L., Haiman, Z., Spergel, D.N. 2001, ApJ, submitted (astro-ph/0106315)
 Viana, P.T.P. & Liddle, A.R. 1999, MNRAS, 303, 535
 White, D.A. & Fabian, A.C. 1995, MNRAS, 273, 72
 White, M., 2001, A&A, 367, 27
 White, S.D.M., Navarro, J.F., Evrard, A.E. & Frenk, C.S. 1993a, Nature, 366, 429
 White, S.D.M., Efstathiou, G. & Frenk, C.S. 1993b, MNRAS, 262, 1023

Simulating the formation of cosmic structure

BY C. S. FRENK

Physics Department, University of Durham, Durham DH1 3LE, England

A timely combination of new theoretical ideas and observational discoveries has brought about significant advances in our understanding of cosmic evolution. Computer simulations have played a key role in these developments by providing the means to interpret astronomical data in the context of physical and cosmological theory. In the current paradigm, our Universe has a flat geometry, is undergoing accelerated expansion and is gravitationally dominated by elementary particles that make up cold dark matter. Within this framework, it is possible to simulate in a computer the emergence of galaxies and other structures from small quantum fluctuations imprinted during an epoch of inflationary expansion shortly after the Big Bang. The simulations must take into account the evolution of the dark matter as well as the gaseous processes involved in the formation of stars and other visible components. Although many unresolved questions remain, a coherent picture for the formation of cosmic structure is now beginning to emerge.

Keywords: Cosmic structure, dark matter, gas dynamics, galaxy formation, computer simulations

1. Introduction

The origin of structure in the Universe is a central problem in Physics. Its solution will not only inform our understanding of the processes by which matter became organized into galaxies and clusters, but it will also help uncover the identity of the dark matter, offer insights into events that happened in the early stages of the Big Bang and provide a useful check on the values of the fundamental cosmological parameters estimated by other means.

Because of its non-linear character, lack of symmetry and general complexity, the formation of cosmic structure is best approached theoretically using numerical simulations. The problem is well posed because the initial conditions – small perturbations in the density and velocity field of matter – are, in principle, known from Big Bang theory and observations of the early Universe, while the basic physical principles involved are understood. The behaviour of the dark matter is governed primarily by gravity, while the formation of the visible parts of galaxies involves gas dynamics and radiative processes of various kinds. Using cosmological simulations it is possible to follow the development of structure from primordial perturbations to the point where the model can be compared with observations.

Over the past few years, there has been huge progress in quantifying observationally the properties of galaxies not only in the nearby universe, but also in the very distant universe. Since the clustering pattern of galaxies is rich with information

about physics and cosmology, much effort is invested in mapping the distribution of galaxies at different epochs. Two large ongoing surveys, the US-based Sloan Digital Sky Survey (York *et al.* 2000), and the Anglo-Australian “2-degree field galaxy redshift survey” (2dFGRS, Colless *et al.* 2001), are revolutionizing our view of the nearby universe with order of magnitude increases in the amount of available data. Similarly, new data collected in the past five years or so have, for the first time, opened up the high redshift universe† to detailed statistical study (Steidel *et al.* 1996).

The advent of large computers, particularly parallel supercomputers, together with the development of efficient algorithms, has enabled the accuracy and realism of simulations to keep pace with observational progress. With the wealth of data now available, simulations are essential to interpret astronomical data and to link them to physical and cosmological theory.

2. Building a model

To build a model of large-scale structure, four key ingredients need to be specified: (i) the content of Universe, (ii) the initial conditions, (iii) the growth mechanism, and (iv) the values of fundamental cosmological parameters. I now discuss each of these in turn.

(a) The content of the Universe

Densities are usually expressed in terms of the cosmological density parameter, $\Omega = \rho/\rho_{crit}$, where the critical density, ρ_{crit} , is the value that makes the geometry of the Universe flat. The main constituents of the Universe and their contribution to Ω are listed in Table 1.

Table 1. *The content of the Universe*

Component	Contribution to Ω
CMB radiation	$\Omega_r = 4.7 \times 10^{-5}$
massless neutrinos	$\Omega_\nu = 3 \times 10^{-5}$
massive neutrinos	$\Omega_\nu = 6 \times 10^{-2} \left(\frac{m_\nu}{1\text{eV}} \right)$
baryons	$\Omega_b = 0.037 \pm 0.009$
(of which stars)	$\Omega_s = (0.0023 - 0.0041) \pm 0.0004$
dark matter	$\Omega_{dm} \simeq 0.3$
dark energy	$\Omega_\Lambda \simeq 0.7$

The main contribution to the extragalactic radiation field today is the cosmic microwave background (CMB), the redshifted radiation left over from the Big Bang. These photons have been propagating freely since the epoch of “recombination”, approximately 300,000 years after the Big Bang. The CMB provides a direct observational window to the conditions that prevailed in the early Universe. The Big

† In cosmology, distances to galaxies are estimated from the redshift of their spectral lines; higher redshifts correspond to more distant galaxies and thus to earlier epochs.

Bang also produced neutrinos which today have an abundance comparable to that of photons. We do not yet know for certain what, if any, is the mass of the neutrino, but even for the largest masses that seem plausible at present, $\sim 0.1\text{eV}$, neutrinos make a negligible contribution to the total mass budget (although they could be as important as baryons). The abundance of baryons is now known with reasonable precision from comparing the abundance of deuterium predicted by Big Bang theory with observations of the absorption lines produced by intergalactic gas clouds at high redshift seen along the line-of-sight to quasars (Tytler *et al.* 2000). Baryons, the overwhelming majority of which are *not* in stars today, are also dynamically unimportant (except, perhaps, in the cores of galaxies).

Dark matter makes up most of the matter content of the Universe today. To the now firm dynamical evidence for its existence in galaxy halos, even more direct evidence has been added by the phenomenon of gravitational lensing which has now been detected around galaxy halos (e.g. Fischer *et al.* 2000, McKay *et al.* 2001, Wilson *et al.* 2001), in galaxy clusters (e.g. Clowe *et al.* 2000), and in the general mass field (e.g. Van Waerbeke *et al.* 2001 and references therein). The distribution of dark matter in rich clusters can be reconstructed in fair detail from the weak lensing of distant background galaxies in what amounts virtually to imagining the cluster dark matter. Various dynamical tests are converging on a value of $\Omega_{\text{dm}} \simeq 0.3$, which is also consistent with independent determinations such as those based on the baryon fraction in clusters (White *et al.* 1993, Evrard 1997), and on the evolution in the abundance of galaxy clusters (Eke *et al.* 1998, Borgani *et al.* 2001). Since Ω_{dm} is much larger than Ω_b , it follows that the dark matter cannot be made of baryons. The most popular candidate for the dark matter is a hypothetical elementary particle like those predicted by supersymmetric theories of particle physics. These particles are referred to generically as cold dark matter or CDM. (Hot dark matter is also possible, for example, if the neutrino had a mass of $\sim 5\text{ eV}$. However, early cosmological simulations showed that the galaxy distribution in a universe dominated by hot dark matter would not resemble that observed in our Universe (White, Frenk and Davis 1983).)

A recent addition to the cosmic budget is the dark energy, direct evidence for which was first provided by studies of type Ia supernovae (Riess *et al.* 1998, Perlmutter *et al.* 1999)[†]. These presumed ‘standard candles’ can now be observed at redshifts between 0.5 and 1 and beyond. The more distant ones are fainter than would be expected if the universal expansion were decelerating today, indicating that the expansion is, in fact, accelerating. Within the standard Friedmann cosmology, there is only one agent that can produce an accelerating expansion. This is nowadays known as dark energy, a generalization of the cosmological constant first introduced by Einstein, which could, in principle, vary with time. The supernova evidence is consistent with the value $\Omega_\Lambda \simeq 0.7$. Further, independent evidence for dark energy is provided by a recent joint analysis of CMB data (see next section) and the 2dFGRS (Efstathiou *et al.* 2002).

Amazingly, when all the components are added together, the data are consistent with a flat universe:

[†] The possibility that dark energy might be the dynamically dominant component had been anticipated by theorists from studies of the cosmic large-scale structure (see e.g. Efstathiou *et al.* 1990), and was considered in the first simulations of structure formation in cold dark matter universes (Davis *et al.* 1985).

$$\Omega = \Omega_b + \Omega_{dm} + \Omega_\Lambda \simeq 1 \quad (2.1)$$

(b) *The initial conditions*

The idea that galaxies and other cosmic structures are the result of the slow amplification by the force of gravity of small primordial perturbations present in the mass density at early times goes back, at least, to the 1940s (Lifshitz 1946). However, it was only in the early 1980s that a physical mechanism capable of producing small perturbations was identified. This is the mechanism of inflation, an idea due to Guth (1981), which changed the face of modern cosmology. Inflation is produced by the dominant presence of a quantum scalar field which rolls slowly from a false to the true vacuum, maintaining its energy density approximately constant and causing the early Universe to expand exponentially for a brief period of time. Quantum fluctuations in the inflaton field are blown up to macroscopic scales and become established as genuine adiabatic ripples in the energy density. Simple models of inflation predict the general properties of the resulting fluctuation field: it has Gaussian distributed amplitudes and a near scale-invariant power spectrum (Starobinskii 1982).

After three decades of ever more sensitive searches, evidence for the presence of small fluctuations in the early universe was finally obtained in 1992. Since prior to recombination the matter and radiation fields were coupled, fluctuations in the mass density are reflected in the temperature of the radiation. Temperature fluctuations in the CMB were discovered by the COBE satellite (Smoot *et al.* 1992) and are now being measured with ever increasing accuracy, particularly by detectors deployed in long-flight balloons (de Bernardis *et al.* 2000, Hanany *et al.* 2000, Leitch *et al.* 2002). The spectrum of temperature fluctuations is just what inflation predicts: it is scale invariant on large scales and shows a series of “Doppler” or “acoustic” peaks which are the result of coherent acoustic oscillations experienced by the photon-baryon fluid before recombination. The characteristics of these peaks depend on the values of the cosmological parameters. For example, the location of the first peak is primarily determined by the large-scale geometry of the Universe and thus by the value of Ω . Current data imply a flat geometry, consistent with eqn. 2.1.

The spectrum of primordial fluctuations generated, for example, by inflation evolves with time in a manner that depends on the content of the Universe and the values of the cosmological parameters. The dark matter acts as a sort of filter, inhibiting the growth of certain wavelengths and promoting the growth of others. Following the classical work of Bardeen *et al.* (1986), transfer functions for different kinds of dark matter (and different types of primordial fluctuation fields, including non-Gaussian cases) have been computed. In Gaussian models, the product (in Fourier space) of the primordial spectrum and the transfer function, together with the growing mode of the associated velocity field, provides the initial conditions for the formation of cosmic structure.

(c) *Growth mechanism*

Primordial fluctuations grow by gravitational instability: overdense fluctuations expand linearly, at a retarded rate relative to the Universe as a whole, until even-

tually they reach a maximum size and collapse non-linearly to form an equilibrium (or ‘virialized’) object whose radius is approximately half the physical size of the perturbation at maximum expansion. The theory of fluctuation growth is lucidly explained by Peebles (1980).

Although gravitational instability is now widely accepted as the primary growth mechanism responsible for the formation of structure, it is only very recently that firm empirical evidence for this process was found. Gravitational instability causes inflow of material around overdense regions. From the perspective of a distant observer, this flow gives rise to a characteristic infall pattern which is, in principle, measurable in a galaxy redshift survey by comparing the two-point galaxy correlation function along and perpendicular to the line-of-sight. In this space, the infall pattern resembles a butterfly (Kaiser 1987). This pattern has been clearly seen for the first time in the 2dFGRS (Peacock *et al.* 2001)†.

(d) Cosmological parameters

After decades of debate, the values of the fundamental cosmological parameters are finally being measured with some degree of precision. The main reason for this is the accurate measurement of the acoustic peaks in the CMB temperature anisotropy spectrum whose location, height and shape depend on the values of the cosmological parameters. Some parameter degeneracies exist but some of these can be broken using other data, for example, the distant Type Ia supernovae or the 2dFGRS (eg. Efstathiou *et al.* 2002). The CMB data alone do not constrain the Hubble constant, but there is a growing consensus from the HST key project (Freedman *et al.* 2001), and other methods, that its value, in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ is $h = 0.70 \pm 0.07$. In addition to h and the other parameters listed in Table 1, the other important number in studies of large-scale structure is the amplitude of primordial density fluctuations which is usually parametrized by the quantity σ_8 (the linearly extrapolated value of the top-hat filtered fluctuation amplitude on the fiducial scale of $8 h^{-1} \text{ Mpc}$). The best estimate of this quantity comes from the observed abundance of rich galaxy clusters which gives $\sigma_8 \Omega^{0.6} = 0.5$, with an uncertainty of about 10% (Eke, Cole & Frenk 1996, Viana & Liddle 1996, Pierpaoli *et al.* 2001).

3. Cosmological simulations

Operationally, the problem of the cosmic large-scale structure can be divided into two parts: understanding the clustering evolution of the dark matter and understanding the gaseous and radiative processes that lead to the formation of galaxies. Specialized simulation techniques have been developed to tackle both aspects of the problem. The evolution of the dark matter is most often calculated using N-body techniques, implemented through a variety of efficient algorithms, such as P³M (Particle-particle/particle-mesh; Efstathiou *et al.* 1985), AP³M (the adaptive mesh version of P³M; Couchman *et al.* 1995) and hierarchical trees (Barnes & Hut 1986, Springel *et al.* 2001, Stadel 2000). Gaseous and radiative processes

† Strictly speaking the ‘butterfly’ pattern does not prove the existence of infall since the continuity equation would ensure a similar pattern even if velocities were induced by non-gravitational processes. However, it can be shown that such velocities, if present, would rapidly decay.

are followed by combining a hydrodynamics code with an N-body code. Numerical hydrodynamic techniques used in cosmology include Eulerian methods (Cen 1992), Lagrangian codes based on Smooth Particle Hydrodynamics (SPH) (Gingold & Monaghan 1977), and hybrid codes (e.g. Gnedin 1995, Pen 1998). These techniques have different strengths and weaknesses, but they all give similar results in the simplest cosmological problems where a detailed comparison has been performed (Frenk *et al.* 1999).

There has been a rapid growth in the size and power of cosmological simulations in the two and a half decades since this technique was introduced into the subject by Peebles (1970). One way to measure this growth is by the number of particles employed in the simulations. The size of the largest simulations has grown exponentially, in a manner reminiscent of the well-known “Moore’s law” that describes the increase in cpu speed with time, except that the advent of massively parallel supercomputers led to a sudden order-of-magnitude jump in size towards the end of the past decade. The largest simulations carried out to date are the 1-billion particle “Hubble volume,” N-body simulations performed by the Virgo consortium, an international collaboration of researchers in the UK, Germany and Canada.

(a) *Large-scale structure*

Figure 1 illustrates the spatial distribution of *dark matter* at the present day, in a series of simulations covering a large range of scales. Each panel is a thin slice of the cubical simulation volume and shows the slightly smoothed density field defined by the dark matter particles. In all cases, the simulations pertain to the “ Λ CDM” cosmology, a flat cold dark matter model in which $\Omega_{\text{dm}} = 0.3$, $\Omega_{\Lambda} = 0.7$ and $h = 0.7$. The top-left panel illustrates the Hubble volume simulation: on these large scales, the distribution is very smooth. To reveal more interesting structure, the top right panel displays the dark matter distribution in a slice from a volume approximately 2000 times smaller. At this resolution, the characteristic filamentary appearance of the dark matter distribution is clearly visible. In the bottom-right panel, we zoom again, this time by a factor of 5.7 in volume. We can now see individual galactic-size halos which preferentially occur along the filaments, at the intersection of which large halos form that will host galaxy clusters. Finally, the bottom-left panel zooms into an individual galactic-size halo. This shows a large number of small substructures that survive the collapse of the halo and make up about 10% of the total mass (Klypin *et al.* 1999, Moore *et al.* 1999)

For simulations like the ones illustrated in Figure 1, it is possible to characterize the statistical properties of the dark matter distribution with very high accuracy. For example, Figure 2 shows the 2-point correlation function, $\xi(r)$, of the dark matter (a measure of its clustering strength) in the simulation depicted in the top-right of Figure 1 (Jenkins *et al.* 1998). The statistical error bars in this estimate are actually smaller than the thickness of the line. Similarly, higher order clustering statistics, topological measures, the mass function and clustering of dark matter halos and the time evolution of these quantities can all be determined very precisely from these simulations (e.g. Jenkins *et al.* 2001, Evrard *et al.* 2002). In a sense, the

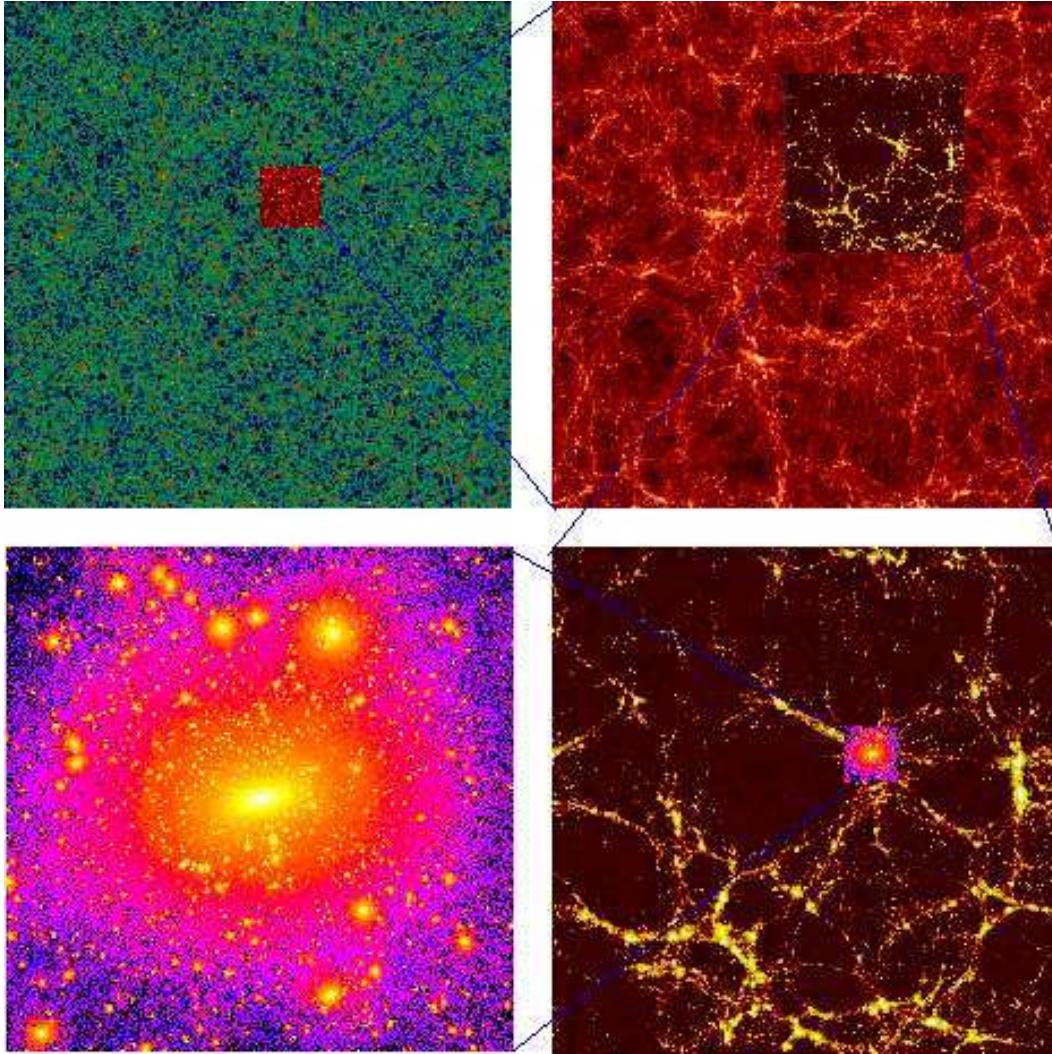


Figure 1. Slices through 4 different simulations of the dark matter in the ‘ Λ CDM’ cosmology. Denoting the number of particles in each simulation by N , the length of the simulation cube by L , the thickness of the slice by t , and the particle mass by m_p , the characteristics of each panel are as follows. Top-left (the Hubble volume simulation, Evrard *et al.* 2002): $N = 10^9$, $L = 3000 h^{-1} \text{Mpc}$, $t = 30 h^{-1} \text{Mpc}$, $m_p = 2.2 \times 10^{12} h^{-1} M_\odot$. Top-right (Jenkins *et al.* 1998): $N = 16.8 \times 10^6$, $L = 250 h^{-1} \text{Mpc}$, $t = 25 h^{-1} \text{Mpc}$, $m_p = 6.9 \times 10^{10} h^{-1} M_\odot$. Bottom-right (Jenkins *et al.* 1998): $N = 16.8 \times 10^6$, $L = 140 h^{-1} \text{Mpc}$, $t = 14 h^{-1} \text{Mpc}$, $m_p = 1.4 \times 10^{10} h^{-1} M_\odot$. Bottom-left (Navarro *et al.* 2002): $N = 7 \times 10^6$, $L = 0.5 h^{-1} \text{Mpc}$, $t = 1 h^{-1} \text{Mpc}$, $m_p = 6.5 \times 10^5 h^{-1} M_\odot$.

problem of the distribution of dark matter in the Λ CDM model can be regarded as largely solved†.

† However, the innermost structure of halos like those in the bottom-left of Figure 1 is still a matter of controversy.

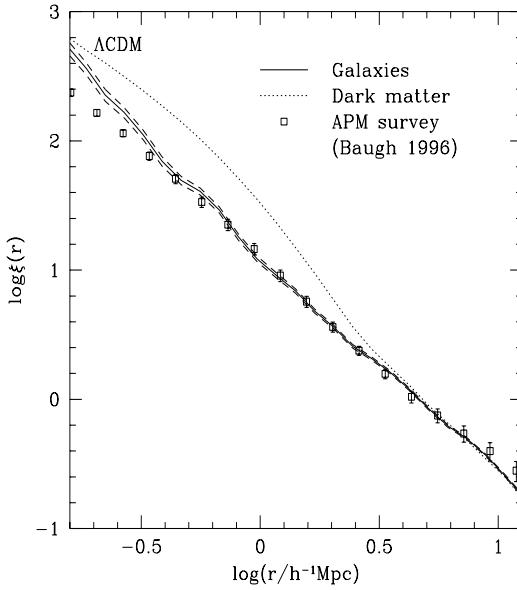


Figure 2. Two-point correlation functions. The dotted line shows the dark matter $\xi_{dm}(r)$ (Jenkins *et al.* 1998). The solid line shows the galaxy predictions of Benson *et al.* (2000), with Poisson errors indicated by the dashed lines. The points with errorbars show the observed galaxy $\xi_{gal}(r)$ (Baugh 1996). The galaxy data are discussed in §3(b). (Adapted from Benson *et al.* (2001a).

In contrast to the clustering of the dark matter, the process of galaxy formation is still poorly understood. How then can dark matter simulations like those of Figure 1 be compared with observational data which, for the most part, refer to galaxies? On large scales a very important simplification applies: for Gaussian theories like CDM, it can be shown that if galaxy formation is a local process, that is, if it depends only upon local physical conditions (density, temperature, etc), then, on scales much larger than that associated with individual galaxies, the galaxies must trace the mass, i.e. on sufficiently large scales, $\xi_{gal}(r) \propto \xi_{dm}(r)$ (Coles 1993). It suffices therefore to identify a random subset of the dark matter particles in the simulation to obtain an accurate prediction for the properties of galaxy clustering on large scales. This idea (complemented on small scales by an empirical prescription in the manner described by Cole *et al.* 1998) has been used to construct the mock versions of a region of the APM galaxy survey and of a slice of the 2dFGRS displayed in Figures 3 and 4 which also show the real data for comparison in each case. By eye at least, it is very difficult to distinguish the mocks from the real data.

A quantitative comparison between simulations and the real world is carried out in Figure 5. The symbols show the estimate of the power spectrum in the 2dFGRS survey (Percival *et al.* 2001). This is the raw power spectrum convolved with the survey window function and can be compared directly with the line showing the theoretical prediction obtained from the mock catalogues which have exactly the same window function. The agreement between the data and the ΛCDM model is remarkably good.

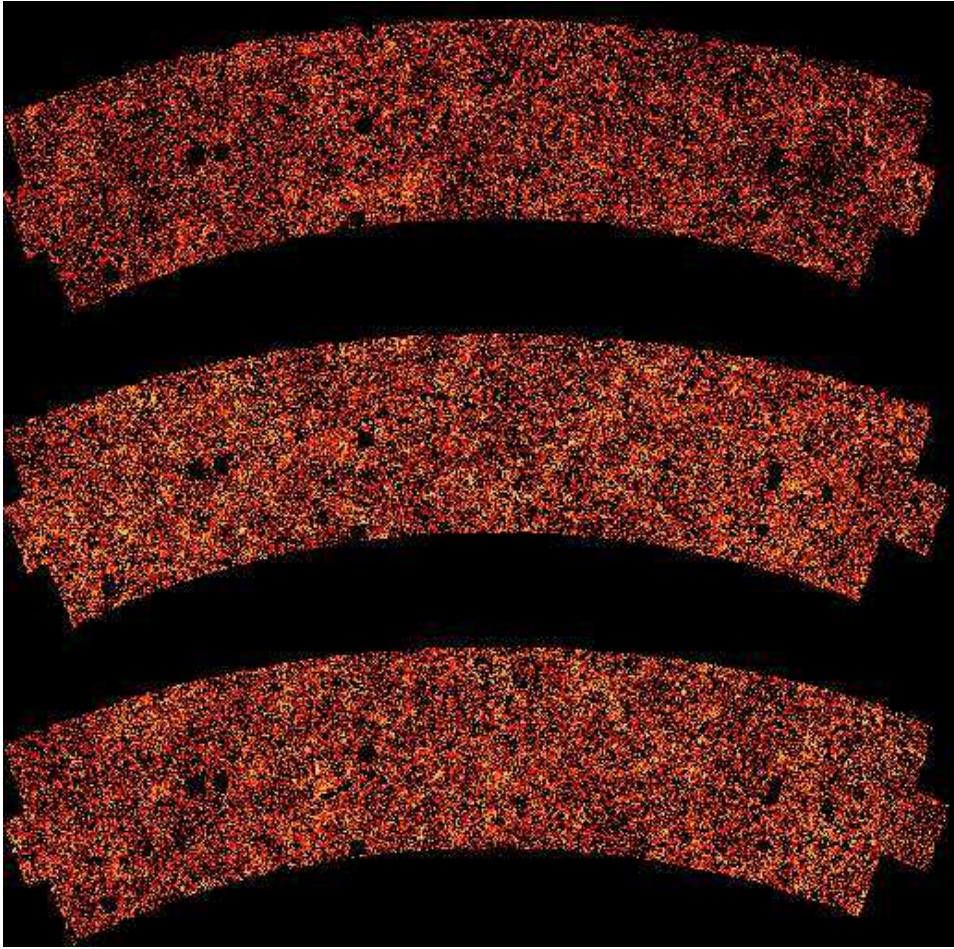


Figure 3. The region of the APM projected galaxy survey from which the 2dFGRS is drawn. Only galaxies brighter than $m_{b_J} = 19.35$ are plotted. The top panel is the real data and the other two panels are mock catalogues constructed from the Hubble volume simulations.

(b) Galaxy formation

Understanding galaxy formation is a much more difficult problem than understanding the evolution of the dark matter distribution. In the CDM theory, galaxies form when gas, initially well mixed with the dark matter, cools and condenses into emerging dark matter halos. In addition to gravity, a non-exhaustive list of the processes that now need to be taken into account includes: the shock heating and cooling of gas into dark halos, the formation of stars from cold gas and the evolution of the resulting stellar population, the feedback processes generated by the ejection of mass and energy from evolving stars, the production and mixing of heavy elements, the extinction and reradiation of stellar light by dust particles, the formation of black holes at the centres of galaxies and the influence of the associated quasar emission. These processes span an enormous range of length and mass scales. For

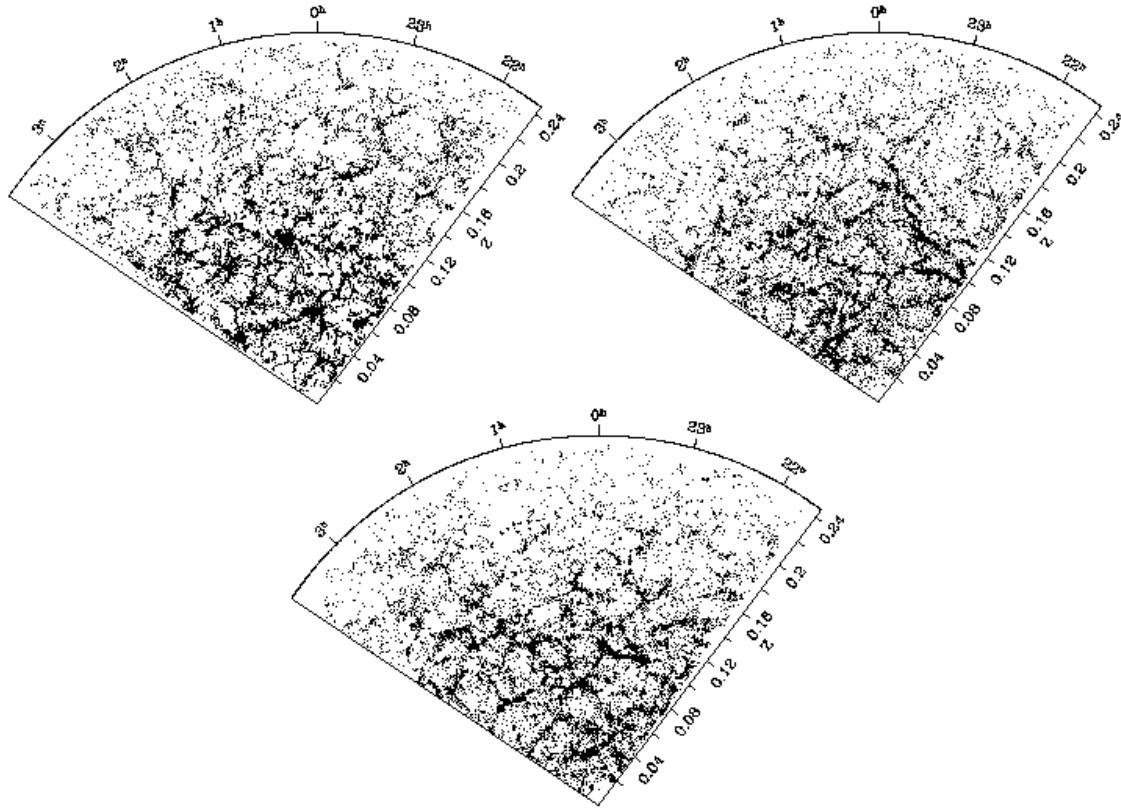


Figure 4. A 1° thick slice through the 2dF galaxy redshift survey. The radial coordinate is redshift and the angular coordinate is right ascension. The top-left panel is the real data and the other two panels are mock catalogues constructed from the Hubble volume simulations.

example, the parsec scale relevant to star formation is a factor of 10^8 smaller than the scale of a galaxy supercluster.

The best that can be done with current computing techniques is to model the evolution of dark matter and gas in a cosmological volume with resolution comparable to a single galaxy. Subgalactic scales must then be regarded as “sub-grid” scales and followed by means of phenomenological models based either on our current physical understanding or on observations. In the approach known as “semi-analytic” modelling (White & Frenk 1991), even the gas dynamics is treated phenomenologically using a simple, spherically symmetric model to describe the accretion and cooling of gas into dark matter halos. It turns out that this simple model works surprisingly well as judged by the good agreement with results of full N-body/gas-dynamical simulations (Benson *et al.* 2001b, Helly *et al.* 2002, Yoshida *et al.* 2002).

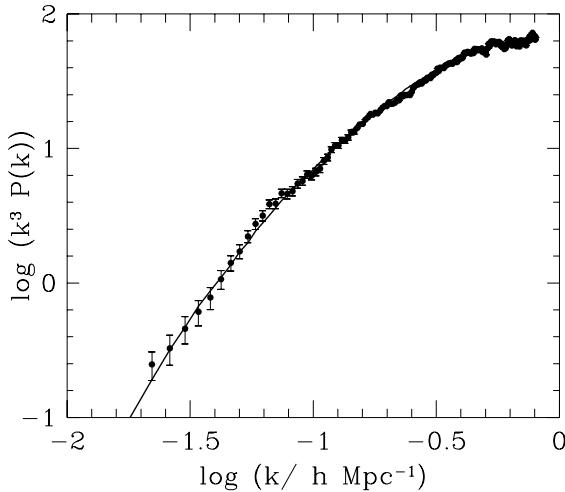


Figure 5. The power spectrum of the 2dFGRS (symbols) compared with the power spectrum predicted in the Λ CDM model (line). Both power spectra are convolved with the 2dFGRS window function. The model predictions come from dark matter simulations and assume that, on large scales, the distribution of galaxies traces the distribution of mass. (Adapted from Percival *et al.* 2001).

The main difficulty encountered in cosmological gas dynamical simulations arises from the need to suppress a cooling instability present in hierarchical clustering models like CDM. The building blocks of galaxies are small clumps that condense at early times. The gas that cools within them has very high density, reflecting the mean density of the Universe at that epoch. Since the cooling rate is proportional to the square of the gas density, in the absence of heat sources, most of the gas would cool in the highest levels of the mass hierarchy leaving no gas to power star formation today or even to provide the hot, X-ray emitting plasma detected in galaxy clusters. Known heat sources are photoionisation by early generations of stars and quasars and the injection of energy from supernovae and active galactic nuclei. These processes, which undoubtedly happened in our Universe, belong to the realm of subgrid physics which cosmological simulations cannot resolve. Different treatments of this “feedback” result in different amounts of cool gas and can lead to very different predictions for the properties of the galaxy population. This is a fundamental problem that afflicts cosmological simulations even when they are complemented by the inclusion of semi-analytic techniques. In this case, the resolution of the calculation can be extended to arbitrarily small mass halos, perhaps allowing a more realistic treatment of feedback. Although they are less general than full gasdynamical simulations, simulations in which the evolution of gas is treated semi-analytically make experimentation with different prescriptions relatively simple and efficient (Kauffmann White & Guiderdoni 1993, Somerville & Primack 1999, Cole *et al.* 2000)

The outcome of an N-body dark matter simulation in a Λ CDM universe in which the visible properties of the galaxies have been calculated using the semi-analytic model of Cole *et al.* (2000) is illustrated in Fig. 6 (Benson *et al.* 2001a). Galaxies

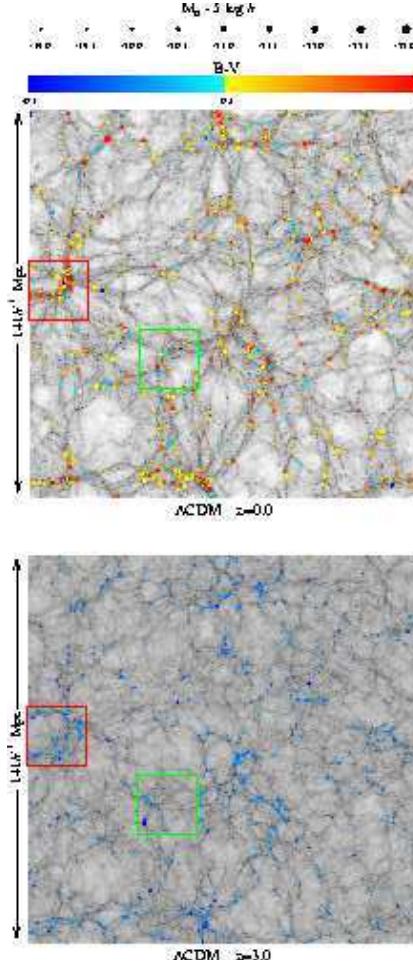


Figure 6. A slice $10 h^{-1}$ Mpc thick of a simulation of a cubic region of side $141 h^{-1}$ Mpc in the ΛCDM cosmology. The grey scale shows a slightly smoothed representation of the dark matter in the N-body simulation. The coloured dots show galaxies; the size of the dots is proportional to the B-band luminosity of the galaxy and the colour represents the B-V colour as given on the scale on the top. The top panel corresponds to redshift $z = 0$ and the bottom panel to $z = 3$. (Adapted from Benson *et al.* 2001a).

form mostly along the filaments delineated by the dark matter. Red galaxies predominate in the most massive dark matter halos, just as observed in real galaxy clusters. This segregation is a natural outcome of hierarchical clustering from CDM initial conditions. It reflects the fact that the progenitors of rich clusters form substantially earlier than a typical dark matter halo of the same mass. Fig. 7 shows the galaxy luminosity function which describes the abundance of galaxies of different luminosities. The theoretical predictions, shown by the line, agree remarkably well with the observations but this should not be regarded as a spectacular success of the theory because the free parameters in the semi-analytic star formation and feedback model have been tuned to achieve as good a match as possible to this

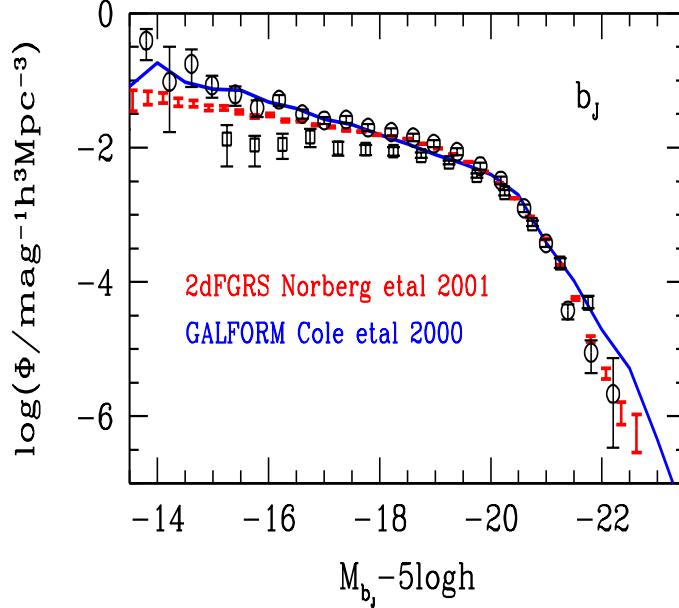


Figure 7. The galaxy luminosity function. The symbols show the number of galaxies per unit volume and per unit magnitude measured in various surveys, as a function of galaxy magnitude (open circles: Zucca *et al.* 1997; open squares: Loveday *et al.* 1992; thick error bars: Norberg *et al.* 2001b). The solid line shows the predictions of the semi-analytic model of Cole *et al.* (2000).

specific observational dataset. In particular, the feedback model has been tuned to produce a relatively flat function at the faint end.

Having fixed the model parameters by reference to a small subset of the data such as the galaxy luminosity function, we can ask whether the same model accounts for other basic observational data. The galaxy autocorrelation function, $\xi_{gal}(r)$, in the simulations is plotted in Fig. 2 above. On large scales, it follows $\xi_{dm}(r)$ quite closely, but on small scales it dips below the mass autocorrelation function. This small scale ‘‘antibias’’ has also been seen in N-body/gasdynamical simulations of the Λ CDM cosmology (Pearce *et al.* 1999, 2001, Dave *et al.* 1999), and in dark matter simulations that resolve individual galactic halos (Klypin *et al.* 1999). The galaxy autocorrelation function in the simulations of Benson *et al.* (2000) agrees remarkably well with the observational data (see also Kauffmann *et al.* 1999a). This is a genuine success of the theory because no model parameters have been adjusted in this comparison. The differences between the small-scale clustering of galaxies and dark matter result from the interplay between the clustering of dark matter halos and the occupation statistics of galaxies in halos which, in turn, are determined by the physics of galaxy formation. This conclusion, discussed in detail by Benson *et al.* (2000), has led to the development of an analytic formulation known as the ‘‘halo model’’ (e.g. Seljak 2000, Peacock & Smith 2000, Berlind & Weinberg 2002).

Another genuine prediction of the model is the dependence of the strength of clustering on the luminosity of different subsamples. It can be seen in Fig. 6 that

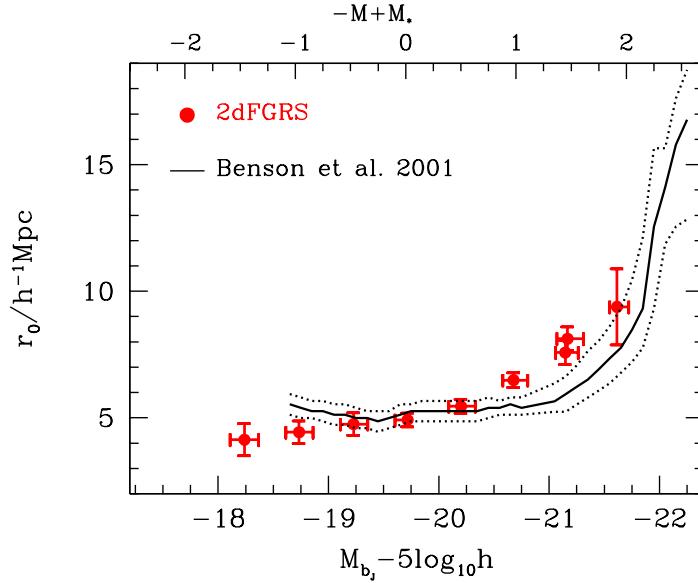


Figure 8. The correlation length as a function of the luminosity of different galaxy subsamples. The correlation length is defined as the pair separation for which $\xi(r) = 1$. The symbols show the results from the 2dFGRS and the line the predictions of the simulations of Benson *et al.* (2000). (Adapted from Norberg *et al.* 2001a).

the brightest galaxies are concentrated in the most massive clusters, leading one to suspect that their autocorrelation function must be stronger than average. This is indeed the case, as illustrated in Fig. 8 which compares the variation of the clustering length (defined as the pair separation for which $\xi(r) = 1$) of galaxy samples of different intrinsic luminosity in the simulations of Benson *et al.* (2001a) with the observational data obtained from the 2dFGRS by Norberg *et al.* (2001a). The agreement between theory and observations is remarkable considering that there are no adjustable parameters in this comparison. The reason for the strong clustering of bright galaxies is related to the colour-density relation seen in Fig. 6: the brightest galaxies form in the highest peaks of the density distribution which, in initially Gaussian fields, are more strongly clustered than average peaks which produce less extreme galaxies.

The patch of model universe illustrated in the top panel of Fig. 6 is shown at the earlier epoch corresponding to redshift $z = 3$ (when the universe was only about 20% of its current age) in the bottom panel of this figure. The galaxies are now blue, reflecting the colour of their younger stellar population. There are fewer galaxies in this plot than in the $z = 0$ slice. In fact, this is the epoch when the first substantial population of bright galaxies formed in the simulation. As Baugh *et al.* (1998) argued, the properties of these model galaxies resemble those of the ‘Lyman-break’ galaxies discovered by Steidel *et al.* (1996), even though different models make somewhat different predictions for their exact properties (Somerville *et al.* 2001). Most models, however, predict that the brightest galaxies at $z = 3$ should be strongly clustered (Kauffmann *et al.* 1999b) and, indeed, the models of Baugh *et al.* (1998) correctly anticipated that the Lyman-break galaxies would

have a clustering length comparable to that of bright galaxies today (Adelberger *et al.* 1998). This too should be regarded as a significant success of this kind of modelling in the Λ CDM cosmology. As Fig. 6 shows, in contrast to the galaxies, the dark matter is much more weakly clustered at $z = 3$ than at $z = 0$, indicating that galaxies were strongly biased at birth.

4. Conclusions

Unlike most computational problems in many areas of science, the cosmological problem is blessed with known, well-specified initial conditions. Within a general class of models, it is possible to calculate the properties of primordial perturbations in the cosmic energy density generated by quantum processes during an early inflationary epoch. In a wide family of inflationary models, these perturbations are adiabatic, scale-invariant and have Gaussian-distributed Fourier amplitudes. The model also requires an assumption about the nature of the dark matter and the possibilities have now been narrowed down to non-baryonic candidates of which cold dark matter particles seem the most promising. An empirical test of the initial conditions for the formation of structure predicted by the model is provided by the cosmic microwave background radiation. The tiny temperature fluctuations it exhibits have exactly the properties expected in the model. Furthermore, the CMB data can be used to fix some of the key model parameters such as Ω and Ω_b , while these data, combined with other recent datasets such as the 2dFGRS, allow the determination of many of the remaining parameters such as Ω_m , Ω_Λ and h . It is this specificity of the cosmological problem that has turned simulations into the primary tool for connecting cosmological theory to astronomical observations.

In addition to well-specified initial conditions, the cosmological dark matter problem has the advantage that the only physical interaction that is important is gravity. The problem can thus be posed as a gravitational N-body problem and approached using the many sophisticated techniques that have been developed over the past two decades to tackle this problem. Although on small scales there remain a number of unresolved issues, it is fair to say that on scales larger than a few megaparsecs, the distribution of dark matter in CDM models is essentially understood. The inner structure of dark matter halos, on the other hand, is still a matter of debate and the mass function of dark matter halos has only been reliably established by simulations down to masses of order $10^{11} M_\odot$. Resolving these outstanding issues is certainly within reach, but this will require carefully designed simulations and large amounts of computing power.

The frontier of the subject at present lies in simulations of the formation, evolution and structure of galaxies. This problem requires first of all a treatment of gas dynamics in a cosmological context and a number of techniques, relying on direct simulations or on semi-analytical approximations, are being explored. There are quite a few different approaches to cosmological gasdynamics, but it is reassuring that they all give similar results in the simplest relevant problem, the evolution of non-radiative gas during the formation of a galaxy cluster. No detailed comparisons exist yet for the more complicated case in which the gas is allowed to cool, but at least one of the gasdynamic simulation techniques, SPH, gives quite similar results to a simple semi-analytic approach. Realistic models of galaxy formation, however, will require much more than a correct treatment of cooling gas. Such mod-

els will necessarily have to include a plethora of astrophysical phenomena such as star formation, feedback, metal enrichment, etc. The huge disparity between the submegaparsec scales on which these processes operate and the gigaparsec scale of the large-scale structure makes it impossible to contemplate a comprehensive *ab initio* calculation. The way forward is clearly through a hybrid approach combining direct simulation of processes operating on a limited range of scales with a phenomenological treatment of the others. There is currently a great deal of activity in the phenomenology of galaxy formation.

In spite of the uncertainties that remain, all the indications are that our Universe is well described by a model in which

- (i) the overall geometry is flat;
- (ii) the dominant dynamical components are cold dark matter ($\sim 30\%$) and dark energy ($\sim 70\%$) with baryons playing very much a supporting role ($\sim 4\%$);
- (iii) the initial conditions are quantum fluctuations in the primordial energy density generated during inflation and
- (iv) structure has grown primarily as a result of the gravitational instability experienced by mass fluctuations in an expanding universe.

A skeptic is entitled to feel that the current paradigm is odd, to say the least. Not only is there a need to invoke vast amounts of as yet undetected non-baryonic cold dark matter, but there is also the need to account for the dominant presence of a dark energy whose very existence is a mystery within conventional models of fundamental physics. Odd as it may seem, however, this model accounts remarkably well for a large and diverse collection of empirical facts that span 13 billion years of evolution.

5. Acknowledgements

I am grateful to my collaborators for their contribution to the work reviewed here, especially Carlton Baugh, Andrew Benson, Shaun Cole, Adrian Jenkins, Cedric Lacey, Peder Norberg, John Peacock, Will Percival, and Simon White.

6. References

- Adelberger K. L., Steidel C. C., Giavalisco M., Dickinson M., Pettini M., Kellogg M. (1998), ApJ, 505, 18
- Bardeen, J. M., Bond, J. R., Kaiser, N. and Szalay, A. S. (1986), Ap.J., 304, 15
- Barnes, J., Hut, P. (1986), Nature, 324, 446
- Baugh C. M. (1996), MNRAS, 280, 26
- Baugh C. M., Cole S., Frenk C. S. (1996b), MNRAS, 283, 1361
- Baugh C. M., Cole S., Frenk C. S. Lacey C. G., (1998), ApJ, 498, 504
- Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G. (2000), MNRAS, 311, 79
- Benson, A. J., Frenk, C. S., Baugh, C. M., Cole, S., Lacey, C. G. (2001a), MNRAS, 327, 1041

- Benson, A. J., Pearce, F. R., Frenk, C. S., Baugh, C. M. & Jenkins, A. R. (2001b), MNRAS, 320, 261
- Berlind, A. & Weinberg, D. (2002), ApJ, in press (astro-ph/0109001)
- de Bernardis, P. et al (2000), Nature 404, 995
- Borgani, S. et al (2001), ApJ, 561, 13
- Cen, R. (1992), ApJ Suppl 78, 341
- Clowe, D., Luppino, G. A., Kaiser, N., Gioia, I. M. (2000), ApJ, 539, 540
- Cole, S., Hatton, S., Weinberg, D. H., Frenk, C. S. (1998), MNRAS, 300, 945
- Cole, S., Lacey, C. G., Baugh, C. M., Frenk, C.S. (2000), MNRAS, 319, 168
- Coles, P. (1993) MNRAS, 262, 1065
- Colless, M. et al (the 2dFGRS team) (2001), MNRAS, 328, 1039
- Couchman, H. M. P., Thomas, P. A., Pearce, F. R. (1995), ApJ, 452, 797
- Dave, R., Hernquist, L., Katz, N. & Weinberg, D. (1999), in Proceedings of Rencontres Internationales de l'IGRAP, Clustering at High Redshift, Marseille 1999, astro-ph/9910221
- Davis, M., Efstathiou, G., Frenk, C. S. & White, S. D. M. (1985), ApJ., 292, 371
- Efstathiou, G., Davis, M., Frenk, C. S. & White, S. D. M. (1985), ApJ Suppl, 57, 241
- Efstathiou, G. et al (The 2dFGRS team) (2002), MNRAS, in press, (astro-ph/0109152)
- Efstathiou, G., Sutherland, W. J., Maddox, S. J. (1990), Nature, 348, 705
- Eke, V. R., Cole, S. & Frenk, C. S. (1996), MNRAS, 282, 263
- Eke, V. R., Cole, S. Frenk, C. S. & Henry, J. P. (1998), MNRAS, 298, 1145
- Evrard, A. E. (1997), MNRAS, 292, 289.
- Evrard, A.E., MacFarland, T., Couchman, H.M.P., Colberg, J.M., Yoshida, N., White, S.D.M., Jenkins, A., Frenk, C.S., Pearce, F.R., Efstathiou, G., Peacock, J. & Thomas, P., (The Virgo Consortium) (2002), ApJ, in press.
- Fischer, P. et al (2000), AJ 120, 1198
- Frenk, C. S. et al (1999), ApJ, 525, 554
- Freedman, W. L. et al (2001), ApJ, 553, 47
- Gingold, R. A. & Monaghan, J. J. (1977), MNRAS 181, 375
- Gnedin, N. Y. (1995), ApJ Suppl, 97, 231
- Guth, A. H. (1981), Phys Rev D, 23, 347
- Hanany, S. et al (2000), ApJ Lett, 545, 5
- Helly, J., Cole, S. M., Frenk, C. S., Baugh, C., Benson, A. & Lacey, C. G. (2002), submitted to MNRAS (astro-ph/0202485)
- Jenkins, A., Frenk, C. S., Pearce, F. R., Thomas, P., Colberg, J., White, S. D. M., Couchman, H., Peacock, J., Efstathiou, G. & Nelson, A. (1998), ApJ, 499, 20
- Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Yoshida N. (2001), MNRAS, 321,372
- Kaiser, N. (1987), MNRAS, 227, 1
- Kauffmann, G., Colberg, J., Diaferio, A., White, S. D. M. (1999a), MNRAS, 303, 188

- Kauffmann, G., Colberg, J., Diaferio, A., White, S. D. M. (1999b), MNRAS, 307, 529
- Kauffmann G., White S. D. M., Guiderdoni B. (1993), MNRAS, 264, 201
- Klypin, A., Gottlober, S., Kravstov, A., Khokhlov, A.M. (1999), ApJ, 516, 530
- Leitch, E.M. et al (2002), ApJ, 568, 28L
- Lifshitz, E. M. (1946), J Phys, 10, 116
- Loveday, J., Peterson, B., Efstathiou, G. & Maddox, S. (1992), ApJ, 319, 338
- McKay, T. et al (2001) ApJ, in press (astro-ph/0108013)
- Moore, B., Ghigna, S., Governato, F., Lake, G., Quinn, T., Stadel, J., Tozzi, P. (1999), ApJ, 524, 19
- Navarro, J. F., Power, C., Jenkins, A. R., Frenk, C. S. & White, S. D. M. (2002), in preparation
- Norberg, P., Baugh, C. M., Hawkin, E., Maddox, S., Peacock, J., Cole, S., Frenk, C. S., et al (The 2dFGRS team) (2001a), MNRAS, 328, 64
- Norberg, P., Cole, S., Baugh, C. M., Frenk, C. S. et al (The 2dFGRS team) (2001b), submitted to MNRAS, (astro-ph/0111011)
- Peacock, J. A. & Smith, R. (2000) 318, 1144
- Peacock, J. et al (The 2dFGRS team) (2001), Nature, 410, 169
- Pearce, F. R., Jenkins, A. R., Frenk, C. S., Colberg, J., White, S. D. M., Thomas, P., Couchman, H. M. P., Peacock, J. A., Efstathiou, G. (the Virgo Consortium) (1999), ApJ, 521, 992
- Pearce, F. R., Jenkins, A. R., Frenk, C. S., White, S. D. M., Thomas, P., Couchman, H. M. P., Peacock, J. A., Efstathiou, G. (the Virgo Consortium) (2001), MNRAS, 326, 649
- Peebles, P. J. E. (1970), AJ, 75, 13
- Peebles, P. J. E. (1980), The large scale structure of the Universe. (Princeton: Princeton University Press).
- Pen, U-L. (1998), ApJ Suppl, 115, 19
- Percival, W. et al (the 2dFGRS team) 2001, MNRAS, 327, 1297
- Perlmutter, S. et al (The Supernova Cosmology Project) (1999), ApJ. 517, 565
- Pierpaoli, E., Scott, D. & White, M. (2001), MNRAS 325, 77
- Riess, A. et al (1998), AJ, 116, 1009
- Seljak, U. (2000), MNRAS, 318, 203
- Smoot, G. et al (1992), Apj 396, 1
- Somerville R. S., Primack J. R. (1999), MNRAS, 310, 1087
- Somerville, Rachel S., Primack, Joel R., Faber, S. M. (2001), MNRAS, 320, 504
- Springel, V., Yoshida, N. & White, S. D. M. (2001), New Astronomy, 6, 79
- Stadel, J. (2000), PhD thesis, University of Washington, Seattle
- Starobinskii, A. (1982), Phys. Lett. 117B, 175
- Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. (1996), ApJ, 462, L17
- Tytler, D., O'Meara, J. M., Suzuki, N. & Lubin, D. (2000), Physica Scripta, 85, 12.
- Van Waerbeke, L. et al (2001), A&A, 374, 757.

- Viana, P. T. P. & Liddle, A. (1996), MNRAS, 281, 323
Wilson, G., Kaiser, N., Luppino, G., & Cowie, L. L. (2001), ApJ, 555, 572
White, S. D. M. & Frenk, C. S. (1991), ApJ, 379, 52
White, S. D. M., Frenk, C. S. & Davis, M. (1983) ApJ 274, L1.
White, S. D. M., Navarro, J. F., Evrard, A. E., Frenk, C. S. (1993), Nature, 366, 429.
York, D.G, etal (the SDSS collaboration) (2000) AJ, 120, 1579
Yoshida, N., Stoehr, F., Springel, V. & White, S. D. M. (2002), submitted to MNRAS (astro-ph/0202341)
Zucca, E. etal (1997), AA, 326, 477.

The picture of our universe: A view from modern cosmology

David D. Reid, Daniel W. Kittell, Eric E. Arsznov, and Gregory B. Thompson

Department of Physics and Astronomy, Eastern Michigan University, Ypsilanti,

MI 48197

In this paper we give a pedagogical review of the recent observational results in cosmology from the study of type Ia supernovae and anisotropies in the cosmic microwave background. By providing consistent constraints on the cosmological parameters, these results paint a concrete picture of our present-day universe. We present this new picture and show how it can be used to answer some of the basic questions that cosmologists have been asking for several decades. This paper is most appropriate for students of general relativity and/or relativistic cosmology.

I. INTRODUCTION

Since the time that Einstein pioneered relativistic cosmology, the field of cosmology has been dominated by theoretical considerations that have ranged from straightforward applications of well-understood physics to some of the most fanciful ideas in all of science. However, in the last several years observational cosmology has taken the forefront. In particular, the results of recent observations on high-redshift supernovae and anisotropies in the radiation from the cosmic microwave background (CMB) have pinned down the major cosmological parameters to sufficient accuracy that a precise picture of our universe has now emerged. In this paper, we present this picture as currently suggested by the beautiful marriage of theory and experiment that now lies at the heart of modern cosmology.

We begin our discussion, in sections II and III, with a review of the standard theory of the present-day universe that persisted, virtually unaltered, from the time of Einstein until the mid 1990s. This review will lay most of the theoretical groundwork needed for sections

IV and V on the two experimental efforts that has had such a major impact over the last few years. Once the new results have been explained, we present, in section VI, the picture of our universe that has emerged from the recent results. We then conclude this paper with some brief comments on the implications of these results for our understanding of not just the present-day universe, but of its past and future. To help make this discussion more accessible, we use SI units with time measured in seconds instead of meters and with all factors of G and c explicitly shown unless otherwise noted.

II. REVIEW OF THE STANDARD PRESENTATION OF COSMOLOGY

The basic tenet that governs cosmology is known as the *cosmological principle*. This principle states that, on large scales, the present universe is homogeneous and isotropic. Homogeneity means that the properties of the universe are the same everywhere in the universe; and isotropy means that from every point, the properties of the universe are the same in every direction. It can be shown that the cosmological principle alone requires that the metric tensor of the universe must take the form of the Robertson-Walker metric [1]. In co-moving, spherical coordinates, this metric tensor leads to the well known line element

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

where the dimensionless function $a(t)$ is called the cosmic scale factor. This line element describes an expanding (or contracting) universe that, at the present (or any) instant in time t_0 , is a three-dimensional hypersphere of constant scalar curvature $K(t_0) = k/a^2(t_0)$. The parameter k represents the sign of this constant which can either be positive ($k = +1$ m $^{-2}$), negative ($k = -1$ m $^{-2}$), or zero ($k = 0$).

The dynamics of the universe is governed by the Einstein field equations

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu}, \quad (2)$$

where $R_{\mu\nu}$ is the Ricci tensor, R is the scalar curvature, and $T_{\mu\nu}$ is the stress-energy tensor.

On large scales, the stress-energy tensor of the universe is taken to be that of a perfect fluid (since homogeneity and isotropy imply that there is no bulk energy transport)

$$T_{\mu\nu} = (p + \rho)u_\mu u_\nu/c^2 - p g_{\mu\nu}, \quad (3)$$

where u_μ is the four-velocity of the fluid, ρ is its energy density, and p is the fluid pressure. In Eqs. (2) and (3), $g_{\mu\nu}$ is relative to the Cartesian coordinates ($x^0 = ct, x^1 = x, x^2 = y, x^3 = z$).

Under the restrictions imposed by the cosmological principle, the field equations (2) reduce to the Friedmann equations for the cosmic scale factor

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3c^2}(\rho + 3p) \quad (4)$$

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3c^2}\rho - \frac{kc^2}{a^2}, \quad (5)$$

where the dot notation represents a derivative with respect to time, as is customary. Observationally, it is known that the universe is expanding. The expansion of the universe follows the Hubble law

$$v_r = \frac{\dot{a}}{a}d = Hd, \quad (6)$$

where v_r is the speed of recession between two points, d is the proper distance between these points, and H ($\equiv \dot{a}/a$) is called the *Hubble parameter*. One of the key features of the Hubble law is that, at any given instant, the speed of recession is directly proportional to the distance. Therefore, by analyzing the Doppler shift in the light from a distant source we can infer its distance provided that we know the present value of the Hubble parameter H_0 , called the *Hubble constant* [2].

One of the principal questions that the field of cosmology hopes to answer concerns the ultimate fate of the universe. Will the universe expand forever, or will the expansion halt and be followed by a contraction? This question of the long-term fate of the expansion is closely connected to the sign of k in the Friedmann equation (5). General relativity teaches us that the curvature of spacetime is determined by the density of matter and energy. Therefore,

the two terms on the right-hand-side of Eq. (5) are not independent. The value of the energy density will determine the curvature of spacetime and, consequently, the ultimate fate of the expansion. Recognizing that the left-hand-side of Eq. (5) is H^2 , we can rewrite this expression as

$$1 = \frac{8\pi G\rho}{3c^2H^2} - \frac{kc^2}{H^2a^2}, \quad (7)$$

and make the following definition:

$$\Omega \equiv \frac{8\pi G\rho}{3c^2H^2}, \quad (8)$$

called the *density parameter*.

Since the sum of the density and curvature terms in Eq. (7) equals unity, the case for which $\Omega < 1$ corresponds to a negative curvature term requiring $k = -1 \text{ m}^{-2}$. The solution for negative curvature is such that the universe expands forever with excess velocity $\dot{a}_{t=\infty} > 0$. This latter case is referred to as an *open* universe. The case for which $\Omega > 1$ corresponds to a positive curvature term requiring $k = +1 \text{ m}^{-2}$. The solution for positive curvature (a *closed* universe) is such that the expansion eventually halts and becomes a universal contraction leading to what is known as the *big crunch*. Finally, the case for which $\Omega = 1$ corresponds to zero curvature (a *flat* universe) requiring $k = 0$. The solution for a flat universe is the critical case that lies on the boundary between an open and closed universe. In this case, the universe expands forever, but the rate of expansion approaches zero asymptotically, $\dot{a}_{t=\infty} = 0$. The value of the energy density for which $\Omega = 1$ is called the *critical density* ρ_c , given by

$$\rho_c = \frac{3c^2H^2}{8\pi G}. \quad (9)$$

The density parameter, then, is the ratio of the energy density of the universe to the critical density $\Omega = \rho/\rho_c$.

There is one final parameter that is used to characterize the universal expansion. Notice that Eq. (4) expresses the basic result that in a matter-dominated universe (in which $\rho+3p >$

0) the expansion should be decelerating, $\ddot{a} < 0$, as a result of the collective gravitational attraction of the matter and energy in the universe. This behavior is characterized by the *deceleration parameter*

$$q \equiv -\frac{\dot{\ddot{a}}/a}{\dot{a}/a^2}. \quad (10)$$

The matter in the present universe is very sparse, so that it is effectively noninteracting (i.e., dust). Therefore, it is generally assumed that the fluid pressure, p , is negligible compared to the energy density. Under these conditions, Eq. (4) shows that the deceleration parameter has a straightforward relationship to the density parameter

$$q = \Omega/2. \quad (11)$$

Collectively, the Hubble constant H_0 and the present values of the density parameter Ω_0 and the deceleration parameter q_0 are known as the *cosmological parameters*. These parameters are chosen, in part, because they are potentially measurable. The range of values that correspond to the different fates of the universe, within this traditional framework, are summarized in Table 1.

TABLE 1. The ranges of the main cosmological parameters for the three models of the universe in the standard presentation of cosmology

Model		Parameters		
50 < H_0 < 100 km·s ⁻¹ ·Mpc ⁻¹				
Open	$\Omega < 1$	$\rho_m < \rho_c$	$q_0 < 1/2$	
Closed	$\Omega > 1$	$\rho_m > \rho_c$	$q_0 > 1/2$	
Flat	$\Omega = 1$	$\rho_m = \rho_c$	$q_0 = 1/2$	

The understanding of cosmology, as outlined above, left several questions unanswered, including the fate of the universal expansion. It was entirely possible to formulate reasonable arguments for whether or not the universe is open, closed, or flat that covered all three

possibilities. The observational data has always suggested that the density of visible matter is insufficient to close the universe and researchers choosing to side with the data could easily take the position that the universe is open. However, it has been known for several decades that a substantial amount of the matter in the universe, perhaps even most of it, is not visible. The existence of large amounts of *dark matter* can be inferred from its gravitational effects both on and within galaxies [3]. Therefore, the prospects of dark matter (and neutrino mass) rendered any conclusion based solely on the amount of visible matter premature. Einstein's view was that the universe is closed, apparently for reasons having to do with Mach's principle [4], and many researchers preferred this view as well for reasons that were sometimes more philosophical than scientific. Then, there were also hints that the universe may be flat; consequentially, many researchers believed that this was most likely true.

Belief that the universe is flat was partly justified by what is known in cosmology as the *flatness problem* having to do with the apparent need of the universe to have been exceedingly close to the critical density shortly after the big bang. The fate of the universe and the flatness problem are just two of several puzzles that emerge from this standard model of the universe. Another important puzzle has to do with the existence, or nonexistence, of the *cosmological constant*, Λ . It turns out that Λ plays a very important role in our story, and *its* story must be told before we can explain how cosmologists have pinned down some of the basic properties of the universe.

III. THE COSMOLOGICAL CONSTANT

In 1915 Albert Einstein introduced his theory of General Relativity. Like Newton before him, Einstein's desire was to apply his theory to cosmology. Einstein embraced the prevailing view at that time that the universe is static. Therefore, he attempted to find solutions of the form $\dot{a} = 0$. It soon became apparent that even with Einstein's theory of gravity, as with Newton's, the gravitational attraction of the matter in the universe causes a static universe

to be unstable. Furthermore, as can be seen from Eq. (4), the subsequent requirement of $\ddot{a} = 0$ implies a negative pressure such that $p = -\rho/3$. For ordinary stellar matter and gas, this relationship is not physically reasonable.

To remedy such problems, Einstein modified his original field equations from Eq. (2) to the more general form

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R - \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}, \quad (12)$$

where Λ is the cosmological constant mentioned in the previous section. Equation (12) is the most general form of the field equations that remains consistent with the physical requirements of a relativistic theory of gravity. The cosmological constant term, for $\Lambda > 0$, can be viewed as a repulsive form of gravity that is independent of the curvature of spacetime. The modern approach is to treat Λ as a form of energy present even in empty space - vacuum energy [5]. This interpretation implies modifying Eq. (12) to

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4} \left(T_{\mu\nu} + \frac{\Lambda c^4}{8\pi G}g_{\mu\nu} \right). \quad (13)$$

In the perfect fluid approximation, this leads to an effective fluid pressure and energy density given by

$$p = p_m - \frac{\Lambda c^4}{8\pi G} \quad (14)$$

$$\rho = \rho_m + \frac{\Lambda c^4}{8\pi G}, \quad (15)$$

where p_m and ρ_m are the pressure and energy density of the *matter* content of the universe. As Eq. (14) shows, the cosmological constant contributes a negative term to the pressure in the universe. This effect of the cosmological constant allowed Einstein to find a static, albeit unstable, solution for the dynamics of the universe.

Once it became known that the universe is expanding, Einstein discarded the cosmological constant term having no other physical reason to include it. However, the possible existence of a non-zero cosmological constant has been a subject of debate ever since. With

the cosmological constant in the picture, the equations for the dynamics of the universe, Eqs. (4) and (5), generalize to

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3c^2}(\rho_m + 3p_m) + \frac{\Lambda c^2}{3} \quad (16)$$

$$\left(\frac{\dot{a}}{a}\right)^2 = H^2 = \frac{8\pi G}{3c^2}\rho_m - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3}. \quad (17)$$

Besides Einstein's static model of the universe, another interesting, and important, solution to Eqs. (16) and (17), known as the *de Sitter solution*, applies to the case of a spatially flat, empty universe ($\rho_m = 0, p_m = 0, k = 0$). In this case, Λ supplies the only contribution to the energy density

$$\rho = -p = \frac{\Lambda c^4}{8\pi G}. \quad (18)$$

Equation (16) shows that under these conditions the universe would be accelerating $\ddot{a} > 0$, and Eq. (17) shows that the Hubble parameter would be given by

$$H = \left(\frac{\Lambda c^2}{3}\right)^{1/2}. \quad (19)$$

The de Sitter solution for the cosmic scale factor shows that the effect of the cosmological constant is to cause the accelerating universe to expand exponentially with time according to

$$a(t) = a_0 e^{Ht}. \quad (20)$$

Obviously, the universe is not completely empty, but the de Sitter solution remains important because it is possible for a cosmological constant term to be sufficiently large as to dominate the dynamics of the universe. The dominant components of the universe are determined by the relative values of the corresponding density parameters. Dividing Eq. (17) by H^2 produces the analog of Eq. (7)

$$1 = \frac{8\pi G\rho_m}{3H^2c^2} + \frac{kc^2}{a^2H^2} + \frac{\Lambda c^2}{3H^2}, \quad (21)$$

where the sign of k has been separated out. The first term on the right-hand-side is called the matter term and, in analogy with Eq. (8), also gives the matter density parameter Ω_m . The second term is the curvature term and is characterized by the curvature density parameter Ω_k . Finally, the last term is known as the vacuum-energy density parameter Ω_Λ . Thus, in a universe with a cosmological constant, the primary density parameters are

$$\Omega_m = \frac{8\pi G\rho_m}{3H^2c^2}, \quad \Omega_k = -\frac{kc^2}{a^2H^2}, \quad \Omega_\Lambda = \frac{\Lambda c^2}{3H^2}. \quad (22)$$

The present values of these parameters, together with the Hubble constant, would determine the dynamics of the universe in this model [6].

IV. DETERMINING COSMOLOGICAL PARAMETERS FROM TYPE IA SUPERNOVAE

A. Type Ia Supernovae

Throughout their lives, stars remain in stable (hydrostatic) equilibrium due to the balance between outward pressures (from the fluid and radiation) and the inward pressure due to the gravitational force. The enormously energetic nuclear fusion that occurs in stellar cores causes the outward pressure. The weight of the outer region of the star causes the inward pressure. A supernova occurs when the gravitational pressure overcomes the internal pressure, causing the star to collapse, and then violently explode. There is so much energy released (in the form of light) that we can see these events out to extremely large distances.

Supernovae are classified into two types according to their spectral features and light curves (plot of luminosity vs. time). Specifically, the spectra of type Ia supernovae are hydrogen-poor, and their light curves show a sharp rise with a steady, gradual decline. In addition to these spectroscopic features, the locations of these supernovae, and the absence of planetary nebulae, allow us to determine the genesis of these events. Based on these facts, it is believed that the progenitor of a type Ia supernova is a binary star system consisting

of a *white dwarf* with a *red giant* companion [7]. Other binary systems have been theorized to cause these supernovae, but are not consistent with spectroscopic observation [8].

Although the Sun is not part of a binary system, approximately half of all stellar systems are. Both members are gravitationally bound and therefore revolve around each other. While a binary star system is very common, the members of the progenitor to a type Ia supernova have special properties. White dwarf stars are different from stars like the Sun in that nuclear fusion does not take place within these objects. Electron degeneracy pressure, which is related to the well known Pauli exclusion principle, holds the white dwarf up against its own weight. For electron degeneracy pressure to become important, an object must be extremely dense. White dwarf stars have the mass of the Sun, but are the size of the Earth. Also, the physics of this exotic form of pressure produces a strange effect: heavier white dwarfs are actually smaller in size ($\text{mass} \times \text{volume} = \text{constant}$) [9]. Red giant stars, on the other hand, are the largest known stars and contain a relatively small amount of mass. As a result, gravity is relatively weak at the exterior region of red giant stars.

In such a binary system, the strong gravitational attraction of the white dwarf overcomes the weaker gravity of the red giant. At the outer edge of the red giant, the gravitational force from the white dwarf is stronger than that from the red giant. This causes mass from the outer envelope of the red giant to be accreted onto the white dwarf. As a result, the mass of the white dwarf increases, causing its size to decrease. This process continues until the mass of the white dwarf reaches the *Chandrasekhar limit* (1.44 solar masses) beyond which electron degeneracy pressure is no longer able to balance the increasing pressure due to the gravitational force. At the center of the white dwarf, the intense pressure and temperature ignites the fusion of Carbon nuclei. This sudden burst of energy produces an explosive deflagration (subsonic) wave that destroys the star. This violently exploding white dwarf is what we see as a type Ia supernova.

The use of type Ia supernovae for determining cosmological parameters rests on the ability of these supernovae to act as standard candles. Standard candles have been used to determine distances to celestial objects for many years. They are luminous objects whose

intrinsic (or absolute) brightness can be determined independent of their distance. The intrinsic brightness, together with the observed apparent brightness (which depends on the distance to the object), can be used to calculate distances. The distance calculated from measurements of the luminosity (power output) of an object is appropriately termed the *luminosity distance*

$$d = 10^{(m-M-25)/5}, \quad (23)$$

where m is the apparent brightness measured in magnitudes (apparent magnitude), M is the absolute magnitude, and d is the luminosity distance in units of megaparsecs. The quantity, $m - M$ is commonly known as the *distance modulus*. For the reader who is unfamiliar with the magnitude scale see chapter 3 of Ref. 9.

As explained above, all type Ia supernovae are caused by the same process, a white dwarf reaching 1.44 solar masses by accretion from a red giant. As a result of this consistency, we not only expect to see extremely consistent light curves from these events, but we also expect that these light curves will reach the same peak magnitude. If this latter point is true, type Ia supernovae can be used as standard candles and, therefore, distance indicators.

Methods for determining the absolute magnitude of a type Ia supernova can be divided into two categories depending on whether or not we know the distance to the event. If we know the distance to the host galaxy of the supernova, by means of a Cepheid variable for example, and we observe the apparent magnitude of the event m , then we can use the distance modulus to calculate the absolute magnitude directly

$$m - M = 5 \log(d) + 25. \quad (24)$$

If the distance is not known, the peak luminosity must be inferred from observational data. The techniques for making this inference often involve corrections for many processes that would otherwise adversely affect the results. These processes include interstellar extinction within the host galaxy, redshift of the light from the expansion of the universe, gravitational lensing, and an apparently natural scatter in the peak brightness; see Ref. 10 for a discussion of these corrections. Once the luminosity L of a supernova has been determined, this

luminosity, together with the luminosity L' , and absolute magnitude M' , of a well-known object (such as the Sun) will yield the absolute magnitude of the supernova

$$M = M' - 2.5 \log(L/L'). \quad (25)$$

Taking all of this into account, it has been determined that the peak absolute magnitude of type Ia supernovae is [11]

$$M_{Ia} = -19.5 \pm 0.2 \text{ mag.} \quad (26)$$

B. Measuring the Hubble Constant

As stated previously, the expansion of the universe follows the Hubble law given by Eq. (6). Observationally, we measure the recession velocity as a redshift, z , in the light from the supernova ($v_r = cz$). Since every type Ia supernovae has about the same absolute magnitude, Eq. (26), the apparent magnitude provides an indirect measure of its distance. Therefore, for nearby supernovae ($z \leq 0.3$) the Hubble Law is equivalent to a relationship between the redshift and the magnitude. Inserting (26) into (24), using (6), and applying to the current epoch, yields the *redshift-magnitude relation*

$$m = M_{Ia} + 5 \log(cz) - 5 \log(H_0) + 25. \quad (27)$$

Defining the $z = 0$ intercept as

$$\tilde{M} \equiv M_{Ia} - 5 \log(H_0) + 25, \quad (28)$$

we can write equation (27) as

$$m = \tilde{M} + 5 \log(cz). \quad (29)$$

As shown in Fig. 1, low-redshift data can be used to find \tilde{M} and Eq. (28) to solve for the Hubble constant. Studies on type Ia supernova [12] consistently suggest a value for the Hubble constant of about $63 \text{ km}\cdot\text{s}^{-1}\cdot\text{Mpc}^{-1}$.

The result for H_0 , found from low-redshift supernovae, tends to set the lower bound when compared with other methods for obtaining H_0 . For example, if the distances to enough galaxies can be accurately found, then the Hubble law can be used directly to obtain a value of H_0 . This has partly been the goal of the *Hubble Space Telescope Key Project* [13]. This project has shown that a careful consideration of the type Ia supernova results in combination with the other methods for obtaining H_0 produces what has become a widely accepted value for the Hubble constant

$$H_0 = 72 \pm 8 \text{ km} \cdot \text{s}^{-1} \cdot \text{Mpc}^{-1}. \quad (30)$$

The value given in Eq. (30) is the one that we shall adopt in this paper.

C. Measuring Ω_m , Ω_Λ , and q_0

In order to determine the other cosmological parameters from the supernova data we must consider supernova at large distances ($z \geq 0.3$). Just as large distance measurements on Earth show us the curvature (geometry) of Earth's surface, so do large distance measurements in cosmology show us the geometry of the universe. Since, as we have seen, the geometry of the universe depends on the values of the cosmological parameters, measurements of the luminosity distance for distant supernova can be used to extract these values.

To obtain the general expression for the luminosity distance, consider photons from a distant source moving radially toward us. Since we are considering photons, $ds^2 = 0$, and since they are moving radially, $d\theta^2 = d\phi^2 = 0$. The Robertson-Walker metric, Eq. (1), then reduces to $0 = c^2 dt^2 - a^2 dr^2(1 - kr^2)^{-1}$, which implies

$$dt = \frac{adr}{c(1 - kr^2)^{1/2}}. \quad (31)$$

To get another expression for dt , we multiply Eq. (17) by $a^2(t)$ which produces an expression for $(da/dt)^2$. Furthermore, we note that since the universe is expanding, the matter density is a function of time. Given that lengths scale as $a(t)$, volumes scale as $a^3(t)$ and therefore,

$$\rho_m(t) \propto 1/a^3(t). \quad (32)$$

Using these facts, together with the definitions of the density parameters in Eq. (22), Eq. (17) becomes

$$\left(\frac{da}{dt}\right)^2 = H_0^2 \left[\Omega_{m,0} \frac{a_0}{a} + \Omega_{k,0} + \Omega_{\Lambda,0} \left(\frac{a}{a_0}\right)^2 \right]. \quad (33)$$

As previously mentioned, it is better to write things in terms of measurable quantities, and in this case we can directly relate the cosmic scale factor to the redshift z . The redshift is defined such that

$$1+z = \frac{\lambda_0}{\lambda}, \quad (34)$$

where λ_0 is the current (received) value of the wavelength and λ is the wavelength at the time of emission. The redshift is a direct result of the cosmic expansion and it can be shown that [14] $\lambda \propto a(t)$; therefore,

$$\frac{a_0}{a} = 1+z. \quad (35)$$

Using Eq. (35) and the fact that $\Omega_k = 1 - \Omega_m - \Omega_\Lambda$ from Eq. (21), Eq. (33) can be rewritten as

$$dt = H_0^{-1}(1+z)^{-1} \left[(1+z)^2(1+\Omega_{m,0}z) - z(z+2)\Omega_{\Lambda,0} \right]^{-1/2} dz. \quad (36)$$

Equating the expressions in Eqs. (31) and (36) and integrating, leads to an expression for the radial coordinate r of the star. The luminosity distance is then given by [15] $d = (1+z)a_0r$. Therefore,

$$d = \frac{c(1+z)}{H_0|\Omega_{k,0}|^{1/2}} \text{sinn} \left\{ |\Omega_{k,0}|^{1/2} \int_0^z \left[(1+z')^2(1+\Omega_{m,0}z') - z'(z'+2)\Omega_{\Lambda,0} \right]^{-1/2} dz' \right\}, \quad (37)$$

where $\text{sinn}(x)$ is $\sinh(x)$ for $k < 0$, $\sin(x)$ for $k > 0$, and if $k = 0$ neither sinn nor $|\Omega_{k,0}|$ appear in the expression. We see that the functional dependence of the luminosity distance is $d(z; \Omega_m, \Omega_\Lambda)$.

Inserting Eq. (37) into Eq. (24), and using the intercept from Eq. (28), we get a redshift-magnitude relation valid at high z

$$m - \tilde{M} = 5 \log[d(z; \Omega_m, \Omega_\Lambda)] \quad (38)$$

In practice, astronomers observe the apparent magnitude and redshift of a distant supernova. The density parameters are then determined by those values that produce the best fit to the observed data according to Eq. (38) for different cosmological models.

Under the continued assumption that the fluid pressure of the matter in the universe is negligible ($p_m \approx 0$), Eq. (16) implies that the deceleration parameter at the present time is given by

$$q_0 = \Omega_{m,0}/2 - \Omega_{\Lambda,0}. \quad (39)$$

Therefore, once the density parameters have been determined by the above procedure, the deceleration parameter can then be found.

Figure 2 illustrates how high-redshift data can be used to estimate the cosmological parameters and provide evidence in favor of a nonzero cosmological constant. In this figure, the abscissa is the difference between the distance moduli for the observed supernovae and what would be expected for a traditional cosmological model such as those represented in Table 1. The case shown is based on the data of Riess *et. al.* [16] using a traditional model with $\Omega_m = 0.2$ and $\Omega_\Lambda = 0$ represented by the central line $\Delta(m - M) = 0$. The figure shows that the data points lie predominantly above the zero line. This result means that the supernovae are further away (or equivalently, dimmer) than traditional, decelerating cosmological models allow. The conclusion then is that the universe must be accelerating. As suggested by Eq. (39), the most straightforward explanation of this conclusion is the presence of a nonzero, positive cosmological constant. The solid curve, above the zero line in Fig. 2, represents a best-fit curve to the data that corresponds to a universe with $\Omega_m = 0.24$ and $\Omega_\Lambda = 0.72$.

Typical values for the cosmological parameters as determined by detailed analysis of the type just discussed are the following [16]:

$$\begin{aligned}\Omega_{m,0} &= 0.24^{+0.56}_{-0.24} \\ \Omega_{\Lambda,0} &= 0.72^{+0.72}_{-0.48} \\ q_0 &= -1.0 \pm 0.4.\end{aligned}\tag{40}$$

Note that the negative deceleration parameter is consistent with an accelerating universe. Furthermore, these values imply that the universe is effectively flat predicting a curvature parameter roughly centered around $\Omega_k \approx 0.04$.

V. DETERMINING COSMOLOGICAL PARAMETERS FROM ANISOTROPIES IN THE CMB

The theory of the anisotropies in the CMB is rich with details about the contents and structure of the early universe. Consequently, this theory can become quite complicated. However, because of this same richness, this branch of cosmology holds the potential to provide meaningful constraints on a very large number of quantities of cosmological interest. Our focus here is to provide the reader with a conceptual understanding of why and how CMB anisotropies can be used to determine cosmological parameters. We will place particular emphasis on the density parameters corresponding to the spatial curvature of the universe Ω_k , and the baryon density Ω_b . The reader seeking more detail should consult Ref. 17 and the references therein.

A. Anisotropies in the CMB

The “hot big bang” model is widely accepted as the standard model of the early universe. According to this idea, our universe started in a very hot, very dense state that suddenly began to expand, and the expansion is continuing today. All of space was contained in that dense point. It is not possible to observe the expansion from an outside vantagepoint and it is not correct to think of the big bang as happening at one point in space. The big bang happened everywhere at once.

During the first fraction of a second after the big bang, it is widely believed that the universe went through a brief phase of exponential expansion called *inflation* [18]. Baryonic matter formed in about the first second; and the nuclei of the light elements began to form (nucleosynthesis) when the universe was only several minutes old. Baryons are particles made up of three quarks; the most familiar baryons are the protons and neutrons in the nuclei of atoms. Since all of the matter that we normally encounter is made up of atoms, baryonic matter is considered to be the “ordinary” matter in the universe.

The very early universe was hot enough to keep matter ionized, so the universe was filled with nucleons and free electrons. The density of free electrons was so high that Thomson scattering effectively made the universe opaque to electromagnetic radiation. The universe remained a baryonic plasma until around 300,000 years after the big bang when the universe had expanded and cooled to approximately 3000 K. At this point, the universe was sufficiently cool that the free electrons could join with protons to form neutral hydrogen. This process is called *recombination*. With electrons being taken up by atoms, the density of free electrons became sufficiently low that the mean free path of the photons became much larger (on the order of the size of the universe); and light was free to propagate. The light that was freed during recombination has now cooled to a temperature of about $T_o = 2.73$ K. This light is what we observe today as the cosmic microwave background. We see the CMB as if it were coming from a spherical shell called the *surface of last scattering* (Fig. 3). This shell has a finite thickness because recombination occurred over a finite amount of time.

Today, over very large scales, the universe is homogeneous. However, as evidenced by our own existence, and the existence of galaxies and groups of galaxies, etc., inhomogeneities exist up to scales on the order of 100 Mpc. Theories of structure formation require that the seeds of the structure we observe today must have been inhomogeneities in the matter density of the early universe. These inhomogeneities would have left their imprint in the CMB which we would observe today as temperature anisotropies. So, in order to explain the universe in which we live, there should be bumps in the CMB; and these bumps should occur over angular scales that correspond to the scale of observed structure. In 1992, the

COBE satellite measured temperature fluctuations δT in the CMB, $\delta T/T \sim 10^{-5}$ on a 7° angular scale [19], where T is the ambient temperature of the CMB. The anisotropies detected by COBE are considered to be large-scale variations caused by nonuniformities generated at the creation of the universe. However, recent observations [20-22] have found small-scale anisotropies that correspond to the physical scale of today's observed structure. It is believed that these latter anisotropies are the result of quantum fluctuations in density that existed prior to inflation which were greatly amplified during inflation. These amplified fluctuations became the intrinsic density perturbations which are the seeds of structure formation.

The small-scale anisotropies in the CMB can be separated into two categories: primary and secondary. Primary anisotropies are due to effects that occur at the time of recombination and are “imprinted” in the CMB as the photons leave the surface of last scattering. Secondary anisotropies arise through scattering along the line of sight between the surface of last scattering and the observer. In this paper, we will only be concerned with the primary anisotropies. There are three main sources for primary anisotropies in the microwave background. These are the Sachs-Wolfe effect, intrinsic (adiabatic) perturbations, and a Doppler effect.

For the largest of these primary anisotropies the dominant mechanism is the Sachs-Wolfe effect. At the surface of last scattering, matter density fluctuations will lead to perturbations in the gravitational potential, $\delta\Phi$. These perturbations cause a gravitational redshift of the photons coming from the surface of last scattering as they “climb out” of the potential wells. This effect is described by, $\delta T/T = \delta\Phi/c^2$. These same perturbations in the gravitational potential also cause a time dilation at the surface of last scattering, so these photons appear to come from a younger, hotter universe. This effect is described by, $\delta T/T = -2(\delta\Phi)/3c^2$. Combining these two processes gives the Sachs-Wolfe effect [23],

$$\frac{\delta T}{T} = \frac{\delta\Phi}{3c^2}. \quad (41)$$

On intermediate scales, the main effect is due to adiabatic perturbations. Recombination

occurs later in regions of higher density, so photons emanating from overly dense regions experience a smaller redshift from the universal expansion and thus appear hotter. The observed temperature anisotropy resulting from this process is given by [23],

$$\left(\frac{\delta T}{T}\right)_{obs} = -\frac{\delta z}{1+z} = \frac{\delta\rho}{\rho}. \quad (42)$$

Finally, on smaller scales there is a Doppler effect that becomes important. This effect arises because the photons are last scattered in a moving plasma. The temperature anisotropy corresponding to this effect is described by [23],

$$\frac{\delta T}{T} = \frac{\delta \vec{v} \cdot \hat{r}}{c}, \quad (43)$$

where \hat{r} denotes the direction along the line of sight and \vec{v} is a characteristic velocity of the material in the scattering medium.

B. Acoustic Peaks and the Cosmological Parameters

The early universe was a plasma of photons and baryons and can be treated as a single fluid [24]. Baryons fell into the gravitational potential wells created by the density fluctuations and were compressed. This compression gave rise to a hotter plasma thus increasing the outward radiation pressure from the photons. Eventually, this radiation pressure halted the compression and caused the plasma to expand (rarefy) and cool producing less radiation pressure. With a decreased radiation pressure, the region reached the point where gravity again dominated and produced another compression phase. Thus, the struggle between gravity and radiation pressure set up longitudinal (acoustic) oscillations in the photon-baryon fluid. When matter and radiation decoupled at recombination the pattern of acoustic oscillations became frozen into the CMB. Today, we detect the evidence of the sound waves (regions of higher and lower density) via the primary CMB anisotropies.

It is well known that any sound wave, no matter how complicated, can be decomposed into a superposition of wave modes of different wavenumbers k , each k being inversely

proportional to the physical size of the corresponding wave (its wavelength), $k \propto 1/\lambda$. Observationally, what is seen is a projection of the sound waves onto the sky. So, the wavelength of a particular mode λ is observed to subtend a particular angle θ on the sky. Therefore, to facilitate comparison between theory and observation, instead of a Fourier decomposition of the acoustic oscillations in terms of sines and cosines, we use an angular decomposition (multipole expansion) in terms of Legendre polynomials $P_\ell(\cos \theta)$. The order of the polynomial ℓ (related to the multipole moments) plays a similar role for the angular decomposition as the wavenumber k does for the Fourier decomposition. For $\ell \geq 2$ the Legendre polynomials on the interval [-1,1] are oscillating functions containing a greater number of oscillations as ℓ increases. Therefore, the value of ℓ is inversely proportional to the characteristic angular size of the wave mode it describes

$$\ell \propto 1/\theta. \quad (44)$$

Experimentally, temperature fluctuations can be analyzed in pairs, in directions \hat{n} and \hat{n}' that are separated by an angle θ so that $\hat{n} \cdot \hat{n}' = \cos \theta$. By averaging over all such pairs, under the assumption that the fluctuations are Gaussian, we obtain the two-point correlation function, $C(\theta)$, which is written in terms of the multipole expansion

$$\langle \delta T(\hat{n}) \cdot \delta T(\hat{n}') \rangle \equiv C(\theta) = \sum_{\ell} \frac{(2\ell + 1)}{4\pi} C_{\ell} P_{\ell}(\cos \theta), \quad (45)$$

the C_{ℓ} coefficients are called the multipole moments.

As predicted, analysis of the temperature fluctuations does in fact reveal patterns corresponding to a harmonic series of longitudinal oscillations. The various modes correspond to the number of oscillations completed before recombination. The longest wavelength mode, subtending the largest angular size for the primary anisotropies, is the fundamental mode – this was the first mode detected. There is now strong evidence that both the 2nd and 3rd modes have also been observed [20-22].

The distance sound waves could have traveled in the time before recombination is called the *sound horizon*, r_s . The sound horizon is a fixed physical scale at the surface of last scattering. The size of the sound horizon depends on the values of the cosmological parameters.

The distance to the surface of last scattering, d_{sls} , also depends on cosmological parameters. Together, they determine the angular size of the sound horizon (see Fig. 3)

$$\theta_s \approx \frac{r_s}{d_{sls}}, \quad (46)$$

in the same way that the angle subtended by the planet Jupiter depends on both its size and distance from us. Analysis of the temperature anisotropies in the CMB determine θ_s and the cosmological parameters can be varied in r_s and d_{sls} to determine the best-fit results.

We can estimate the sound horizon by the distance that sound can travel from the big bang, $t = 0$, to recombination t_*

$$r_s(z_*; \Omega_b, \Omega_r) \approx \int_0^{t_*} c_s dt, \quad (47)$$

where z_* is the redshift parameter at recombination ($z_* \approx 1100$) [25], Ω_r is the density parameter for radiation (photons), c_s is the speed of sound in the photon-baryon fluid, given by [26]

$$c_s \approx c [3(1 + 3\Omega_b/4\Omega_r)]^{-1/2}, \quad (48)$$

which depends on the baryon-to-photon density ratio, and dt is determined by an expression similar to Eq. (36), except at an epoch in which radiation plays a more important role. The energy density of radiation scales as $\rho_r \propto a^{-4}$ [27], so with the addition of radiation, Eq. (33) generalizes to

$$\left(\frac{da}{dt}\right)^2 = H_0^2 \left[\Omega_{r,0} \left(\frac{a_0}{a}\right)^2 + \Omega_{m,0} \frac{a_0}{a} + \Omega_{k,0} + \Omega_{\Lambda,0} \left(\frac{a}{a_0}\right)^2 \right], \quad (49)$$

which, upon using Eq. (35) and $\Omega_r + \Omega_m + \Omega_\Lambda + \Omega_k = 1$, leads to

$$dt = H_0^{-1}(1+z)^{-1} \left\{ (1+z)^2(1+\Omega_{m,0}z) + z(z+2) \left[(1+z)^2\Omega_{r,0} - \Omega_{\Lambda,0} \right] \right\}^{-1/2} dz. \quad (50)$$

The distance to the surface of last scattering, corresponding to its angular size, is given by what is called the angular diameter distance. It has a simple relationship to the luminosity distance d [15] given in Eq. (37)

$$d_{sls} = \frac{d(z_*; \Omega_m, \Omega_\Lambda)}{(1 + z_*)^2}. \quad (51)$$

The location of the first acoustic peak is given by $\ell \approx d_{sls}/r_s$ and is most sensitive to the curvature of the universe Ω_k .

To get a feeling for this result, we can consider a very simplified, heuristic calculation. We will consider a prediction for the first acoustic peak for the case of a flat universe. To leading order, the speed of sound in the photon-baryon fluid, Eq. (48), is constant $c_s = c/\sqrt{3}$. We further make the simplifying assumption that the early universe was matter-dominated (there is good reason to believe that it was which will be discussed in the next section). With these assumptions, Eqs. (49) and (47) yield (dropping the ‘0’ from the density parameters)

$$r_s = \frac{c_s}{H_0 \sqrt{\Omega_m}} \int_{z_*}^{\infty} (1 + z)^{-5/2} dz, \quad (52)$$

which gives

$$r_s = \frac{2c_s}{3H_0 \sqrt{\Omega_m}} (1 + z_*)^{-3/2}. \quad (53)$$

The distance to the surface of last scattering, in our flat universe model, will depend on both Ω_m and Ω_Λ . Following a procedure similar to that which lead to Eq. (37), the radial coordinate of the surface of last scattering, r_{sls} (not to be confused with r_s), is determined by

$$r_{sls} = \frac{c}{H_0} \int_0^{z_*} [\Omega_m(1 + z)^3 + \Omega_\Lambda]^{-1/2} dz, \quad (54)$$

which does not yield a simple result. Using a binomial expansion, the integrand can be approximated as $\Omega_m^{-1/2}(1 + z)^{-3/2} - (\Omega_\Lambda/2\Omega_m^{3/2})(1 + z)^{-9/2}$ and the integral is more easily handled. The distance is then determined by $d_{sls} = r_{sls}/(1 + z_*)$ which gives

$$d_{sls} = \frac{2c}{7H_0(1 + z_*)} \left\{ 7\Omega_m^{-1/2} - 2\Omega_\Lambda\Omega_m^{-3/2} + O[(1 + z_*)^{-1/2}] \right\}. \quad (55)$$

Using $\Omega_\Lambda = 1 - \Omega_m$ and neglecting the higher order terms gives

$$d_{sls} \approx \frac{2c\Omega_m^{-1/2}}{7H_0(1 + z_*)} \left\{ 9 - 2\Omega_m^3 \right\}. \quad (56)$$

Combining Eqs. (53) and (56) to get our prediction for the first acoustic peak gives

$$\ell \approx \frac{d_{sls}}{r_s} \approx 0.74\sqrt{(1+z_*)}\left\{9 - 2\Omega_m^3\right\} \approx 221. \quad (57)$$

This result is consistent with the more detailed result that [28]

$$\ell \approx 200/\sqrt{1-\Omega_k}, \quad (58)$$

where, in our calculation $\Omega_k = 0$. Equation (58) suggests that a measurement of $\ell \approx 200$ implies a flat universe. The BOOMERanG [22] collaboration found $\ell \approx 197 \pm 6$, and the MAXIMA-1 [21] collaboration measured $\ell \approx 220$. Additional simplified illustrations for how the cosmological parameters can be obtained from the acoustic peak can be found in Ref. 29.

Experimental results, such as those quoted above, are determined by plotting the power spectrum (power per logarithmic interval), $(\delta T_\ell)^2$, given by

$$(\delta T_\ell)^2 = \frac{\ell(\ell+1)}{2\pi} C_\ell, \quad (59)$$

or by the square root of this quantity. The power spectrum may be quickly calculated for a given cosmological model using a code such as CMBFAST which is freely available online [30]. The solid curve in Fig. 4 was calculated using CMBFAST and the data points are only a representative few included to show the kind of agreement between theory and experiment that exists.

While the location of the first acoustic peak helps to fix Ω_k , other features of the power spectrum help to determine the baryon density. Since baryons are the primary cause of the gravitational potential wells that help generate the acoustic oscillations, they affect the power spectrum in several ways. The relative heights of the peaks are an indication of Ω_b in that an increase in baryon density results in an enhancement of the odd peaks. An increase in baryon density also leads to enhanced damping at higher multipoles [31].

It is important to recognize that the constraints on cosmological parameters obtained through this sort of analysis are correlated so that the range of possible values of Ω_Λ ,

for example, depends on what is assumed for the possible range of values of the Hubble constant. Therefore, it is customary to incorporate results from other observational (or theoretical) work in the analysis of the CMB data. With this in mind, we use the value of the Hubble constant stated in Eq. (30). Given this assumption, a combined study of the CMB anisotropy data from the BOOMERanG [22], MAXIMA-1 [21], and COBE-DMR [32] collaborations suggests the following values for the two cosmological parameters being considered here [33]:

$$\begin{aligned}\Omega_{k,0} &= 0.11 \pm 0.07 \\ \Omega_{b,0} &= 0.062 \pm 0.01.\end{aligned}\tag{60}$$

As with the type Ia supernova results, the best-fit CMB results predict an essentially flat universe. In fact, it is quite possible to adopt a model with $\Omega_k \equiv 0$ and still obtain a very good fit to the data along with reasonable values for the other cosmological parameters [33]. Again, the CMB data also provides values for additional cosmological parameters, but the curvature and baryon densities are perhaps the most accurately constrained at this time.

Even though the recent revolution in cosmology was ignited by the type Ia supernova and CMB anisotropy results, it is also important to acknowledge prior work toward constraining the cosmological parameters. This work includes investigations on gravitational lensing [34], large-scale structure [35], and the ages of stars, galaxies, and globular clusters [36]. Without this work, the ability to use the supernova and CMB data to place fairly tight restrictions on the major cosmological parameters would be significantly diminished.

VI. THE PICTURE OF OUR UNIVERSE

Given the results from observational cosmology discussed in the previous two sections we are now able to present a concrete picture of the universe, as opposed to the traditional array of models with very different properties. Taking a more comprehensive view, in Table 2 we present a set of cosmological parameters (without errors) that might be taken as the “best estimates” based on various observational and theoretical studies [37].

TABLE 2. Our best estimates of the cosmological parameters for the present-day universe and the primary sources we used to obtain them. If “theory” is listed as the source, we derived the value from other estimates by using the stated equation.

Parameter	Value	Primary Sources
Hubble Constant	$H_0 = 72 \text{ km}\cdot\text{s}^{-1}\cdot\text{Mpc}^{-1}$	[13]
Cosmological Constant	$\Omega_\Lambda = 0.70$	[16, 33]
Matter	$\Omega_m = 0.30$	[16, 33]
Baryonic matter	$\Omega_b = 0.04$	[33]
Dark matter	$\Omega_{CDM} = 0.26$	theory: Eq. (61)
Curvature	$\Omega_k = 0.00$	[16, 20-22, 33]
Deceleration parameter	$q_0 = -0.55$	theory: Eq. (39)

This set of parameters describes a flat universe the dynamics of which is dominated by two mysterious forms of energy, most prominently, the cosmological constant. So then, the long-standing debate over whether or not the cosmological term should be included in Einstein’s theory is over; not only should it be included, it dominates the universe. Although the debate over the existence of the cosmological constant has ended, the debate over its physical implications has just begun. Further comments about this debate will be discussed in the conclusion.

The other mysterious form of energy listed in Table 2, Ω_{CDM} , is dark matter where “CDM” stands for “cold dark matter.” Recall that ordinary matter made up of atomic nuclei only contributes to the baryon content of the universe with $\Omega_b \approx 0.04$. However, since the total matter content is $\Omega_m \approx 0.30$, the rest of the matter in the universe must be in some exotic, unseen form which is why we call it dark matter

$$\Omega_{CDM} = \Omega_m - \Omega_b \quad (61)$$

We have known about dark matter for several decades now, having been first discovered through anomalous rotation curves of galaxies [3]. The results from the CMB anisotropies only help to confirm that not only does dark matter exist, but that it comprises roughly 90% of the matter in the universe.

Given values of the cosmological parameters, we can now solve for the dynamics of the universe. The Friedmann equations (16) and (17) for our present ($p_m = 0, k = 0$) universe can be combined to give

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 = \Lambda c^2. \quad (62)$$

This equation can be solved exactly giving the result [38]

$$a(t) = A^{1/3} \sinh^{2/3} \left(\frac{t}{t_\Lambda} \right), \quad (63)$$

where $A = \Omega_{m,0}/\Omega_{\Lambda,0} \approx 0.43$ and $t_\Lambda = (4/3\Lambda c^2)^{1/2} \approx 3.4 \times 10^{17}$ s. The cosmic scale factor is plotted in Fig. 5 and compared to the purely de Sitter universe described by Eq. (20). From this comparison, we see that, today, the qualitative behavior of our universe is that of a de Sitter universe except that the presence of matter has caused the universe to expand less than in the de Sitter case.

With $a(t)$ in hand, we can now write a precise metric for the universe

$$ds^2 = c^2 dt^2 - A^{2/3} \sinh^{4/3}(t/t_\Lambda) [dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2]. \quad (64)$$

This tells us that we can visualize the universe as an expanding Euclidean sphere with the expansion governed by $a(t)$ as given in Eq. (63). Note, however, that in this visualization the universe is represented as the entire volume of the sphere and not just the surface.

Another interesting feature that emerges from this picture is that if Λ is truly constant, the universe would have once been matter-dominated. To see why this is, recall that because the size of the universe changes, the density parameters are functions of time. As we go back in time, the universe gets smaller so that the energy density of matter ρ_m gets larger while the energy density associated with Λ , see Eq. (15), remains constant. Using Eq. (22) we can see that the ratio of matter-to-cosmological constant is

$$\frac{\Omega_m}{\Omega_\Lambda} = \frac{\Omega_{m,0}}{\Omega_{\Lambda,0}} a^{-3}(t). \quad (65)$$

Therefore, at some finite time in the past the universe was such that $\Omega_m/\Omega_\Lambda > 1$. Since the expansion of a matter-dominated universe would be decelerating, this implies that the universe underwent a transition from decelerated expansion to accelerated expansion. This behavior is reflected in the deceleration parameter as a function of time, which, given the current cosmological parameters becomes

$$q(t) = \frac{1}{2} \left[1 - 3 \tanh^2(t/t_\Lambda) \right]. \quad (66)$$

Figure 6 is a plot of $q(t)$ and shows that the deceleration parameter was once positive and that a transition to $q(t) < 0$ occurred around the time at which $\Omega_m/\Omega_\Lambda = 1$.

Having a specific model of the universe allows us to determine specific answers to questions that cosmologists have been asking for decades. While we cannot address all such questions in this paper we will tackle a few of the most common: (a) What is the age of the universe? (b) Will the universe expand forever or will the expansion eventually stop followed by a re-collapse? (c) Where is the edge of the observable universe?

The age of the universe can be calculated by integrating dt from now, $z = 0$, back to the beginning $z = \infty$. For our universe, the steps leading to Eq. (36) produces

$$dt = H_0^{-1} (1+z)^{-1} \left[\Omega_{m,0} (1+z)^3 + \Omega_{\Lambda,0} \right]^{-1/2} dz. \quad (67)$$

Making the definition $x \equiv 1+z$, the present age of the universe is given by

$$t_0 = H_0^{-1} \int_1^\infty \left[\Omega_{m,0} x^5 + \Omega_{\Lambda,0} x^2 \right]^{-1/2} dx. \quad (68)$$

The solution to Eq. (68) is complex. Taking only the real part gives

$$t_0 = \frac{2}{3H_0 \Omega_{\Lambda,0}^{1/2}} \tanh^{-1} \left[\left(1 + \frac{\Omega_{m,0}}{\Omega_{\Lambda,0}} \right)^{1/2} \right] = 13.1 \times 10^9 \text{ yr}. \quad (69)$$

The question of whether or not the universe will expand forever is determined by the asymptotic behavior of $a(t)$. Since $\sinh(x)$ diverges as $x \rightarrow \infty$, it is clear that the universe will continue to expand indefinitely unless some presently unknown physical process drastically alters its dynamics.

Finally, concerning the question of the size of the observable universe, there are two types of horizons that might fit this description, the *particle horizon* and the *event horizon*. The particle horizon is the position of the most distant event that can presently be seen, that is, from which light has had enough time to reach us since the beginning of the universe. Unfortunately, since current evidence suggests that the universe was not always dominated by the cosmological constant, we cannot extend the current model back to the beginning. We can, however, extend it into the future. The event horizon is the position of the most distant event that we will ever see. If we consider a photon moving radially toward us from this event, then Eq. (31) describes its flight. Since we are interested in those events that will occur from now t_0 , onward, Eq. (31) leads to

$$\int_0^{r_{EH}} dr = cA^{-1/3} \int_{t_0}^{\infty} \sinh^{-2/3}(t/t_{\Lambda}) dt, \quad (70)$$

where r_{EH} is the radial coordinate of our event horizon. Performing a numerical solution to the integral yields

$$r_{EH} \approx 1.2ct_{\Lambda} = 16 \times 10^9 \text{ light years}. \quad (71)$$

This result suggests that 16 billion light years is the furthest that we will ever be able to see. As far as we are aware, the most distant object ever observed (besides the CMB) is currently the galaxy RD1 at a redshift of $z = 5.34$, which places it approximately 12.2 billion light years away [39].

VII. CONCLUSIONS

In summary, the resent observational results in cosmology strongly suggest that we live in a universe that is spatially flat, expanding at an accelerated rate, homogeneous and isotropic on large scales, and is approximately 13 billion years old. The expansion of the universe is described by Eq. (63), and its metric by Eq. (64). We have seen that roughly 96% of the matter and energy in the universe consists of cold dark matter and the cosmological

constant. We now know basic facts about the universe much more precisely than we ever have. However, since we cannot speak with confidence about the nature of dark matter or the cosmological constant, perhaps the most interesting thing about all of this is that knowing more about the universe has only shown us just how little we really understand.

As mentioned previously, the most common view of the cosmological constant is that it is a form of vacuum energy due, perhaps, to quantum fluctuations in spacetime [5]. However, within the context of general relativity alone there is no need for such an interpretation; Λ is just a natural part of the geometric theory [40]. If, however, we adopt the view that the cosmological constant belongs more with the energy-momentum tensor than with the curvature tensor, this opens up a host of possibilities including the possibility that Λ is a function of time [41].

In conclusion, it is also important to state that although this paper emphasizes what the recent results say about our present universe, these results also have strong implications for our understanding of the distant past and future of the universe. For an entertaining discussion of the future of the universe see Ref. 42. Concerning the past, the results on anisotropies in the CMB have provided strong evidence in favor of the inflationary scenario, which requires a Λ -like field in the early universe to drive the inflationary dynamics. To quote White and Cohn, “Of dozens of theories proposed before 1990, only inflation and cosmological defects survived after the COBE announcement, and only inflation is currently regarded as viable by the majority of cosmologists” [17].

VIII. ACKNOWLEDGEMENTS

We would like to acknowledge (and recommend) the excellent website of Dr. Wayne Hu [43]. This resource is very useful for learning about the physics of CMB anisotropies. We are also grateful to Dr. Manasse Mbonye for making several useful suggestions.

- [1] Steven Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*, (John Wiley & Sons, New York, 1972).
- [2] It should be noted that what we have called the Hubble parameter is also often called the Hubble constant. This name is inappropriate, however, because the Hubble parameter is time dependent.
- [3] Vera Rubin, *Bright Galaxies, Dark Matter*, (Springer Verlag/AIP Press, New York, 1997).
- [4] P. J. E. Peebles, *Principles of Physical Cosmology*, (Princeton University Press, Princeton, 1993) p. 14.
- [5] Sean M. Carroll, “The Cosmological Constant,” Living Rev. Rel. **4**, 1-80 (2001); astro-ph/0004075.
- [6] For additional discussion on the cosmological constant see A. Harvey and E. Schucking, “Einstein’s mistake and the cosmological constant,” Am. J. Phys. **68**, 723-727 (2000).
- [7] Stephen Webb, *Measuring the Universe: The Cosmological Distance Ladder*, (Praxis, Chichester, UK, 1999), pp. 229-230.
- [8] P. Hoflich, Thielemann, F. K., and Wheeler J. C., “Type Ia Supernovae: Influence of the Initial Composition on the Nucleosynthesis, Light Curves, Spectra, and Consequences for the Determination of Ω_M & Λ ,” Astrophys. J. **495**, 617-629 (1998); astro-ph/9709233.
- [9] Bradley W. Carroll and Dale A. Ostlie, *An Introduction to Modern Astrophysics* (Addison-Wesley, Reading, MA, 1996) pp. 588-590.
- [10] M. Hamuy, M. M. Phillips, J. Maza *et al.*, “A Hubble Diagram of Distant Type Ia Supernovae,” The Astron. Journal, **109**(1), 1-13 (1995).
- [11] See p. 231 of Ref. 7.
- [12] M. Hamuy, M. M. Phillips, N. Suntzeff, and R. Schommer, “The Hubble Diagram of the Calán/Tololo Type Ia Supernovae and the value of H_0 ,” The Astron. Journal, **112**, 2398-2407

- (1996); astro-ph/9609062. A. G. Kim, S. Gabi, G. Goldhaber *et al.* “Implications for the Hubble Constant from the First Seven Supernovae at $z \leq 0.35$,” *Astrophys. J.* **476**, L63-66 (1997).
- [13] Wendy L. Freedman, B. F. Madore, B. K. Gibson *et al.*, “Final Results from the Hubble Space Telescope Key Project to Measure the Hubble Constant,” *Astrophys. J.*, **553**, 47-72 (2001).
- [14] See p. 97 of Ref. 4.
- [15] S. M. Carroll, W. H. Press, and E. L. Turner, “The Cosmological Constant,” *Ann. Rev. Astron. & Astrophys.*, **30**, 499-542 (1992).
- [16] Adam G. Riess, Alexei V. Filippenko, Peter Challis *et al.*, “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant,” *Astron. J.*, **116**, 1009-1038 (1998).
- [17] Martin White and J. D. Cohn, “Resource Letter: TACMB-1: The theory of anisotropies in the cosmic microwave background,” *Am. J. Phys.* **70**, 106-118 (2002).
- [18] Alan H. Guth, *The Inflationary Universe*, (Perseus Books, Reading, MA, 1997).
- [19] G. F. Smoot, C. L. Bennett, A. Kogut *et al.*, “Structure in the COBE differential microwave radiometer first-year maps,” *Astrophys. J.* **396**, L1-5 (1992).
- [20] A. D. Miller, R. Caldwell, M. J. Devlin *et al.*, “A Measurement of the Angular Power Spectrum of the Cosmic Microwave Background from $l = 100$ to 400,” *Astrophys. J.*, **524**, L1-4 (1999); astro-ph/9906421.
- [21] S. Hanany, P. Ade, A. Balbi *et al.*, “MAXIMA-1: A Measurement of the Cosmic Microwave Background Anisotropy on Angular Scales of $10' - 5^\circ$,” *Astrophys. J.*, **545**, L5-9 (2000); astro-ph/0005123.
- [22] P. de Bernardis, P. A. R. Ade, J. J. Bock *et al.*, “A Flat Universe from High-resolution Maps of the Cosmic Microwave Background Radiation,” *Nature*, **404**, 955-959 (2000); astro-ph/0004404.
- [23] J. A. Peacock, *Cosmological Physics*, (Cambridge University Press, New York, 1999) pp. 591-592.

- [24] P. J. E. Peebles and J. T. Yu, “Primeval Adiabatic Perturbation in an Expanding Universe,” *Astrophys. J.*, **162**, 815-836 (1970).
- [25] Martin White, Douglas Scott, and Joseph Silk, “Anisotropies in the Cosmic Microwave Background,” *Ann. Rev. Astron. Atrophys.*, **32**, 319-370 (1994).
- [26] W. Hu and M. White, “Acoustic Signatures in the Cosmic Microwave Background,” *Astrophys. J.*, **471**, 30-51 (1996); astro-ph/9602019.
- [27] Hans C. Ohanian and Remo Ruffini, *Gravitation and Spacetime*, 2nd ed., (W. W. Norton & Co., New York, 1994) chap. 10.
- [28] See Ref. 21.
- [29] Neil J. Cornish, “Using the acoustic peak to measure cosmological parameters,” *Phys. Rev. D*, **63**, 027302-1 - 4 (2001); astro-ph/0005261.
- [30] See <http://physics.nyu.edu/matiasz/CMBFAST/cmbfast.html>
- [31] See Ref. 26.
- [32] C. L. Bennett, A. J. Banday, K. M. Górski *et al.*, “Four-Year COBE DMR Cosmic Microwave Background Observations: Maps and Basic Results,” *Astrophys. J.*, **464**, L1-4 (1996).
- [33] A. H. Jaffe, P. A. R. Ade, A. Balbi *et al.*, “Cosmology from MAXIMA-1, BOOMERANG & CODE/DMR CMB Observations,” *Phys. Rev. Lett.*, **86**, 3475-3479 (2001); astro-ph/0007333.
- [34] Kyu-Hyun Chae, “New Modeling of the Lensing Galaxy and Cluster of Q0957+561: Implications for the Global Value of the Hubble Constant,” *Astrophys. J.*, **524**, 582-590 (1999).
- [35] J. A. Peacock, S. Cole, P. Norberg *et al.*, “A measurement of the cosmological mass density from clustering in the 2dF Galaxy Redshift Survey,” *Nature*, **410**, 169-173 (2001); astro-ph/0103143.
- [36] L. M. Krauss, “The Age of Globular Clusters,” *Phys. Rept.*, **333**, 33-45 (2000); astro-ph/9907308.

- [37] The values that we have used are predominantly consistent with those of C. H. Linewater, “Cosmological Parameters,” astro-ph/0112381.
- [38] Ø. Grøn, “A new standard model of the universe,” Eur. J. Phys. **23**, 135-144 (2002).
- [39] A. Dey, H. Spinrad, D. Stern *et al.*, “A galaxy at $z = 5.34$,” Astrophys. J., **498**, L93-97 (1998).
- [40] See, for example, Alex Harvey, “Is the Universe’s Expansion Accelerating?,” Phys. Today (Letters), **55**, 73-74 (2002).
- [41] See for example, M. Doran and C. Wetterich, “Quintessence and the cosmological constant,” astro-ph/0205267; and M. Mbonye, “Matter fields from a decaying background Lambda vacuum,” astro-ph/0208244.
- [42] Fred Adams and Greg Laughlin, *The Five Ages of the Universe: Inside the Physics of Eternity*, (Simon & Schuster, New York, 1999).
- [43] See <http://background.uchicago.edu/~whu/>.

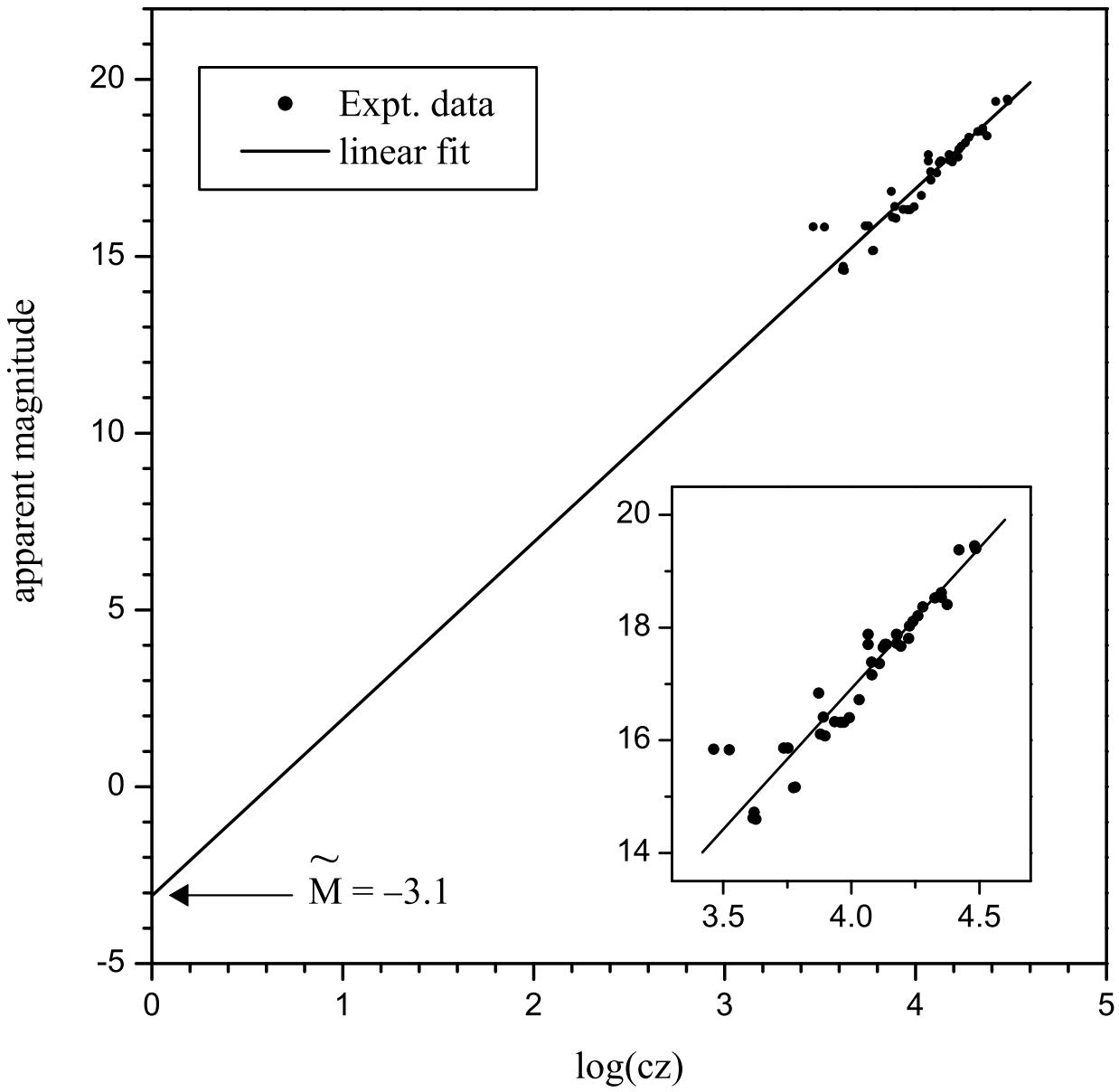


FIGURE 1. Using low-redshift supernovae to determine the Hubble constant. The data points are from Ref. 12. The inset graph shows a close-up view of the data and the best-fit line. The best-fit line determines the intercept which can be used in Eq. (28) to determine the Hubble constant.

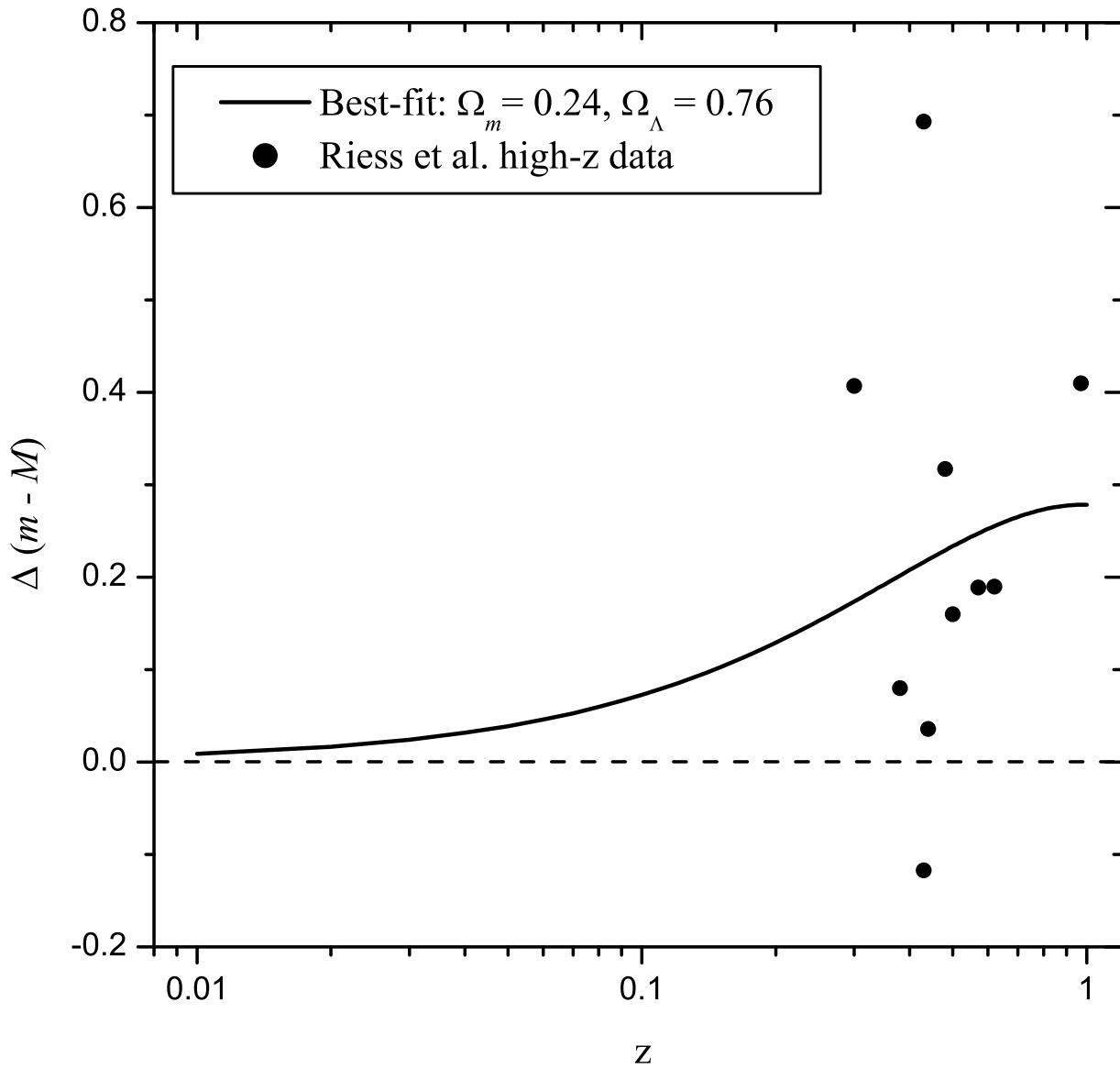


FIGURE 2. Using high-redshift data to determine cosmological parameters and provide evidence for a nonzero cosmological constant. The zero line corresponds to a traditional decelerating model of the universe with $\Omega_m = 0.2$, $\Omega_\Lambda = 0$, and $\Omega_k = 0.8$. The data points are the high-redshift supernovae from Ref. 16. The solid curve corresponds to those cosmological parameters that produce a best fit to the data points as determined in Ref. 16.

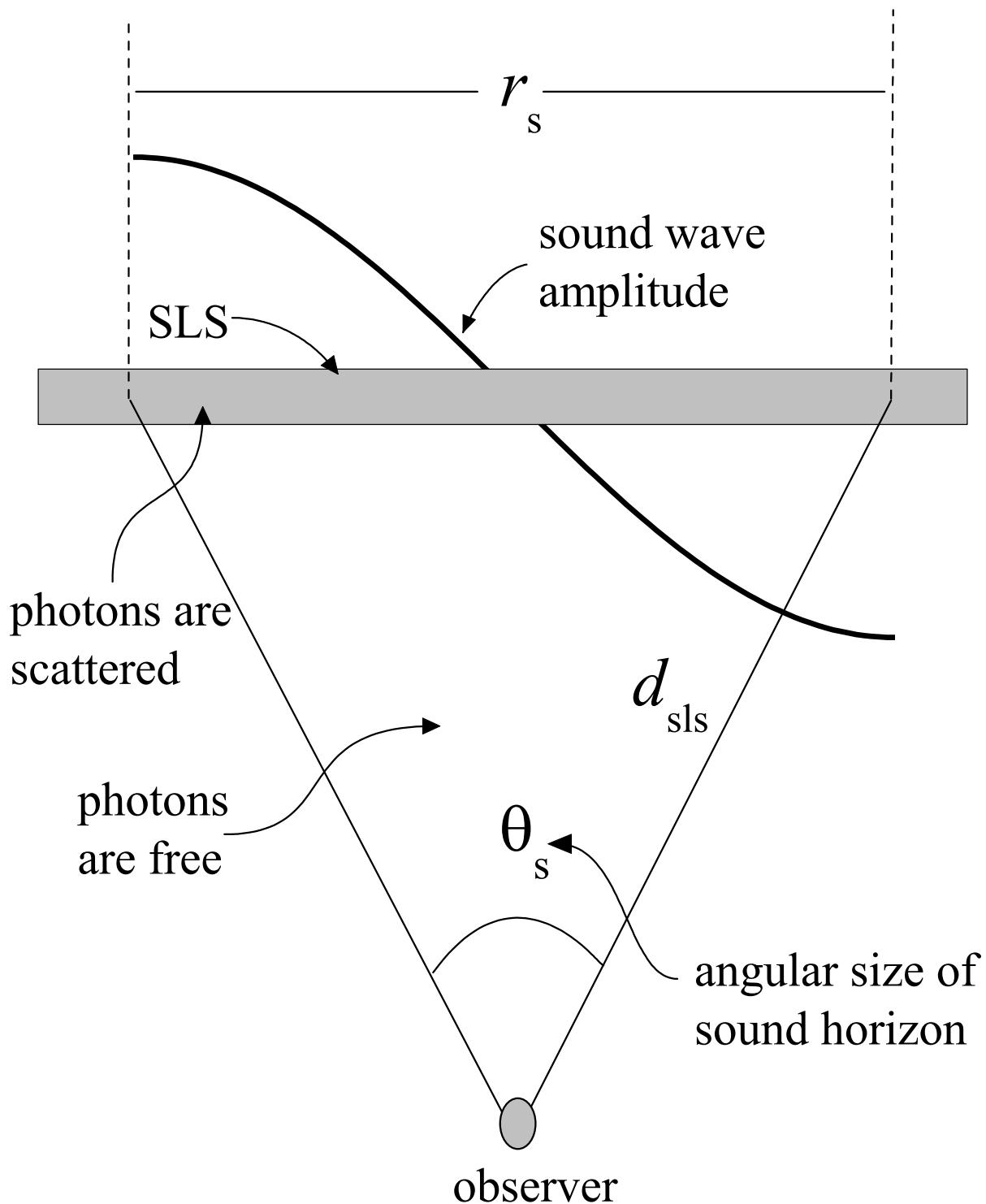


FIGURE 3. The surface of last scattering (SLS), the fundamental acoustic mode, and the sound horizon (r_s , θ_s). The photons of the CMB underwent Thomson scattering in the early universe and acoustic oscillation left their imprint at recombination.

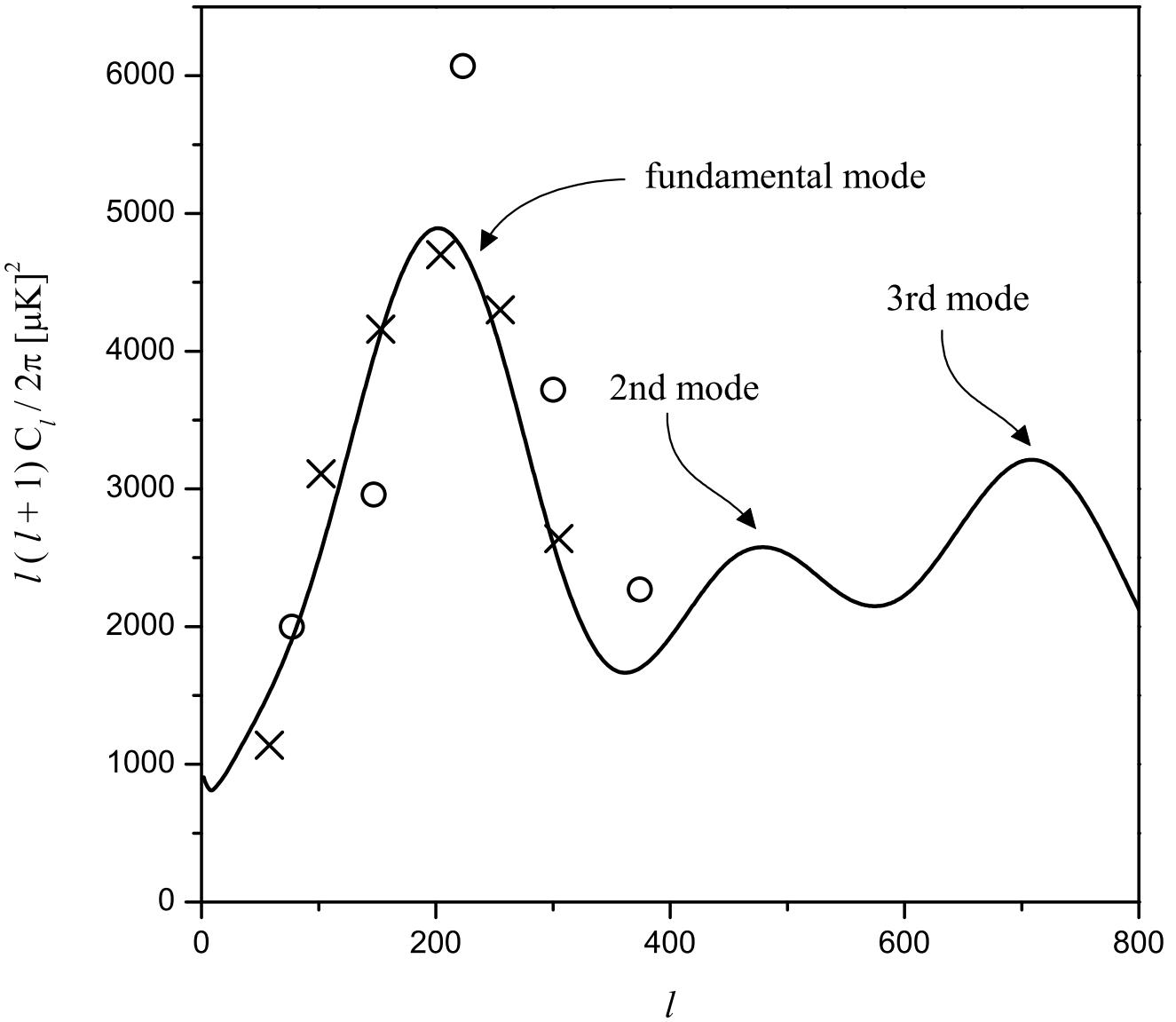


FIGURE 4. The power spectrum. The solid curve is a theoretical power spectrum calculated using CMBFAST [30]. The open circles are from Ref. 20 and the crosses are from Ref. 21. Notice that the first peak corresponding to the fundamental acoustic mode occurs near $l = 200$, signifying a flat universe.

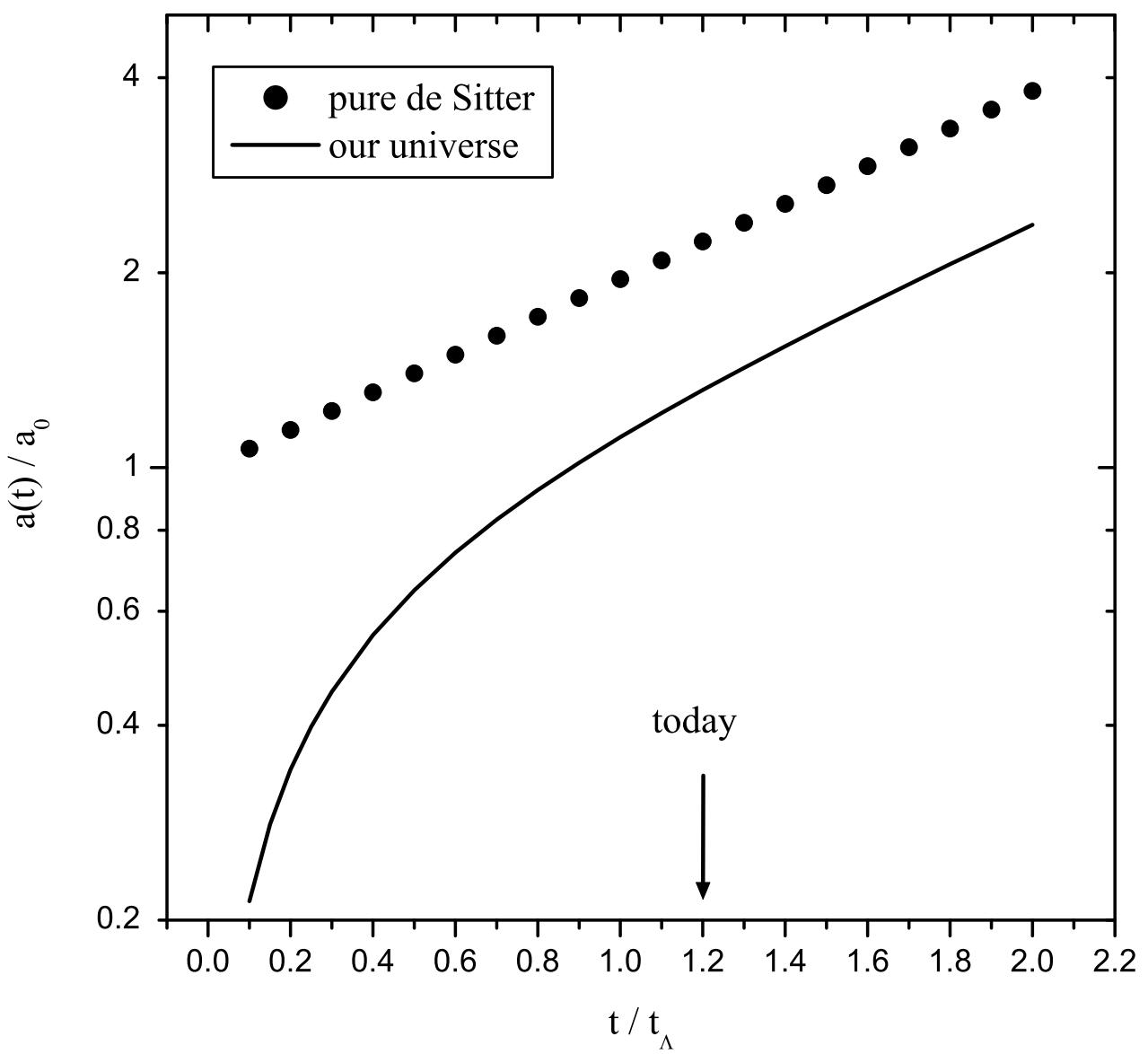


FIGURE 5. The cosmic scale factor for our universe compare to the de Sitter model.

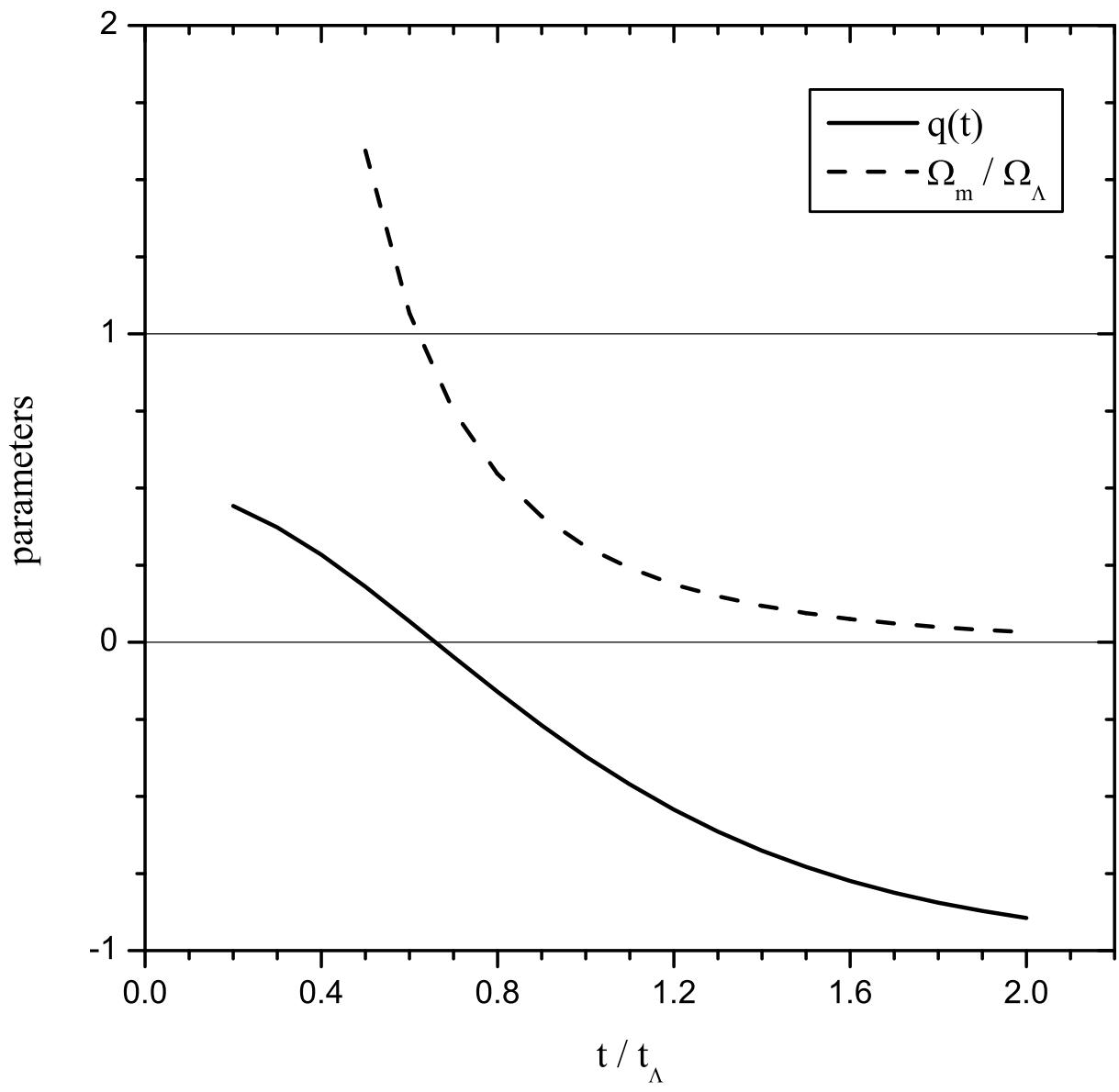


FIGURE 6. The deceleration parameter of our universe. The sign of $q(t)$ switches from positive to negative at around the same time that the universe goes from matter-dominated to Λ -dominated.

Gravitational lensing as a probe of structure

By PETER SCHNEIDER¹

¹Institut für Astrophysik und Extraterrestrische Forschung, Universität Bonn, Germany

Gravitational lensing has become one of the most interesting tools to study the mass distribution in the Universe. Since gravitational light deflection is independent of the nature and state of the matter, it is ideally suited to investigate the distribution of all (and thus also of dark) matter in the Universe. Lensing results have now become available over a wide range of scales, from the search for MACHOs in the Galactic halo, to the mass distribution in galaxies and clusters of galaxies, and the statistical properties of the large-scale matter distribution in the Universe. Here, after introducing the concepts of strong and weak lensing, several applications are outlined, from strong lensing by galaxies, to strong and weak lensing by clusters and the lensing properties of the large-scale structure.

1. Introduction

Light rays are deflected in gravitational fields, just like massive particles are. Hence, the deflection of light probes the gravitational field, and therefore the matter distribution that causes it. Since the field is independent of the state and nature of the matter generating it, it provides an ideal tool for studying the total (that is, luminous and dark) matter in cosmic objects. As we shall see, gravitational light deflection is used to study cosmic mass distributions on scales ranging from stars to galaxies, and from clusters of galaxies to the large-scale matter distribution in the Universe. In this contribution, I will concentrate on those aspects which are of particular relevance for learning about the dark matter distribution in the Universe.

Gravitational lensing describes phenomena of gravitational light deflection in the weak-field, small deflection limit; strong-field light deflection (important for light propagation near black holes or neutron stars) are not covered by gravitational lens (hereafter GL) theory. The basic theory of gravity, and of light propagation in a gravitational field is General Relativity, which says that photons travel along null geodesics of the spacetime metric (these are described by a second-order differential equation). In GL theory, several simplifications apply, owing to restriction to weak fields, and thus small deflections. We shall see the convenience of those further below.

Gravitational lensing as a whole, and several particular aspects of it, has been reviewed previously. Two extensive monographs (Schneider, Ehlers & Falco 1992, hereafter SEF; Petters, Levine & Wambsganss 2001, hereafter PLW) describe lensing in all depth, in particular providing a derivation of the gravitational lensing equations from General Relativity. Fort & Mellier (1994) describe the giant luminous arcs and arclets in clusters of galaxies (see Sect. 5.3), Paczyński (1996) and Roulet & Mollerach (1997) review the effects of gravitational microlensing in the Local Group, whereas the reviews by Narayan & Bartelmann (1999) and Wambsganss (1998) provide a concise and didactical account of GL theory and observations. Much of this contribution will be focused on weak gravitational lensing, which has been reviewed recently by Mellier (1999), Bartelmann & Schneider (2001), Wittman (2002), van Waerbeke & Mellier (2003) and Refregier (2003).

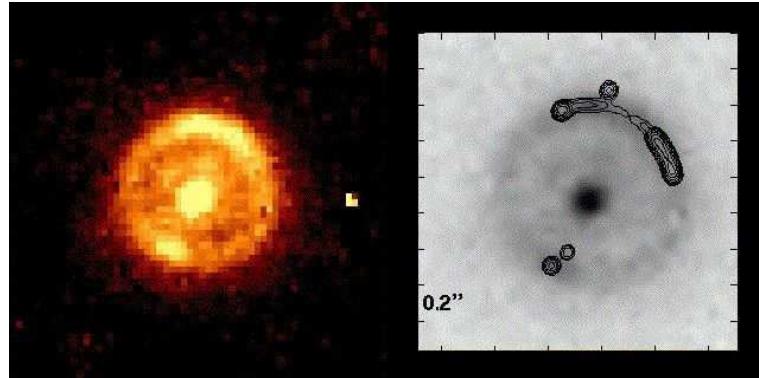


FIGURE 1. The radio source JVAS B1938+666 shows two radio sources (contours in the right panel), one of which is mapped into four components, the other shows a double image; furthermore, the outer radio contours merge into an arc around the lensing galaxy. The underlying grey-scale figure, and the left panel, shows a near-IR iamge of the field, revealing the lens galaxy, as well as the Einstein-ring image of the galaxy hosting the radio-AGN (from King et al. 1998)

2. Basics of gravitational lensing

2.1. Very brief history of lensing

The investigation of gravitational light deflection dates back more than 200 years to Mitchel, Cavendish, Laplace and Soldner (see SEF, PLW for references and much more detail). At that time a metric theory of gravity was not known, and light was treated as massive particles moving with the velocity of light. General Relativity, finalized in 1915, predicts a deflection angle twice as large as ‘Newtonian’ theory, and was verified in 1919 by measuring the deflection of light near the Solar limb during an eclipse. Soon after, the ‘lens effect’ was discussed by Lodge, Eddington and Chwolson, i.e. the possibility that light deflection leads to multiple images of sources behind mass concentrations, or even yields a ring-like image. Einstein, in 1936, considered in detail the lensing of a source by a star (or a point-mass lens), and concluded that the angular separation between the two images would be far too small (of order milliarcseconds) to be resolvable, so that “there is no great chance of observing this phenomenon”. In 1937, Zwicky, instead of looking at lensing by stars in our Galaxy, considered “extragalactic nebulae” (nowadays called galaxies) as lenses. He noted that they produce angular separations than can be separated with telescopes. Observing such an effect, he noted, would furnish an additional test of GR, would allow one to see galaxies at larger distances (due to the magnification effect), and to determine the masses of these nebulae acting as lenses. He furthermore considered the probability of such lens effects and concluded that about 1 out of 400 distant sources should be affected by lensing, and therefore predicted that “the probability that nebulae which act as gravitational lenses will be found becomes practically a certainty”. His visions were right on (nearly) all accounts.

In the mid-1960’s, Klimov, Liebes, and Refsdal independently formulated the basic theory of gravitational lensing, and focused on astrophysical applications, like determination of masses and cosmological parameters. For example, Refsdal noticed that the light travel time along the two rays corresponding to two images of a source, is proportional to the size of the Universe, thus to H_0^{-1} , and that a measurement of the time delay, possible if the source varies intrinsically, would allow the determination of the Hubble constant.

In 1979, the first GL system was discovered by Walsh, Carswell & Weymann: The

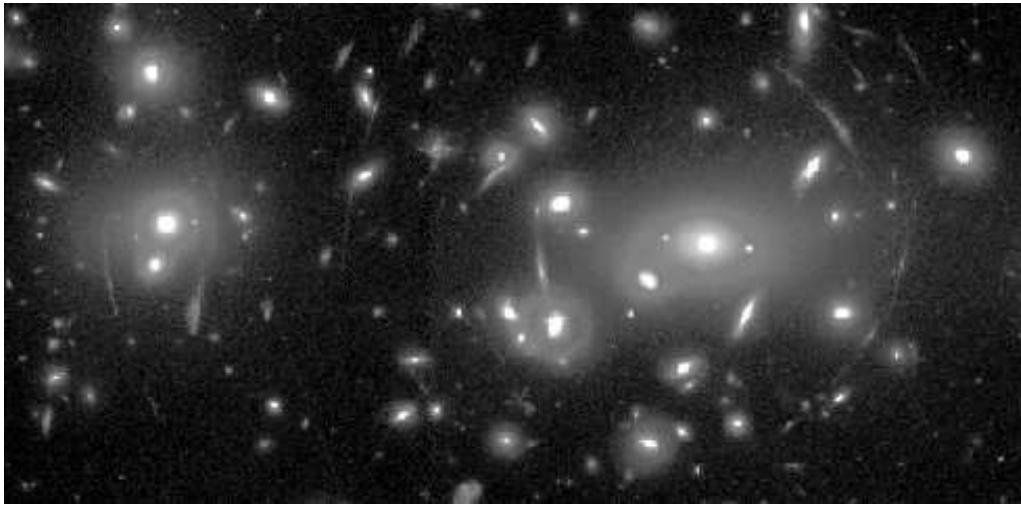


FIGURE 2. The redshift $z = 0.18$ cluster Abell 2218 displays an enormously rich structure of arcs, highly stretched images of background galaxies which curve around the main cluster center (seen to the right of the image center), but also around the secondary mass peak near the left edge of this WFPC2@HST image. Together with the identification of several multiply-imaged background sources, these lensing phenomena have yielded a very detailed mass map of the inner part of this cluster (Courtesy J.-P. Kneib)

two images of the QSO 0957+561 are separated by about 6 arcseconds, having identical colors, redshifts ($z_s = 1.41$) and spectra; both images are radio sources with a core-jet structure on milli-arcsec scales. Soon thereafter, a galaxy situated between the two quasar images was detected, with redshift $z_d = 0.36$, being member of a cluster. 1980 marks the discovery of the first GL system with four QSO images, QSO 1115+080, two of which are very closely spaced. 1986 saw the discovery of a new lensing phenomenon, which had been predicted long before: the detection of a radio ring, in which an extended radio source is mapped into a complete ring by an intervening galaxy (see also Fig. 1). Such *Einstein rings* turn out to yield the most accurate mass determinations in extragalactic astronomy. At present, some 72 multiple-image systems are known where the major lens component is a galaxy, including about 46 doubles, 20 four-image systems, but also one 3-image, one 5-image, and one 6-image system. The first of these were discovered serendipitously, but since the 1990's, large systematic searches for such systems were successfully conducted in the optical and, in particular, radio wavebands.

In 1986, a new lensing phenomenon was discovered by two independent groups: strongly elongated, curved features around two clusters of galaxies. Their extreme length-to-width ratios made them difficult to interpret; the measurement of the redshift of one of them placed the source of the arc at a distance well behind the corresponding cluster. Hence, these giant luminous arcs are images of background galaxies, highly distorted by the tidal gravitational field of the cluster. By now, many clusters with giant arcs are known and investigated in detail, for which in particular the high-resolution of the HST was essential. Less extremely distorted images of background galaxies have been named *arclets* and can be identified in many clusters.

If some sources are so highly distorted as those seen in Fig. 2, one expects to see many more sources which are distorted to a much smaller degree – such that they can not be identified individually as lensed images, but that nearby images are distorted in a similar way, so that the distortion can be identified statistically. This forms the basis of weak

lensing; coherent image distortions around massive clusters were detected in the early 1990's by Tyson and his group. As shown by Kaiser & Squires, there image distortions can be used to obtain a parameter-free mass map of clusters. Further weak lensing phenomena, such as galaxy-galaxy lensing, have been detected in the last decade; the weak lensing effect by the large-scale matter distribution in the Universe, the so-called cosmic shear, was discovered by several independent teams simultaneously in 2000, and this has opened up a new window in observational cosmology.

Last but not least, gravitational microlensing in the local group has been suggested in 1986 by Paczynski as a test of whether the dark matter in the halo of our Galaxy is made up of compact objects; the first microlensing events were discovered in 1993 by three different groups. I refer to the lectures of Prof. Sadoulet for a discussion of microlensing and the results concerning the dark matter in our Milky Way.

2.2. Deflection angle and lens equation

We shall provide here the basic lensing relations; the reader is encouraged to refer to one of the reviews or books listed in the introduction for a full derivation of these relations.

2.2.1. Deflection by a ‘point mass’ M

Consider the deflection of a light ray by the exterior of a spherically symmetric mass M ; from the Schwarzschild metric one finds that a ray with impact parameter ξ is deflected by an angle

$$\hat{\alpha} = \frac{4GM}{c^2 \xi} = \frac{2R_s}{\xi} : \text{Einstein deflection angle; } R_s: \text{Schwarzschild radius} \quad (2.1)$$

valid for $R_s/\xi \ll 1$, or $\hat{\alpha} \ll 1$; note that this also implies that $\Phi/c^2 \ll 1$, where Φ is the Newtonian gravitational potential. The value for $\hat{\alpha}$ is twice the ‘Newtonian’ value derived by Soldner and others and was verified during the Solar Eclipse 1919!

2.2.2. Deflection by a mass distribution

Since the field equations of General Relativity can be linearized if the gravitational field is weak, the deflection angle caused by an extended mass distribution can be calculated as the (vectorial) sum of the deflections due to its individual mass elements. If the deflection angle is small (which is implied by the weak-field assumption), the light ray near the mass distribution will deviate only slightly from the straight, undeflected ray. In this (‘Born’) approximation, valid if the extent of the mass distribution is much smaller than the distances between source, lens, and observer (the ‘geometrically thin lens’), the deflection angle depends solely on the *surface mass density* $\Sigma(\xi)$, defined in terms of the volume density $\rho(\vec{r})$ as

$$\Sigma(\xi) \equiv \int d\vec{r}_3 \rho(\xi_1, \xi_2, r_3), \quad (2.2)$$

with the r_3 -direction along the line-of-sight. Superposing the deflections by the mass elements of the lens, one obtains the deflection angle

$$\hat{\alpha}(\xi) = \frac{4G}{c^2} \int d^2\xi' \Sigma(\xi') \frac{\xi - \xi'}{|\xi - \xi'|^2}. \quad (2.3)$$

The geometrically-thin condition is satisfied in virtually all astrophysically relevant situations (i.e. lensing by galaxies and clusters of galaxies), unless the deflecting mass extends all the way from the source to the observer, as in the case of lensing by the large-scale structure. The relevant deflections are small, e.g., $\hat{\alpha} \lesssim 1''$ for galaxies, $\hat{\alpha} \lesssim 30''$ for clusters.

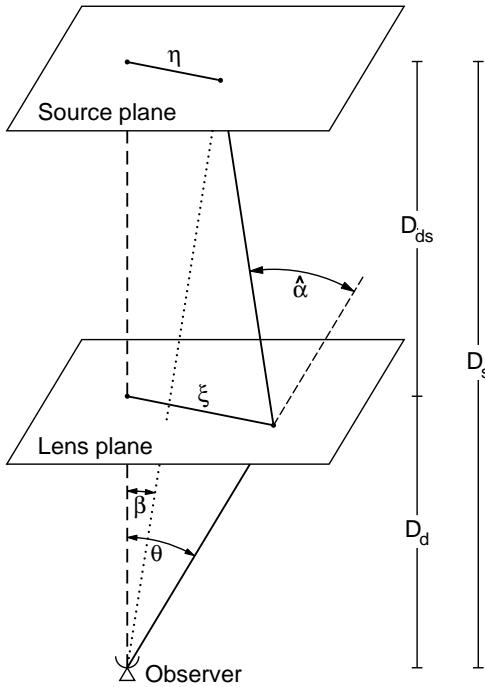


FIGURE 3. Sketch of a typical gravitational lens system.

2.2.3. The lens equation

The lens equation relates the true position of the source to its observed position; we define the lens and source plane as planes perpendicular to the line-of-sight to the deflector, at the distance D_d and D_s of the lens and the source, respectively (see Fig. 3). Furthermore, we define the ‘optical axis’ as a ‘straight’ line through the lens center (the exact definition does not matter; any change of it represents just an unobservable translation in the source plane), and its intersections with the lens and source planes as their respective origins. Denoting ξ as the two-dimensional position of the light ray in the lens plane and η as the position of the source (see Fig. 3), then from geometry,

$$\eta = \frac{D_s}{D_d} \xi - D_{ds} \hat{\alpha}(\xi) . \quad (2.4)$$

Note that the distances D occurring here are the angular-diameter distances, since they relate physical transverse separations to angles. If θ denotes the angle of a light ray relative to the optical axis, β as the angular position of the unlensed source (see Fig. 3),

$$\eta = D_s \beta ; \quad \xi = D_d \theta , \quad (2.5)$$

then

$$\boxed{\beta = \theta - \frac{D_{ds}}{D_s} \hat{\alpha}(D_d \theta) \equiv \theta - \alpha(\theta)} , \quad (2.6)$$

where $\alpha(\theta)$ is the *scaled deflection angle*, which in terms of the *dimensionless surface mass density*

$$\kappa(\theta) := \frac{\Sigma(D_d \theta)}{\Sigma_{\text{cr}}} \quad \text{with} \quad \boxed{\Sigma_{\text{cr}} = \frac{c^2}{4\pi G} \frac{D_s}{D_d D_{ds}}} , \quad (2.7)$$

reads as

$$\alpha(\theta) = \frac{1}{\pi} \int_{\mathbb{R}^2} d^2\theta' \kappa(\theta') \frac{\theta - \theta'}{|\theta - \theta'|^2}. \quad (2.8)$$

Note that the *critical surface mass density* Σ_{cr} depends only on the distances. Lenses with $\kappa \sim 1$ at some points are called *strong lenses*, and those with $\kappa \ll 1$ everywhere are *weak lenses*. The lens equation $\beta = \theta - \alpha(\theta)$ is a mapping $\theta \rightarrow \beta$ from the lens plane to the source plane; but in general, this mapping is non-invertible: for a given source position β , the lens equation can have multiple solutions θ which correspond to multiple images of a source at β .

2.2.4. Deflection and Fermat potentials

Since $\nabla \ln |\theta| = \theta / |\theta|^2$, the deflection angle can be written as

$$\alpha = \nabla \psi, \quad \text{with} \quad \psi(\theta) = \frac{1}{\pi} \int_{\mathbb{R}^2} d^2\theta' \kappa(\theta') \ln |\theta - \theta'|; \quad (2.9)$$

being the *deflection potential*; hence, the lens equation describes a gradient mapping. From $\nabla^2 \ln |\theta| = 2\pi\delta_D(\theta)$, where δ_D denotes Dirac's delta-'function', one finds the 2-D Poisson equation

$$\nabla^2 \psi = 2\kappa. \quad (2.10)$$

Defining the *Fermat potential*

$$\phi(\theta; \beta) = \frac{1}{2} (\theta - \beta)^2 - \psi(\theta), \quad (2.11)$$

where β enters as a parameter, one sees that the lens equation can be written as

$$\nabla \phi(\theta; \beta) = \mathbf{0}. \quad (2.12)$$

Solutions of (2.12) can then be classified, according to whether the potential ϕ has a minimum, maximum, or saddle point at the solution point θ .

2.3. Effects of lensing

2.3.1. Multiple images

Multiple images correspond to multiple solutions θ of the lens equation for fixed source position β . For the case of a point-mass lens,

$$\hat{\alpha}(\xi) = \frac{4GM}{c^2} \frac{\xi}{|\xi|^2}, \quad \beta = \theta - \frac{4GMD_{\text{ds}}}{c^2 D_{\text{d}} D_{\text{s}}} \frac{\theta}{|\theta|^2} = \theta - \theta_E^2 \frac{\theta}{|\theta|^2}, \quad (2.13)$$

where θ_E is the *Einstein angle* $\propto \sqrt{M}$. Note that (2.13) can also be obtained from (2.3) and (2.6) by setting $\Sigma(\xi) = M \delta_D(\xi)$. The lens equation has two solutions, one on either side of the lens (just solve the quadratic equation for θ), with image separation $\Delta\theta \gtrsim 2\theta_E \propto \sqrt{M}$ [†]. Hence, the image separation yields an estimate for the mass of the lens. In general, however, more complicated mass models are needed to fit the observed image positions in a gravitational lens system, i.e., to find a mass model and a source position such that $\beta = \theta_i - \alpha(\theta_i)$ is satisfied for all images θ_i .

2.3.2. Magnification

Gravitational light deflection conserves surface brightness; this follows from Liouville's theorem, noting that light deflection is not associated with emission or absorption pro-

[†] Mathematically, substantially larger separations can occur if $|\beta| \gg \theta_E$, but this case is astronomically irrelevant, as explained shortly.

cesses. Therefore, $I(\boldsymbol{\theta}) = I^{(s)}[\boldsymbol{\beta}(\boldsymbol{\theta})]$, where $I(\boldsymbol{\theta})$ and $I^{(s)}(\boldsymbol{\beta})$ denote the surface brightness in the image and source plane. Differential light bending causes light bundles to get distorted; for very small light bundles, the distortion is described by the Jacobian matrix

$$\mathcal{A}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \delta_{ij} - \frac{\partial^2 \psi(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \end{pmatrix} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}, \quad (2.14)$$

where

$$\gamma_1 = \frac{1}{2}(\psi_{,11} - \psi_{,22}), \quad \gamma_2 = \psi_{,12} \quad (2.15)$$

are the two Cartesian components of the shear (or the tidal gravitational force). For a small source centered on $\boldsymbol{\beta}_0 = \boldsymbol{\theta}_0 - \boldsymbol{\alpha}(\boldsymbol{\theta}_0)$:

$$I(\boldsymbol{\theta}) = I^{(s)} [\boldsymbol{\beta}_0 + \mathcal{A}(\boldsymbol{\theta}_0) \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}_0)]. \quad (2.16)$$

Hence, the image of a small circular source with radius r is an ellipse with semi-axes $\lambda_{1,2} r$ where $\lambda_{1,2}$ are the eigenvalues of \mathcal{A} ; the orientation of the ellipse is determined by the shear components $\gamma_{1,2}$.

The area distortion by differential deflection yields a magnification (since I is unchanged, and flux = $I \times$ solid angle),

$$\boxed{\mu = \frac{S}{S_0} = \frac{1}{\det \mathcal{A}} = \frac{1}{(1 - \kappa)^2 - |\gamma|^2}}, \quad (2.17)$$

with $|\gamma| = \sqrt{\gamma_1^2 + \gamma_2^2}$. Since \mathcal{A} is different for different multiple images, the image fluxes are different; the observed flux ratios yield the image magnification ratios. In particular, if the image separation in a point mass lens system is substantially larger than $2\theta_E$, which occurs for $|\boldsymbol{\beta}| \gg \theta_E$, the secondary image is very strongly demagnified and thus invisible. Flux ratios can in principle be used to constrain lens models, in addition to the image positions, but the magnifications are affected by small-scale structure in the mass distribution (we shall come back to this point below), rendering them less useful in mass model determinations. Note that the final expression in (2.17) can have either sign; images with $\mu > 0$ have positive parity, those with $\mu < 0$ negative parity. Positive parity images correspond to extrema of the Fermat potential ϕ , negative parity images correspond to saddle points of ϕ . In the rest of this article, we will always mean the absolute value of the magnification when writing μ .

2.3.3. Shape distortions

The image shape of extended (resolved) sources is changed by lensing, since the eigenvalues of \mathcal{A} are different in general; rewriting

$$\boxed{\mathcal{A}(\boldsymbol{\theta}) = (1 - \kappa) \begin{pmatrix} 1 - g_1 & -g_2 \\ -g_2 & 1 + g_1 \end{pmatrix}, \quad g_i = \frac{\gamma_i}{(1 - \kappa)}} : \text{reduced shear}, \quad (2.18)$$

one sees that the shape distortion is determined by the reduced shear. This in fact forms the basis of *weak lensing*. It should be noted that giant arcs cannot be described by the linearized lens equation, as they are too big.

2.3.4. Time delay

The light travel time along the various ray paths corresponding to different images is different in general. This implies that variations of the source luminosity will show up as flux variations of the different images at different times, shifted by a time delay Δt ,

$$\Delta t = \frac{D_d D_s}{c D_{ds}} (1 + z_d) \left[\phi(\boldsymbol{\theta}^{(1)}; \boldsymbol{\beta}) - \phi(\boldsymbol{\theta}^{(2)}; \boldsymbol{\beta}) \right], \quad (2.19)$$

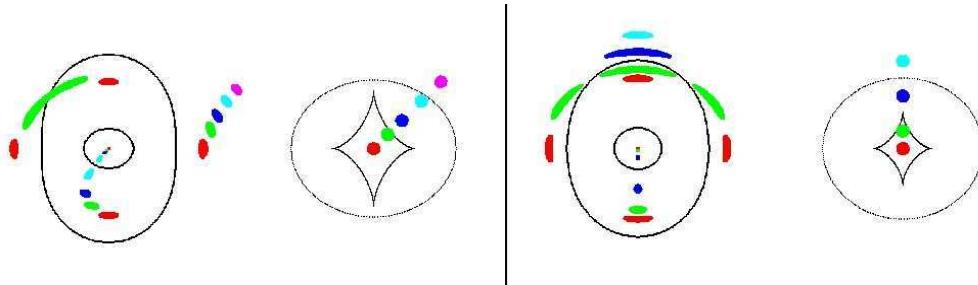


FIGURE 4. Illustration of critical curves and caustics for an elliptical mass distribution, and the image geometry for various source positions (taken from Narayan & Bartelmann 1999). In each of the two panels, the right figure shows the caustic curves, together with several different source positions; the corresponding image positions are displayed on the left, together with the critical curves. One sees that, depending on where the source is located relative to the caustics, different image multiplicities occur

where $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$ are the two image positions considered. In fact, $\phi(\boldsymbol{\theta}; \boldsymbol{\beta})$ is, up to an affine transformation, the light travel time along a ray from the source at $\boldsymbol{\beta}$ which crosses the lens plane at $\boldsymbol{\theta}$. Recalling that $\nabla\phi(\boldsymbol{\theta}; \boldsymbol{\beta}) = 0$ was equivalent to the lens equation, we see from (2.19) that this is Fermat's principle in lensing: the light travel time is stationary at physical images. Note that $\Delta t \propto H_0^{-1}$, since all the distances D are $\propto c/H_0$; hence, a measurement of the time delay can be used to determine the Hubble constant, provided the lens model is sufficiently well known. We shall return to this issue below.

2.3.5. General properties of lenses

If $\Sigma(\boldsymbol{\xi})$ is a smooth function, then for (nearly) every source position $\boldsymbol{\beta}$, the number of images is odd ('odd-number theorem'; Burke 1981). If in addition, $\Sigma(\boldsymbol{\xi})$ is non-negative, then at least one of the images (corresponding to a minimum in light travel time) is magnified, $\mu \geq 1$ ('magnification theorem'; Schneider 1984). The odd-number theorem is violated observationally: one (usually) finds doubles and quads. The missing odd image is expected to be close to the center of the lens, where $\kappa \gg 1$ presumably, meaning that $\mu \ll 1$; hence, this central image is highly demagnified, and thus not observable. Both of these theorems can be generalized even to non-thin lenses (Seitz & Schneider 1992).

The closed and smooth curves where $\det \mathcal{A}(\boldsymbol{\theta}) = 0$ are called *critical curves*. When they are mapped back into the source plane using the lens equation, the corresponding curves in the source plane are called *caustics*. The number of images changes by ± 2 if the source position crosses a caustic; then two images appear or merge. A source close to, and at the inner side of, a caustic produces two closely separated and very bright images near the corresponding critical curve. A source close to, and on the inner side of a cusp has three bright images close to the corresponding point on the critical curve. From singularity theory, one finds that in the limit of very large magnifications, the two close images on either side of the critical curve have equal magnification, and thus should appear equally bright. Similarly, of the three images formed near a cusp, the sum of the magnifications of the outer two images should equal that of the middle image, with corresponding consequences for the flux ratios. As we shall see, these universality relations are strongly violated in observed lens systems, providing a strong clue for the presence of substructure in the mass distribution of lens galaxies.

3. (Strong) Lensing by galaxies

The first lensing phenomena detected were multiple images of distant QSOs caused by the lensing effect of a foreground galaxy. If the lens is a massive (i.e., $\sim L_*$) galaxy, the corresponding image separations are $\sim 1''$. Gravitational lens models can be used to constrain the mass distribution in (the inner part of) these lensing galaxies; in particular, the mass inside a circle traced by the multiple images (or the Einstein ring) can be determined with *very* high precision in some cases. Furthermore, as already mentioned, time-delays can be used to determine H_0 . Mass substructure in these galaxies can be (and has been) detected, and the interstellar medium of lens galaxies can be investigated. We shall describe some of these techniques and results in a bit more detail below.

3.1. Mass determination

To obtain accurate mass estimates, one needs detailed models, obtained by fitting images and galaxy positions (and fluxes). However, even without these detailed models, a simple mass estimate is possible: the mean surface mass density inside the Einstein radius θ_E of a lens is the critical surface mass density, so that

$$\boxed{M(\theta_E) = \pi(D_d\theta_E)^2 \Sigma_{\text{cr}}}. \quad (3.1)$$

An estimate of θ_E is obtained as the radius of the circle tracing the multiple images (or the ring radius in case of Einstein ring images). The estimate (3.1) is exact for axi-symmetric lenses, and also a very good approximation for less symmetrical ones.

3.2. Mass models

The simplest mass model for a galaxy is that of a singular isothermal sphere (SIS), which is an analytic solution of the Vlasov–Poisson equation of stellar dynamics (see Binney & Tremaine 1987) and whose density profile behaves like $\rho(r) \propto r^{-2}$, so that the surface mass density is given by

$$\Sigma(\xi) = \frac{\sigma_v^2}{2G\xi}; \quad \text{with } \sigma_v : \text{1-D velocity dispersion.} \quad (3.2)$$

This model is often good enough for rough estimates, in particular since the inner parts of the radial mass profile of galaxies seem to closely follow this relation. Multiple images occur for $\beta < \theta_E$, and their separation is $\Delta\theta = 2\theta_E$, with

$$\theta_E = 4\pi \left(\frac{\sigma_v}{c} \right)^2 \left(\frac{D_{\text{ds}}}{D_s} \right) \approx 1''.15 \left(\frac{\sigma_v}{200 \text{ km/s}} \right)^2 \left(\frac{D_{\text{ds}}}{D_s} \right). \quad (3.3)$$

Hence, massive ellipticals create image separations of up to $\sim 3''$, whereas for less massive ones, and for spirals, $\Delta\theta$ is of order or below $\sim 1''$.

However, this simple mass model is unrealistic owing to its diverging density for $r \rightarrow 0$ and its infinite mass; furthermore, it (like all axi-symmetric models) cannot account for the occurrence of quadruply imaged sources. More complicated models include some or all of these:

- A finite core size, to remove the central divergence. Applied to observed systems, the models ‘like’ the core to be very small, in particular since the third or fifth image is not seen, which needs high demagnification $\mu \ll 1$ and thus high κ near the center.
- Elliptical isodensity contours that break the axial symmetry, which is needed to explain 4-image systems; those cannot be produced by symmetric lens models.
- External tidal field (shear): A lens galaxy is not isolated, but may be part of a group or a cluster, and in any case the inhomogeneous matter distribution between source and lens, and lens and observer introduces a shear (cosmic shear) of typically a few percent.

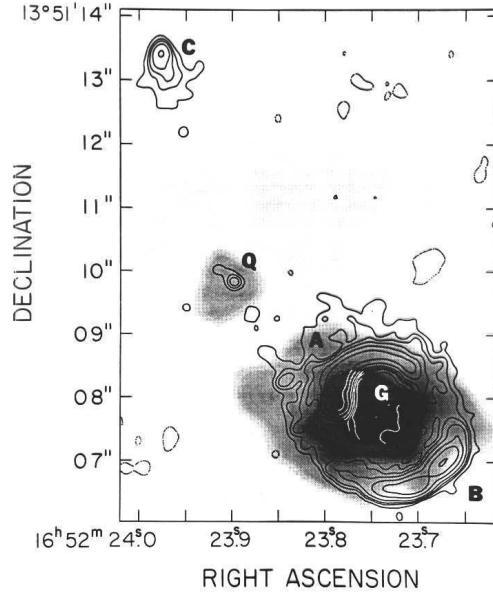


FIGURE 5. The Southern radio lobe (the Northern one is denoted by C) of the quasar MG 1654+134 ($z = 1.7$, central component Q coincident with the optical QSO position), shown here as contours, is mapped into a complete Einstein ring by a foreground galaxy (G), indicated by the grayscale image. For this lens system, probably the most detailed mass analysis was performed – see text (image is from Langston et al. 1990)

This external influence can be linearized over the region of the galaxy where multiple images occur, and this yields a uniform shear term in the lens equation.

In fact, any realistic model of a lens consists of at least an elliptical mass distribution and an external shear. This then yields the necessary number of free parameters for a lens model: 1 for the mass scale of the lens (either the Einstein radius, or σ_v), 2 for the lens position, 2 for the source position, 2 for the lens ellipticity (axis ratio and orientation), and 2 for the external shear (a two-component quantity). This can be compared to the number of observables. In a quad-system, one has 2×4 image positions, and the 2 coordinates of the lens galaxy. In this case, the number of observational constraints is larger by one than the number of free parameters. In addition, one could use the flux ratios of the images (i.e., the magnification ratios) to constrain the lens model, but as we shall discuss below, these are not reliable constraints for fitting a macro-model.

Modeling of strong lens systems which contain Einstein rings yield much better constrained lens parameters and therefore even more accurate mass estimates; a beautiful example of this is the radio ring MG1645+134 (see Kochanek 1995; Wallington et al. 1996) around a foreground elliptical galaxy at $z = 0.25$ shown in Fig. 5.

Results from fitting parametrized mass models to observed multiple image systems include the following: Many lens systems require quite a strong external shear, which may be explained by massive ellipticals being preferentially located in dense regions, i.e. in clusters or groups. The orientation of the mass ellipticity follows closely that of light distribution; however, this is not the case for the magnitude of ellipticity (Keeton et al. 1998). From mass modelling and detailed spectroscopic studies of lens galaxies, the latter in combination with stellar dynamical arguments, one finds that inside the Einstein radius, about half the mass is dark, and half is baryonic (Treu & Koopmans 2002;

Koopmans & Treu 2003); therefore, in massive (lens) galaxies, baryons have strongly affected the mass profile, owing to their cooling and contraction. These authors have also shown that over the radius relevant for lensing, the profile is very well approximated by an isothermal one.

This issue is closely related to the determination of the Hubble constant from measuring time delays in multiple image systems. At present, time delays are known for about 10 lens systems, in some cases with an accuracy of 1%. Therefore, H_0 can in principle be obtained by (2.19). However, reality turns out to be considerably more difficult. The major difficulty lies in the so-called mass-sheet degeneracy, which says that the transformation $\kappa(\theta) \rightarrow \lambda\kappa(\theta) + (1 - \lambda)$ of the surface mass density leaves the image positions and magnification ratios invariant, but changes the time delay by a factor λ (Falco et al. 1985). Essentially, the bracket in (2.19) depends on the mean surface mass density within the annular region around the lens in which the multiple images are located (Kochanek 2002), and the mass-sheet degeneracy changes that value. One therefore requires additional information about the mass profile in galaxies.

As mentioned above, an isothermal profile ($\kappa \propto \theta^{-1}$) provides a reasonable fit to lens systems. Assuming an isothermal profile yields values of the Hubble constant of order $H_0 \approx 50$ km/s/Mpc, consistently for the ‘simple’ lens systems. This is at variance with the value $H_0 \approx 72$ km/s/Mpc obtained from the Hubble Key Project (Freedman et al. 2001). On the other hand, the isothermal profile is at best a reasonable approximation to the real mass profile. Cosmological simulations yield a cuspy profile, such as the one found by Navarro et al. (1997; hereafter NFW). These dark matter profiles are then modified by baryons cooling inside these halos; the larger the baryon fraction, the more are the dark matter profiles affected. Kochanek (2003) pointed out that in order to get a value of H_0 from lensing which is compatible with that from the Hubble Key Project, one would need a baryon fraction as large as 20% of the total dark matter in the halo to cool, in order to render the central mass profiles of lenses steep enough; in effect, that leads to constant M/L-models within the region where the multiple images are found. This high fraction of cold baryons in galaxies is at odds with the local inventories of baryonic mass in galaxies. At present, the origin of this discrepancy is not known.

3.3. Substructure in lens galaxies

Whereas ‘simple’ lens models can usually fit the image positions *very* well, in most lens systems they are unable to provide a good fit to the flux ratios. The best known (but by far not worst) case is QSO 1422+231, where several groups have tried, and failed, to obtain a good lens model explaining image positions and fluxes. Mao & Schneider (1998) have provided an analytical argument, based on the universality of the lens mapping near cusps, why one would not expect to find a smooth model for this system. Since the flux ratios in this system are most reliably and accurately measured in the radio, absorption by the ISM in the (elliptical) lens galaxy is expected to be negligible. We argued that small-scale structure in the mass distribution can change the magnification, but leave the image position essentially unchanged; this is due to the fact that the deflection angle depends on first partial derivatives of the deflection potential ψ , whereas the magnification depends on κ and γ , and thus on second derivatives of ψ ; those are more strongly affected by small-scale structure.

This effect has been known for a long time: the optical and UV radiation from QSOs comes from a region small enough that even stars in the lens galaxy can affect their magnification, whereas the corresponding deflections are of order 10^{-6} arcseconds – this *microlensing* phenomenon (see, e.g., Wambsganss et al. 1990) shows up as uncorrelated brightness variations in the multiple images, and has clearly been detected in the QSO

2237+0305 (Schmidt et al. 2002) and in some of the other lens systems. However, the VLBI images of QSO 1422+231 are extended, and therefore individual stars cannot affect their magnification. But massive structures with $M \gtrsim 10^7 M_\odot$ can change their magnification. Recall that CDM models of structure formation actually predict the presence of sub-halos in each massive galaxy – the (missing, since unobserved) satellites. As shown by Dalal & Kochanek (2002; and references therein), the statistics of mismatches between observed flux ratios and those predicted by simple lens models which fit the image positions in 4-image systems is in agreement with expectations from CDM satellites. Bradac et al. (2002) have explicitly demonstrated that the flux ratio problem in QSO 1422+231 is cured by placing a low-mass halo near one of the QSO images. Kochanek & Dalal (2003) have considered, and ruled out, alternative explanations for the flux ratio mismatches, such as interstellar scattering. One of the signatures of mass substructure, first found in investigations of microlensing (Schechter & Wambsganss 2002), is that the brightest saddle point is expected to be affected most, in the sense that it has a high probability of being substantially demagnified. Kochanek & Dalal have shown that this particular behavior is seen in a sample of 7 quadruple image lens systems. This behavior cannot be explained by absorption, scattering or scintillation by the interstellar medium of the lens galaxy. Hence, lensing has probably detected the ‘missing’ satellites in galaxy halos; the observed flux mismatches require a mass fraction in subclumps of order a few percent of the total lens mass, in accordance with predictions from CDM simulations. Bradac et al. (2002, 2003) have generated synthetic lens systems, using model galaxies as obtained from CDM simulations as deflectors, and have shown that the resulting image fluxes are at variance with the predictions from simple lens models, again due to the substructure in the mass distribution. In a few of the observed lens systems, a fairly massive subclump can be identified directly by its luminosity, yielding further support to this interpretation.

3.4. Other properties of lens galaxies

3.4.1. Evolution

Early-type galaxies are known to be located on the so-called fundamental plane (FP), i.e., there is a relation between their central surface brightness, their effective radius and the velocity dispersion in these galaxies. The FP has been observed even to high redshifts, using early-type galaxies in high-redshift clusters; it is known to evolve with z , mainly due to passive evolution of the stellar population. The lens galaxies form a mass-selected sample of galaxies not selected for cluster membership, and it is therefore of great interest to see whether they also obey a FP relation. In fact, since lens galaxies have a well-determined mass scale (or σ_v , after fitting an isothermal mass model), and are spread over a large redshift range, they are ideal for FP research. Rusin et al. (2003) have found from a sample of 28 lenses that the evolution of the FP is compatible with passive stellar evolution, and that it favours a high redshift for the epoch of star formation in these galaxies.

3.4.2. The interstellar matter in lens galaxies

Multiple image systems provide us with views of the same source along different lines-of-sight. Excluding time-delay effects in connection with spectral variability, as well as differential magnification, spectral differences between the images can then only be caused by propagation effects. In particular, one can study the properties of the dust in lens galaxies, as color differences between images can be attributed to different extinction and reddening along the different lines-of-sight through the lens galaxy. Falco et al. (1999) have investigated 23 gravitational lens galaxies over the range $0 \lesssim z_d \lesssim 1$. Given that

most lens galaxies are early types, they found a small median differential extinction of $\Delta E(B-V) \sim 0.05$, with slightly larger (smaller) values for radio- (optically-)selected lens systems. The lack of a clear correlation with the separation of the image from the lens center points towards patchy extinction. Two spiral lens galaxies show a substantially larger extinction. The extinction law, i.e., the relation between extinction and reddening, varies between different lens galaxies over quite some range; the Galactic extinction law is therefore by no means universally applicable.

4. Weak gravitational lensing

Multiple images, microlensing (with appreciable magnifications) and arcs in clusters are phenomena of *strong lensing*. In *weak gravitational lensing*, the Jacobi matrix \mathcal{A} is very close to the unit matrix, which implies weak distortions and small magnifications. Those cannot be identified in individual sources, but only in a statistical sense; the basics of these effects will be described in this section, and several applications will be discussed in later sections.

4.1. Distortion of faint galaxy images

Images of distant, extended sources are distorted in shape and size; this is described by the locally linearized lens equation around the image center $\boldsymbol{\theta}_0$,

$$\boldsymbol{\beta} - \boldsymbol{\beta}_0 = \mathcal{A}(\boldsymbol{\theta}_0) \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}_0), \quad (4.1)$$

where $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(\boldsymbol{\theta}_0)$, with the Jacobian (2.18), and the invariance of surface brightness (2.16). Recall that the shape distortion is described by the (reduced) shear which is a two-component quantity, most conveniently written as a complex number,

$$\gamma = \gamma_1 + i\gamma_2 = |\gamma| e^{2i\varphi}; \quad g = g_1 + ig_2 = |g| e^{2i\varphi}; \quad (4.2)$$

its amplitude describes the degree of distortion, whereas its phase φ yields the direction of distortion. The reason for the factor ‘2’ in the phase is the fact that an ellipse transforms into itself after a rotation by 180° . Consider a circular source with radius r ; mapped by the local Jacobi matrix, its image is an ellipse, with semi-axes

$$\frac{r}{1 - \kappa - |\gamma|} = \frac{r}{(1 - \kappa)(1 - |g|)} \quad ; \quad \frac{r}{1 - \kappa + |\gamma|} = \frac{r}{(1 - \kappa)(1 + |g|)}$$

and the major axis encloses an angle φ with the positive θ_1 -axis. Hence, if circular sources could be identified, the measured image ellipticities would immediately yield the value of the reduced shear, through the axis ratio

$$|g| = \frac{1 - b/a}{1 + b/a} \quad \Leftrightarrow \quad \frac{b}{a} = \frac{1 - |g|}{1 + |g|}$$

and the orientation of the major axis φ . However, faint galaxies are not intrinsically round, so that the observed image ellipticity is a combination of intrinsic ellipticity and shear. The strategy to nevertheless obtain an estimate of the (reduced) shear consists in locally averaging over many galaxy images, assuming that the intrinsic ellipticities are *randomly oriented*. In order to follow this strategy, one needs to clarify first how to define ‘ellipticity’ for a source with arbitrary isophotes (faint galaxies are not simply elliptical); in addition, seeing caused by atmospheric turbulence will blur – and thus circularize – observed images. We will consider these issues in turn.

4.2. Measurements of shapes and shear

Let $I(\boldsymbol{\theta})$ be the brightness distribution of an image, assumed to be isolated on the sky; the center of the image can be defined as

$$\bar{\boldsymbol{\theta}} \equiv \frac{\int d^2\theta q_I[I(\boldsymbol{\theta})] \boldsymbol{\theta}}{\int d^2\theta q_I[I(\boldsymbol{\theta})]} , \quad (4.3)$$

where $q_I(I)$ is a suitably chosen weight function; e.g., if $q_I(I) = I H(I - I_{\text{th}})$, $\bar{\boldsymbol{\theta}}$ would be the center of light within a limiting isophote of the image (where H denotes the Heaviside step function). We next define the tensor of second brightness moments,

$$Q_{ij} = \frac{\int d^2\theta q_I[I(\boldsymbol{\theta})] (\theta_i - \bar{\theta}_i)(\theta_j - \bar{\theta}_j)}{\int d^2\theta q_I[I(\boldsymbol{\theta})]} , \quad i, j \in \{1, 2\} . \quad (4.4)$$

Note that for an image with circular isophotes, $Q_{11} = Q_{22}$, and $Q_{12} = 0$. The trace of Q describes the size of the image, whereas the traceless part of Q_{ij} contains the ellipticity information.

From Q_{ij} , one defines two complex ellipticities,

$$\boxed{\chi \equiv \frac{Q_{11} - Q_{22} + 2iQ_{12}}{Q_{11} + Q_{22}}} \quad \text{and} \quad \boxed{\epsilon \equiv \frac{Q_{11} - Q_{22} + 2iQ_{12}}{Q_{11} + Q_{22} + 2(Q_{11}Q_{22} - Q_{12}^2)^{1/2}}} . \quad (4.5)$$

Both of them have the same phase (because of the same numerator), but a different absolute value; for an image with elliptical isophotes of axis ratio $r \leq 1$, one obtains

$$|\chi| = \frac{1 - r^2}{1 + r^2} \quad ; \quad |\epsilon| = \frac{1 - r}{1 + r} . \quad (4.6)$$

Which of these two definitions is more convenient depends on the context; one can easily transform one into the other,

$$\epsilon = \frac{\chi}{1 + (1 - |\chi|^2)^{1/2}} , \quad \chi = \frac{2\epsilon}{1 + |\epsilon|^2} . \quad (4.7)$$

4.2.1. From source to image ellipticities

In total analogy, one defines the second-order brightness tensor $Q_{ij}^{(s)}$, and the complex ellipticities $\chi^{(s)}$ and $\epsilon^{(s)}$ for the unlensed source. From

$$Q_{ij}^{(s)} = \frac{\int d^2\beta q_I[I^{(s)}(\boldsymbol{\beta})] (\beta_i - \bar{\beta}_i)(\beta_j - \bar{\beta}_j)}{\int d^2\beta q_I[I^{(s)}(\boldsymbol{\beta})]} , \quad i, j \in \{1, 2\} , \quad (4.8)$$

one finds with $d^2\beta = \det \mathcal{A} d^2\theta$, $\boldsymbol{\beta} - \bar{\boldsymbol{\beta}} = \mathcal{A}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})$ that

$$Q^{(s)} = \mathcal{A} Q \mathcal{A}^T = \mathcal{A} Q \mathcal{A} , \quad (4.9)$$

where $\mathcal{A} \equiv \mathcal{A}(\bar{\boldsymbol{\theta}})$ is the Jacobi matrix of the lens equation at position $\bar{\boldsymbol{\theta}}$. Using the definitions of the complex ellipticities, one finds the transformations:

$$\boxed{\chi^{(s)} = \frac{\chi - 2g + g^2\chi^*}{1 + |g|^2 - 2\Re(g\chi^*)}} ; \quad \boxed{\epsilon^{(s)} = \begin{cases} \frac{\epsilon - g}{1 - g^*\epsilon} & \text{if } |g| \leq 1 ; \\ \frac{1 - g\epsilon^*}{\epsilon^* - g^*} & \text{if } |g| > 1 . \end{cases}} \quad (4.10)$$

The inverse transformations are obtained by interchanging source and image ellipticities, and $g \rightarrow -g$ in the foregoing equations.

4.2.2. Estimating the (reduced) shear

In the following we make the assumption that the intrinsic orientation of galaxies is random,

$$E(\chi^{(s)}) = 0 = E(\epsilon^{(s)}) , \quad (4.11)$$

which is expected to be valid since there should be no direction singled out in the Universe. This then implies that the expectation value of ϵ is [as obtained by averaging the transformation law (4.10) over the phase of the intrinsic source orientation (Schramm & Kaiser 1995; Seitz & Schneider 1997)]

$$E(\epsilon) = \begin{cases} g & \text{if } |g| \leq 1 \\ 1/g^* & \text{if } |g| > 1 . \end{cases}$$

(4.12)

This is a remarkable result, since it shows that each image ellipticity provides an unbiased estimate of the local shear, though a very noisy one. The noise is determined by the intrinsic ellipticity dispersion

$$\sigma_\epsilon = \sqrt{\langle \epsilon^{(s)} \epsilon^{(s)*} \rangle} .$$

This noise can be beaten down by averaging over many galaxy images. Fortunately, we live in a Universe where the sky is ‘full of faint galaxies’, as was impressively demonstrated by the Hubble Deep Field images (Williams et al. 1996). Hence, the accuracy of a shear estimate depends on the local number density of galaxies for which a shape can be measured. In order to obtain a high density, one requires deep imaging observations. As a rough guide, on a 3 hour exposure under excellent observing conditions with a 3-meter class telescope, about 30 galaxies per arcmin² can be used for a shape measurement.

Note that in the weak lensing regime, $\kappa \ll 1$, $|\gamma| \ll 1$, one finds

$$\gamma \approx g \approx \langle \epsilon \rangle \approx \frac{\langle \chi \rangle}{2} . \quad (4.13)$$

4.3. Problems in measuring shear, and their solutions

4.3.1. Major problems

- Seeing, that is the finite size of the point spread function (PSF), circularizes images; this effect is severe since faint galaxies (i.e. those at a magnitude limit for which the number density is of order 30 per arcmin²) are not larger than the typical seeing disk. Therefore, weak lensing requires imaging with very good seeing.
- The PSF is not circular, owing to e.g., wind shake of the telescope, or tracking errors. However, an anisotropic PSF causes round sources to appear elliptical, and thus mimics shear.
- Galaxy images are *not* isolated, and therefore the integrals in the definition of Q_{ij} have to be cut-off at a finite radius. Hence, one usually uses a weight function q which depends explicitly on $|\theta - \bar{\theta}|$; however, this modifies the transformation (4.10) between image and source ellipticity.
- The sky noise, i.e. the finite brightness of the night sky, introduces a noise component in the measurement of image ellipticities from CCD data, so that only for high-S/N objects can a shape be measured.
- Distortion by telescope and camera optics renders the coaddition of exposures complex; one needs to employ remapping, using accurate astrometry, and sub-pixel coaddition.

Depending on the science application, the shear one wants to measure is of order a few

percent, or even smaller. Essentially all of the effects mentioned can introduce ellipticities of the same order in the measured images if they are not carefully taken into account.

4.3.2. Solutions

In order to deal with these issues, specific software has been developed. The one that has been mostly used up to now is the Kaiser et al. (1995, KSB) method, or its implementation IMCAT. Its basic idea is as follows: the image ellipticity will be determined by the intrinsic ellipticity, the (reduced) shear, the size of the PSF and its anisotropy. Note that the PSF can be investigated by identifying stars (that is: point sources) on the images. The response of χ to the shear depends on the size of the source – for small sources, blurring by seeing reduces the response to a large degree. The size of the sources is estimated from the size of the seeing-convolved images. In addition, the response of χ to a PSF anisotropy also depends on the image size. The KSB method, an essential part of which was put forward by Luppino & Kaiser (1997; for a complete derivation, see Sect. 4.6.2 of Bartelmann & Schneider 2001), results in the relation

$$\boxed{\chi_{\alpha}^{\text{obs}} = \hat{\chi}_{\alpha}^0 + P_{\alpha\beta}^{\text{sm}} q_{\beta} + P_{\alpha\beta}^g g_{\beta}}, \quad (4.14)$$

where $\hat{\chi}^0$ is the ellipticity of the source convolved with the isotropic part of the PSF and therefore, its expectation value vanishes, according to the assumption of randomly oriented sources, $E(\hat{\chi}^0) = 0$. The tensor $P_{\alpha\beta}^{\text{sm}}$ describes the response of the image ellipticity to the PSF anisotropy, which is quantified by the (complex) ellipticity q_{β} of the PSF, as measured from stars. $P_{\alpha\beta}^g$ is a tensor which describes the response of the image ellipticity to an applied shear. Both, $P_{\alpha\beta}^{\text{sm}}$ and $P_{\alpha\beta}^g$, are calculated for each image individually; they depend on higher-order moments of the image brightness distribution and the size of the PSF. Detailed simulations (e.g., Erben et al. 2001; Bacon et al. 2001) have shown that the KSB method can measure shear with better than $\sim 10\%$ accuracy. Several other methods for measuring image shapes in order to obtain an estimate of the local shear have been developed and some of them have already been applied to observational data (Bonnet & Mellier 1995; Kuijken 1999; Kaiser 2000; Bernstein & Jarvis 2002; Refregier & Bacon 2003).

4.4. Magnification effects

The magnification caused by the differential light bending changes the apparent brightness of sources; this leads to two effects:

(a) The observed flux S from a source is changed from its unlensed value S_0 according to $S = \mu S_0$; if $\mu > 1$, sources appear brighter than they would without an intervening lens.

(b) A population of sources in the unlensed solid angle ω_0 is spread over the solid angle $\omega = \mu\omega_0$ due to the magnification.

These two effects affect the number counts of sources differently; which one of them wins depends on the slope of the number counts; one finds

$$\boxed{n(>S, \theta, z) = \frac{1}{\mu(\theta, z)} n_0 \left(> \frac{S}{\mu(\theta, z)}, z \right)}, \quad (4.15)$$

where $n(>S, z)$ and $n_0(>S, z)$ are the lensed and unlensed cumulative number counts of sources, respectively. The first argument of n_0 accounts for the change of the flux, whereas the prefactor in (4.15) stems from the change of apparent solid angle.

As an illustrative example, we consider the case that the source counts follow a power law,

$$n_0(>S) = a S^{-\alpha} ; \quad (4.16)$$

one then finds for the lensed counts in a region of the sky with magnification μ :

$$\frac{n(>S)}{n_0(>S)} = \mu^{\alpha-1} , \quad (4.17)$$

and therefore, if $\alpha > 1$ (< 1), source counts are enhanced (depleted); the steeper the counts, the stronger the effect. In the case of weak lensing, where $|\mu - 1| \ll 1$, one probes the source counts only over a small range in flux, so that they can always be approximated (locally) by a power law.

One important example is provided by the lensing of QSOs. The QSO number counts are steep at the bright end, and flat for fainter sources. This implies that in regions of magnification > 1 , bright QSO should be overdense, faint ones underdense. This *magnification bias* is the reason why the fraction of lensed sources is *much* higher in bright QSO samples than in fainter ones!

4.5. Tangential and cross component of shear

4.5.1. The shear components

The shear components γ_1 and γ_2 are defined relative to a reference Cartesian coordinate frame. Note that the shear is *not* a vector, owing to its transformation properties under rotations, which is the same as that of the linear polarization; it is therefore called a *polar*. In analogy with vectors, it is often useful to consider the shear components in a rotated reference frame, that is, to measure them w.r.t. a different direction; for example, the arcs in clusters are tangentially aligned, and so their ellipticity is oriented approximately tangent to the radius vector in the cluster.

If ϕ specifies a direction, one defines the *tangential* and *cross components* of the shear *relative to this direction* as

$$\boxed{\gamma_t = -\mathcal{R}\text{e} [\gamma e^{-2i\phi}] \quad , \quad \gamma_x = -\mathcal{I}\text{m} [\gamma e^{-2i\phi}]} . \quad (4.18)$$

For example, in the case of a circularly-symmetric matter distribution, the shear at any point will be oriented tangent to the direction towards the center of symmetry. Thus in this case, choose ϕ to be the polar angle of a point; then, $\gamma_x = 0$. In full analogy to the shear, one defines the tangential and cross components of an image ellipticity, ϵ_t and ϵ_x .

4.5.2. Minimum lens strength for its weak lensing detection

As a first application of this decomposition, we consider how massive a lens needs to be in order to produce a detectable weak lensing signal. For this purpose, consider a lens modeled as an SIS with one-dimensional velocity dispersion σ_v . In the annulus $\theta_{\text{in}} \leq \theta \leq \theta_{\text{out}}$, centered on the lens, let there be $N = n\pi(\theta_{\text{out}}^2 - \theta_{\text{in}}^2)$ galaxy images with position $\boldsymbol{\theta}_i = \theta_i(\cos\phi_i, \sin\phi_i)$ and (complex) ellipticities ϵ_i . For each one of them, consider the tangential ellipticity

$$\epsilon_{ti} = -\mathcal{R}\text{e} (\epsilon_i e^{-2i\phi_i}) . \quad (4.19)$$

Next we define a statistical quantity to measure the degree of tangential alignment of the galaxy images, and thus the lens strength:

$$X \equiv \sum_{i=1}^N a_i \epsilon_{ti} , \quad (4.20)$$

where the factors $a_i = a(\theta_i)$ are arbitrary at this point, and will later be chosen such as to maximize the signal-to-noise ratio of this estimator. The shear of an SIS is given by

$$\gamma_t(\theta) = \frac{\theta_E}{2\theta} = 2\pi \left(\frac{\sigma_v}{c} \right)^2 \left(\frac{D_{ds}}{D_s} \right) \frac{1}{\theta}. \quad (4.21)$$

The expectation value of the ellipticity is $E(\epsilon_{ti}) = \gamma_t(\theta_i)$, so that $E(X) = \theta_E \sum_i a_i / (2\theta_i)$. Since $\epsilon_{ti} = \epsilon_{ti}^{(s)} + \gamma_t(\theta_i)$ in the weak lensing regime, one has $E(\epsilon_{ti}\epsilon_{tj}) = \gamma_t(\theta_i)\gamma_t(\theta_j) + \delta_{ij}\sigma_\epsilon^2/2$, thus

$$E(X^2) = \sum_{i,j=1}^N a_i a_j E(\epsilon_{ti}\epsilon_{tj}) = [E(X)]^2 + \frac{\sigma_\epsilon^2}{2} \sum_{i=1}^N a_i^2. \quad (4.22)$$

Therefore, the signal-to-noise ratio for a detection of the lens is

$$\frac{S}{N} = \frac{\theta_E}{\sqrt{2}\sigma_\epsilon} \frac{\sum_i a_i \theta_i^{-1}}{\sqrt{\sum_i a_i^2}}. \quad (4.23)$$

The a_i can now be chosen so as to maximize S/N; from differentiation of S/N, one finds a maximum if $a_i \propto 1/\theta_i$. Then, performing the ensemble average over the galaxy positions, one finally obtains:

$$\begin{aligned} \frac{S}{N} &= \frac{\theta_E}{\sigma_\epsilon} \sqrt{\pi n} \sqrt{\ln(\theta_{out}/\theta_{in})} \\ &= 8.4 \left(\frac{n}{30 \text{ arcmin}^{-2}} \right)^{1/2} \left(\frac{\sigma_\epsilon}{0.3} \right)^{-1} \left(\frac{\sigma_v}{600 \text{ km s}^{-1}} \right)^2 \\ &\times \left(\frac{\ln(\theta_{out}/\theta_{in})}{\ln 10} \right)^{1/2} \left\langle \frac{D_{ds}}{D_s} \right\rangle. \end{aligned} \quad (4.24)$$

From this consideration we conclude that clusters of galaxies with $\sigma_v \gtrsim 600 \text{ km/s}$ can be detected with sufficiently large S/N by weak lensing, but individual galaxies ($\sigma_v \sim 200 \text{ km/s}$) are too weak as lenses to be detected individually.

4.6. Galaxy-galaxy lensing

Whereas galaxies are not massive enough to show a weak lensing signal individually, the signal of many galaxies can be (statistically) superposed. Consider sets of foreground (lens) and background galaxies; on average, in a foreground-background galaxy pair, the ellipticity of the background galaxy will be oriented preferentially in the direction tangent to the connecting line. In other words, if φ is the angle between the major axis of the background galaxy and the connecting line between foreground and background galaxy, values $\pi/4 \leq \varphi \leq \pi/2$ should be slightly more frequent than $0 \leq \varphi \leq \pi/4$. The mean tangential ellipticity $\langle \epsilon_t(\theta) \rangle$ of background galaxies relative to the direction towards foreground galaxies measures the mean tangential shear at this separation.

The strength of this mean tangential shear measures mass properties of the galaxy population selected as potential lenses. In order to properly interpret the lensing signal, one needs to know the redshift distribution of the foreground and background galaxies. Furthermore, one needs to assume a relation between the lens galaxies' luminosity and mass properties (such as a Faber-Jackson type of relation), unless the sample is so large that one can finely bin the galaxies with respect to their luminosities; this requires of course redshift information. In this way, the velocity dispersion σ_* of an L_* -galaxy can be determined from galaxy-galaxy lensing. Furthermore, galaxy-galaxy lensing provides a highly valuable tool to study the mass distribution of galaxy halos at distances from

their centers which are much larger than the extent of luminous tracers, such as stars and gas – or ask the question of where the galaxy halos ‘end’.

Whereas the first detection of galaxy-galaxy lensing (Brainerd et al. 1996) was based on a single field with $\sim 9'$ sidelength, much larger surveys have now become available, most notably the SDSS (Fischer et al. 2000; McKay et al. 2001). These large data sets have allowed the splitting of the lens galaxies into subsamples and to investigate their properties separately. From this it was verified that early-type galaxies have a larger mass than spiral galaxies with the same luminosity, and that this behaviour extends to large radii. Furthermore, the lensing signal for early-type galaxies can be detected out to much larger scales than for late-types. The interpretation of this result is not unique: it either can mean that ellipticals have a more extended halo than spirals, or that the lens signal from ellipticals, which tend to be preferentially located inside groups and clusters, arises in fact from the host halo in which they reside (see also Guzik & Seljak 2002). Indeed, when the lens galaxy sample is divided into those living in high- and low-density environments, the former ones have a significantly more extended lensing signal.

What the galaxy-galaxy signal really measures is the relation between light (galaxies) and mass. In its simplest terms, this relation can be expressed by a bias factor b and the correlation coefficient r . Schneider (1998) and van Waerbeke (1998) pointed out that lensing can be used to study the bias factor as a function of scale and redshift, by correlating the lensing signal with the number density of galaxies. In fact, as shown in Hoekstra et al. (2002b), both b and r can be expressed in terms of the galaxy-galaxy lensing signal, the angular correlation function of the (lens) galaxies and the cosmic shear signal (see Sect. 6 below). Applying this method to the combination of the Red-Sequence Cluster Survey and the VIRMOS-DESCART survey, they derived the scale dependence of b and r ; on large (linear) scales, their results are compatible with constant values, whereas on smaller scales it appears that both of these functions vary. Future surveys will allow much more detailed studies on the relation between mass and light, and therefore determine the biasing properties of galaxies from observations directly. This is of course of great interest, since the unknown behavior of the biasing yields the uncertainty in the transformation of the power spectrum of the galaxy distribution, as determined from extensive galaxy redshift surveys, to that of the underlying mass distribution. Hence, weak lensing is able to provide this crucial calibration.

5. Lensing by clusters of galaxies

5.1. Introduction

Clusters are the most massive bound structures in the Universe; this, together with the (related) fact that their dynamical time scale (e.g., the crossing time) is not much smaller than the Hubble time – so that they retain a ‘memory’ of their formation – render them of particular interest for cosmologists. The evolution of their abundance, i.e., their comoving number density as a function of mass and redshift, is an important probe for cosmological models. Furthermore, they form signposts of the dark matter distribution in the Universe. Clusters act as laboratories for studying the evolution of galaxies and baryons in the Universe. In fact, clusters were (arguably) the first objects for which the presence of dark matter has been concluded (by Zwicky in 1933).

5.2. The mass of galaxy clusters

Cosmologists can predict the abundance of clusters as a function of their mass (e.g., using numerical simulations); however, the mass of a cluster is not directly observable, but only its luminosity, or the temperature of the X-ray emitting intra-cluster medium.

Therefore, in order to compare observed clusters with the cosmological predictions, one needs a way to determine their masses. Three principal methods for determining the mass of galaxy clusters are in use:

- Assuming virial equilibrium, the observed velocity distribution of galaxies in clusters can be converted into a mass estimate; this method typically requires assumptions about the statistical distribution of the anisotropy of the galaxy orbits.
- The hot intra-cluster gas, as visible through its Bremsstrahlung in X-rays, traces the gravitational potential of the cluster. Under certain assumptions (see below), the mass profile can be constructed from the X-ray emission.
- Weak and strong gravitational lensing probe the projected mass profile of clusters; this will be described further below.

All three methods are complementary; lensing yields the line-of-sight projected density of clusters, in contrast to the other two methods which probes the mass inside spheres. On the other hand, those rely on equilibrium (and symmetry) conditions.

5.2.1. *X-ray mass determination of clusters*

The intracluster gas emits via Bremsstrahlung; the emissivity depends on the gas density and temperature, and, at lower T , on its chemical composition. Assuming that the gas is in hydrostatic equilibrium in the potential well of cluster, the gas pressure P must balance gravity, or

$$\nabla P = -\rho_g \nabla \Phi,$$

where Φ is the gravitational potential and ρ_g is the gas density. In the case of spherical symmetry, this becomes

$$\frac{1}{\rho_g} \frac{dP}{dr} = -\frac{d\Phi}{dr} = -\frac{GM(r)}{r^2}.$$

From the X-ray brightness profile and temperature measurement, $M(r)$, the total mass inside r (dark plus luminous) can then be determined,

$$M(r) = -\frac{k_B T r^2}{G \mu m_p} \left(\frac{d \ln \rho_g}{dr} + \frac{d \ln T}{dr} \right).$$

(5.1)

However, the two major X-ray satellites currently operating, Chandra & XMM-Newton, have revealed that at least the inner regions of clusters show a considerably more complicated structure than implied by hydrostatic equilibrium. In some cases, the intracluster medium is obviously affected by a central AGN, which produces additional energy and entropy input. Cold fronts, with very sharp edges, and shocks have been discovered, most likely showing ongoing merger events. The temperature and metallicity appear to be strongly varying functions of position. Therefore, mass estimates of central parts of clusters from X-ray observations require special care.

5.3. *Luminous arcs & multiple images*

Strong lensing effects in clusters show up in the form of giant luminous arcs, strongly distorted arclets, and multiple images of background galaxies. Since strong lensing occurs only in the central part of clusters (typically corresponding to $\sim 50 h^{-1}$ kpc), it can be used to probe only their inner mass structure. However, strong lensing yields by far the most accurate central mass determinations; in some favourable cases with many strong lensing features (such as for Abell 2218; see Fig. 2), accuracies better than $\sim 10\%$ can be achieved.

5.3.1. First go: $M(\leq \theta_E)$

Giant arcs occur where the distortion (and magnification) is very large, that is near critical curves. To a first approximation, assuming a spherical mass distribution, the location of the arc relative to the cluster center (which usually is assumed to coincide with the brightest cluster galaxy) yields the Einstein radius of the cluster, so that the mass estimate (3.1) can be applied. Therefore, this simple estimate yields the mass inside the arc radius. However, this estimate is not very accurate, perhaps good to within $\sim 50\%$. Its reliability depends on the level of asymmetry and substructure in the cluster mass distribution (Bartelmann 1995). Furthermore, it is likely to overestimate the mass in the mean, since arcs preferentially occur along the major axis of clusters. Of course, the method is very difficult to apply if the center of the cluster is not readily identified or if it is obviously bimodal. For these reasons, this simple method for mass estimates is not regarded as particularly accurate.

5.3.2. Detailed modelling

The mass determination in cluster centers becomes much more accurate if several arcs and/or multiple images are present, since in this case, detailed modelling can be done. This typically proceeds in an interactive way: First, multiple images have to be identified (based on their colors and/or detailed morphology, as available with HST imaging). Simple (plausible) mass models are then assumed, with parameters fixed by matching the multiple images, and requiring the distortion at the arc location(s) to be strong and have the correct orientation. This model then predicts the presence of further multiple images; they can be checked for through morphology and color. If confirmed, a new, refined model is constructed, which yields further strong lensing predictions etc. Such models have predictive power and can be trusted in quite some detail; the accuracy of mass estimates in some favourable cases can be as high as a few percent.

In fact, these models can be used to predict the redshift of arcs and arclets (Kneib et al. 1994): since the distortion of a lens depends on the source redshift, once a detailed mass model is available, one can estimate the value of the lens strength $\propto D_{ds}/D_s$ and thus infer the redshift. This method has been successfully applied to HST observations of clusters (Ebbels et al. 1998). Of course, having spectroscopic redshifts of the arcs available increases the accuracy of the calibration of the mass models; they are therefore very useful.

5.3.3. Results

The main results of the strong lensing investigations of clusters can be summarized as follows:

- The mass in cluster centers is much more concentrated than predicted by (simple) models based on X-ray observations. The latter usually predict a relatively large core of the mass distribution. These large cores would render clusters sub-critical to lensing, i.e., they would be unable to produce giant arcs or multiple images. In fact, when arcs were first discovered they came as a big surprise because of these expectations. By now we know that the intracluster medium is much more complicated than assumed in these ‘ β -model’ fits for the X-ray emission.
- The mass distribution in the inner part of clusters often shows strong substructure, or multiple mass peaks. These are also seen in the galaxy distribution of clusters, but with the arcs can be verified to also correspond to mass peaks. These are easily understood in the frame of hierarchical mergers in a CDM model; the merged clusters retain their multiple peaks for a dynamical time or even longer, and are therefore not in virial equilibrium.

- The orientation of the (dark) matter appears to be fairly strongly correlated with the orientation of the light in the cD galaxy; this supports the idea that the growth of the cD galaxy is related to the cluster as a whole, through repeated accretion of lower-mass member galaxies. In that case, the cD galaxy ‘knows’ the orientation of the cluster.
- There is in general good agreement between lensing and X-ray mass estimates for those clusters where a ‘cooling flow’ indicates that they are in dynamical equilibrium, provided the X-ray analysis takes the presence of the cooling flow into account.

5.4. Mass reconstructions from weak lensing

Whereas strong lensing probes the mass distribution in the inner part of clusters, weak lensing can be used to study the mass distribution at much larger angular separations from the cluster center. In fact, as we shall see, weak lensing can provide a parameter-free reconstruction of the projected two-dimensional mass distribution in clusters. This discovery (Kaiser & Squires 1993) actually marked the beginning of quantitative weak lensing research.

5.4.1. The Kaiser–Squires inversion

Weak lensing yields an estimate of the local (reduced) shear, as discussed in Sect. 4.2. Here we shall discuss how to derive the surface mass density from a measurement of the (reduced) shear. Starting from (2.9) and the definition (2.15) of the shear, one finds that the latter can be written in the form

$$\begin{aligned}\gamma(\boldsymbol{\theta}) &= \frac{1}{\pi} \int_{\mathbb{R}^2} d^2\boldsymbol{\theta}' \mathcal{D}(\boldsymbol{\theta} - \boldsymbol{\theta}') \kappa(\boldsymbol{\theta}') , \quad \text{with} \\ \mathcal{D}(\boldsymbol{\theta}) &\equiv \frac{\theta_2^2 - \theta_1^2 - 2i\theta_1\theta_2}{|\boldsymbol{\theta}|^4} = \frac{-1}{(\theta_1 - i\theta_2)^2} .\end{aligned}\quad (5.2)$$

Hence, the complex shear γ is a convolution of κ with the kernel \mathcal{D} , or, in other words, \mathcal{D} describes the shear generated by a point mass. In Fourier space this convolution becomes a multiplication,

$$\hat{\gamma}(\boldsymbol{\ell}) = \pi^{-1} \hat{\mathcal{D}}(\boldsymbol{\ell}) \hat{\kappa}(\boldsymbol{\ell}) \quad \text{for } \boldsymbol{\ell} \neq \mathbf{0} .$$

This relation can be inverted to yield

$$\hat{\kappa}(\boldsymbol{\ell}) = \pi^{-1} \hat{\gamma}(\boldsymbol{\ell}) \hat{\mathcal{D}}^*(\boldsymbol{\ell}) \quad \text{for } \boldsymbol{\ell} \neq \mathbf{0} , \quad (5.3)$$

where

$$\hat{\mathcal{D}}(\boldsymbol{\ell}) = \pi \frac{(\ell_1^2 - \ell_2^2 + 2i\ell_1\ell_2)}{|\boldsymbol{\ell}|^2}$$

was used. Fourier back-transformation of (5.3) then yields

$$\kappa(\boldsymbol{\theta}) - \kappa_0 = \frac{1}{\pi} \int_{\mathbb{R}^2} d^2\boldsymbol{\theta}' \mathcal{D}^*(\boldsymbol{\theta} - \boldsymbol{\theta}') \gamma(\boldsymbol{\theta}') = \frac{1}{\pi} \int_{\mathbb{R}^2} d^2\boldsymbol{\theta}' \operatorname{Re} [\mathcal{D}^*(\boldsymbol{\theta} - \boldsymbol{\theta}') \gamma(\boldsymbol{\theta}')] . \quad (5.4)$$

Note that the constant κ_0 occurs since the $\boldsymbol{\ell} = \mathbf{0}$ -mode is undetermined. Physically, this is related to the fact that a uniform surface mass density yields no shear. Furthermore, it is obvious (physically, though not so easily seen mathematically) that κ must be real; for this reason, the imaginary part of the integral should be zero, and taking the real-part only makes no difference. However, in practice it is different, as noisy data, when inserted into the inversion formula, will produce a non-zero imaginary part. What (5.4) shows is that if γ can be measured, κ can be determined.

Before looking at this in more detail, we briefly mention some difficulties with the inversion formula as given above:

- Since γ can at best be estimated at discrete points (galaxy images), smoothing is required. One might be tempted to replace the integral in (5.4) by a discrete sum over galaxy positions, but as shown by Kaiser & Squires (1993), the resulting mass density estimator has infinite noise (due to the θ^{-2} -behavior of the kernel \mathcal{D}).
- It is not the shear γ , but the reduced shear g that can be determined from the galaxy ellipticities; hence, one needs to obtain a mass density estimator in terms of g .
- The integral in (5.4) extends over \mathbb{R}^2 , whereas data are available only on a finite field; therefore, it needs to be seen whether modifications allow the construction of an estimator for the surface mass density from finite-field shear data.
- To get absolute values for the surface mass density, the additive constant κ_0 is of course a nuisance. As will be explained soon, this indeed is the largest problem in mass reconstructions, and carries the name *mass-sheet degeneracy* (note that we mentioned this effect before, in the context of determining the Hubble constant from time-delays in lens systems).

5.4.2. Improvements and generalizations

Smoothing of data is needed to get a shear field from discrete data points. When smoothed with Gaussian kernel of angular scale θ_s , the covariance of the resulting mass map is finite, and given by (Lombardi & Bertin 1998; van Waerbeke 2000)

$$\text{Cov}(\kappa(\boldsymbol{\theta}), \kappa(\boldsymbol{\theta}')) = \frac{\sigma_\epsilon^2}{4\pi\theta_s^2 n} \exp\left(-\frac{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^2}{2\theta_s^2}\right).$$

Thus, the larger the smoothing scale, the less noise does the corresponding mass map have. Note that, since (i) smoothing can be represented by a convolution, (ii) the relation between κ and γ is a convolution, and (iii) convolution operations are transitive, it does not matter whether the shear field is smoothed first and inserted into (5.4), or the noisy inversion obtained by transforming the integral into a sum over galaxy image positions is smoothed afterwards with the same smoothing kernel.

Noting that it is the reduced shear $g = \gamma/(1 - \kappa)$ that can be estimated from the ellipticity of images, one can write:

$$\kappa(\boldsymbol{\theta}) - \kappa_0 = \frac{1}{\pi} \int_{\mathbb{R}^2} d^2\boldsymbol{\theta}' [1 - \kappa(\boldsymbol{\theta}')] \text{Re}[\mathcal{D}^*(\boldsymbol{\theta} - \boldsymbol{\theta}') g(\boldsymbol{\theta}')] ; \quad (5.5)$$

this integral equation for κ can be solved by iteration, and it converges quickly. Note that in this case, the undetermined constant κ_0 no longer corresponds to adding a uniform mass sheet. What the arbitrary value of κ_0 corresponds to can be seen as follows: The transformation

$$\begin{aligned} \kappa(\boldsymbol{\theta}) &\rightarrow \kappa'(\boldsymbol{\theta}) = \lambda\kappa(\boldsymbol{\theta}) + (1 - \lambda) \quad \text{or} \\ [1 - \kappa'(\boldsymbol{\theta})] &= \lambda[1 - \kappa(\boldsymbol{\theta})] \end{aligned} \quad (5.6)$$

changes the shear $\gamma \rightarrow \gamma' = \lambda\gamma$, and thus leaves g invariant; this is the mass-sheet degeneracy! It can be broken if magnification information can be obtained, since

$$\mu \rightarrow \lambda^{-2} \mu .$$

Magnification can in principle be obtained from the number counts of images (Broadhurst et al. 1995), owing to magnification bias, provided the unlensed number density is sufficiently well known. Indeed, magnification effects have been detected in a few clusters as a number depletion of faint galaxy images towards the center of the clusters (e.g., Fort et al. 1997; Taylor et al. 1998; Dye et al. 2002). In principle, the mass sheet degeneracy can also be broken if redshift information of the source galaxies is available and if the sources are widely distributed in redshift; however, even in this case it is only mildly

broken, in the sense that one needs a fairly high number density of background galaxies in order to fix the parameter λ to within $\sim 10\%$.

Finite-field inversions start from the relation (Kaiser 1995)

$$\nabla \kappa = \begin{pmatrix} \gamma_{1,1} + \gamma_{2,2} \\ \gamma_{2,1} - \gamma_{1,2} \end{pmatrix} \equiv \mathbf{u}_\gamma(\boldsymbol{\theta}), \quad (5.7)$$

which is a *local* relation between shear and surface mass density; it can easily be derived from the definitions of κ (2.10) and γ (2.15) in terms of ψ_{ij} . A similar relation can be derived in terms of reduced shear,

$$\nabla K(\boldsymbol{\theta}) = \frac{-1}{1 - g_1^2 - g_2^2} \begin{pmatrix} 1 - g_1 & -g_2 \\ -g_2 & 1 + g_1 \end{pmatrix} \begin{pmatrix} g_{1,1} + g_{2,2} \\ g_{2,1} - g_{1,2} \end{pmatrix} \equiv \mathbf{u}_g(\boldsymbol{\theta}), \quad (5.8)$$

where

$$K(\boldsymbol{\theta}) \equiv \ln[1 - \kappa(\boldsymbol{\theta})] \quad (5.9)$$

is a non-linear function of κ . These equations can be integrated, by formulating them as a von Neumann boundary-value problem on the data field \mathcal{U} (Seitz & Schneider 2001),

$$\nabla^2 \kappa = \nabla \cdot \mathbf{u}_\gamma \quad \text{with} \quad \mathbf{n} \cdot \nabla \kappa = \mathbf{n} \cdot \mathbf{u}_\gamma \quad \text{on} \quad \partial \mathcal{U}; \quad (5.10)$$

where \mathbf{n} is the outward-directed normal on the boundary $\partial \mathcal{U}$ of \mathcal{U} . The analogous equation holds for K in terms of g and \mathbf{u}_g . The numerical solution of these equations is fast, using overrelaxation (see Press et al. 1992). In fact, the foregoing formulation of the problem is equivalent (Lombardi & Bertin 1998) to the minimization of the action

$$A = \int_{\mathcal{U}} d^2\theta |\nabla \kappa(\boldsymbol{\theta}) - \mathbf{u}_\gamma(\boldsymbol{\theta})|^2, \quad (5.11)$$

from which the von Neumann problem can be derived as the Euler equation of the variational principle $\delta A = 0$. These parameter-free mass reconstructions have been applied to quite a number of clusters; it provides a tool to make their dark matter distribution ‘visible’.

5.4.3. Results

The mass reconstruction techniques discussed above have been applied to quite a number of clusters up to now, yielding parameter-free mass maps. It is obvious that the quality of a mass map depends on the number density of galaxies that can be used for a shear estimate, which in turn depends on the depth and the seeing of the observational data. Furthermore, the mass profiles of clusters are much more reliably determined if the data field covers a large region, as boundary effects get minimized.

The first application of the Kaiser & Squires reconstruction technique was done to the X-ray detected cluster MS1224+20 (Fahlman et al. 1994); it resulted in an estimate of the mass-to-light ratio in this $z = 0.33$ cluster of $\sim 800 h$, considerably larger than ‘normal’ values of $M/L \sim 250 h$. This conclusion was later reinforced by a fully independent weak-lensing analysis of this cluster by Fischer (1999). This mass estimate is in fact in conflict with the measured velocity dispersion of the cluster galaxies, which is much smaller than obtained by an SIS fit to the shear data. The line-of-sight to this cluster is fairly complicated, with additional peaks in the redshift distribution of galaxies in the field (Carlberg et al. 1994), all of which are included in the weak lensing measurement. Furthermore, this cluster may not be in a relaxed state, which probably renders the X-ray mass analysis inaccurate.

In fact, non-relaxed clusters are probably more common than naively expected. One example is shown in Fig. 6, a mass reconstruction of a high-redshift cluster based on

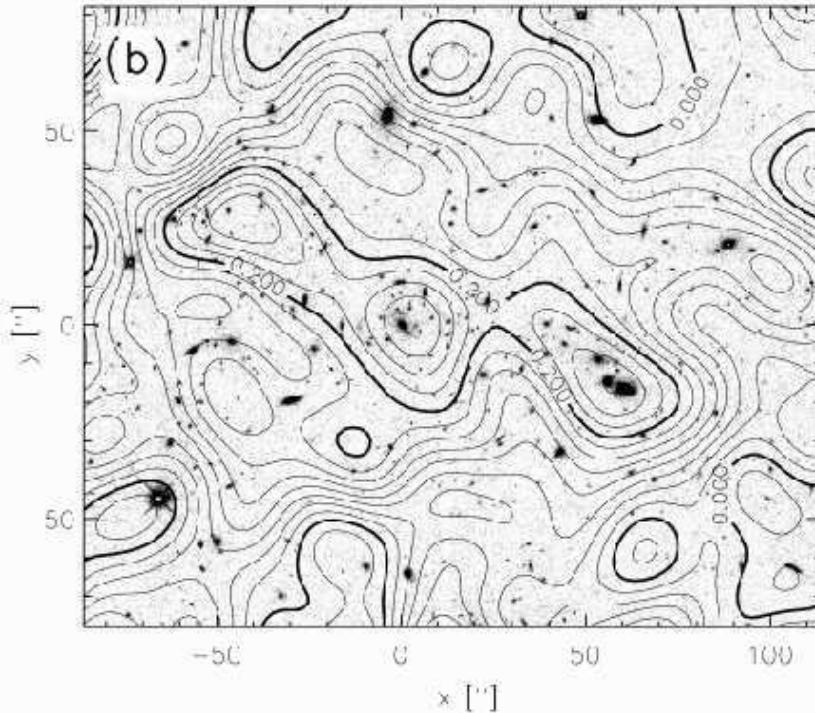


FIGURE 6. Mass reconstruction (contours) of the inner part of the high redshift ($z_d = 0.83$) cluster MS1054–03, based on a mosaic of six pointings obtained with the WFPC2@HST (from Hoekstra et al. 2000). The splitting of the cluster core into three subcomponents, previously seen from ground-based images by Clowe et al. (2000), shows that this cluster is not yet relaxed

HST data. The presence of three mass clumps, which coincide with three concentrations of cluster galaxies, indicates that this cluster is still in the process of merging. When the merging occurs along the line-of-sight, then it is less obvious in the mass maps. One example seems to be the cluster Cl0024+16, for which one obtains a large mass from the distance of its arcs from the cluster center (Colley et al. 1996), but which is fairly underluminous in X-rays for this mass. A detailed investigation of the structure of this cluster in radial velocity space by Czoske et al. (2002) has shown strong evidence for a collision of two clusters along the line-of-sight. Another example is provided by the cluster A1689, where the extended arc structures suggest an Einstein radius of about $40''$ for this cluster, making this the strongest lensing cluster in the sky, but the weak lensing results do not support the enormous mass obtained from the arcs (see Clowe & Schneider 2001, King et al. 2002, and references therein).

However, in many clusters the weak lensing mass estimates are in good agreement with those from dynamical estimates and X-ray determinations (e.g., Squires et al. 1996), provided the inner region of the clusters are omitted – but for them, the weak lensing method does not have sufficient angular resolution anyway. For example, Hoekstra and collaborators have observed three X-ray selected clusters with HST mosaics, and their results, summarized in Hoekstra et al. (2002a), shows that the SIS fit values for the velocity dispersion agree with those from spectroscopic investigations.

The mass maps can also be used to study how well the cluster galaxy distribution traces the underlying dark matter. An HST data based mass reconstruction of Cl0939+47 (Seitz et al. 1996) shows detailed structure that is very well matched with the distribution of

bright cluster galaxies. A more quantitative investigation was performed by Wilson et al. (2001) showing that early-type galaxies trace the dark matter distribution quite well.

One of the predictions of CDM models for structure formation is that clusters of galaxies are located at the intersection points of filaments. In particular, this implies that a physical pair of clusters should be connected by a bridge or filament of (dark) matter, and weak lensing mass reconstructions can in principle be used to search for them. In the investigation of the $z = 0.42$ supercluster MS0302, Kaiser et al. (1998) found an indication of a possible filament connecting two of the three clusters, with the caveat (as pointed out by the authors) that the filament lies just along the boundary of two CCD chips. Gray et al. (2002) saw a filament connecting the two clusters A901A/901B in their mass reconstruction of the A901/902 supercluster field. One of the problems related to the unambiguous detection of filaments is the difficulty to define what a ‘filament’ is, i.e. to device a statistics to quantify the presence of a mass bridge. Because of that, it is difficult to distinguish between noise in the mass maps, the ‘elliptical’ extension of two clusters pointing towards each other, and a true filament.

A perhaps surprising result is the difficulty of distinguishing between the NFW mass profile from, say power-law models, such as the isothermal profile. In fact, even from weak lensing observations covering large fields, out to the virial radius of clusters (Clowe & Schneider 2001, 2002; King et al. 2002), the distinction between NFW and isothermal is present only at the $\lesssim 2\sigma$ level. The reason for this is the mass-sheet degeneracy. Within a family of models (such as the NFW), the model parameters can be determined with fairly high accuracy. One way to improve on the distinction between various mass profiles is to incorporate strong lensing constraints into the mass reconstruction (e.g., using inverse methods; see below). In particular, the multiple images seen in the cores of clusters can be used to determine the central mass profile (see Sand et al. 2002; Gavazzi et al. 2003). On the other hand, one can statistically superpose weak lensing measurements of clusters to obtain their average mass profile, as done by Dahle et al. (2003) for six clusters; even in that case, a (generalized) NFW profile is hardly distinguishable from an isothermal model.

5.4.4. *Inverse methods*

In addition to these ‘direct’ methods for determining κ , inverse methods have been developed, such as a maximum-likelihood fit (Bartelmann et al. 1996) to the data. In these techniques, one parameterizes the lens by the deflection potential ψ on a grid and then maximizes

$$\mathcal{L} = \prod_{i=1}^{N_g} \frac{1}{\pi \sigma_i^2(\boldsymbol{\theta}_i, \{\psi_n\})} \exp \left(-\frac{|\epsilon_i - g(\boldsymbol{\theta}_i, \{\psi_n\})|^2}{\sigma_i^2(\boldsymbol{\theta}_i, \{\psi_n\})} \right) \quad (5.12)$$

with respect to these gridded ψ -values. In order to avoid overfitting, one needs a regularization; entropy regularization (Seitz et al. 1998) seems best suited. It should be pointed out that the deflection potential ψ , and not the surface mass density κ , should be used as a variable, for two reasons: first, shear and κ depend locally on ψ , and are thus readily calculated by finite differencing, whereas the relation between γ and κ is non-local and requires summation over all gridpoints. Second, and more importantly, the surface mass density on a finite field *does not* determine γ on this field, since mass outside the field contributes to γ as well.

There are a number of reasons why inverse methods are in principle preferable to the direct method discussed above. First, in the direct methods, the smoothing scale is set arbitrarily, and in general kept constant. It would be useful to have an objective way how to choose this scale, and perhaps, the smoothing scale be a function of position: e.g.,

in regions with larger number densities of sources, the smoothing scale could be reduced. Second, the direct methods do not allow additional input coming from observations; for example, if both shear and magnification information are available, the latter could not be incorporated into the mass reconstruction. The same is true for clusters where strong lensing constraints are known.

5.5. Aperture mass

In the weak lensing regime, $\kappa \ll 1$, the mass-sheet degeneracy corresponds to adding a uniform surface mass density κ_0 . We shall now consider a quantity, linearly related to κ , that is unaffected by the mass-sheet degeneracy. Let $U(|\boldsymbol{\theta}|)$ be a compensated weight (or filter) function, with $\int d\theta \theta U(\theta) = 0$, then the *aperture mass*

$$M_{\text{ap}}(\boldsymbol{\theta}_0) = \int d^2\theta \kappa(\boldsymbol{\theta}) U(|\boldsymbol{\theta} - \boldsymbol{\theta}_0|) \quad (5.13)$$

is independent of κ_0 , as can be easily seen. The important point to notice is that M_{ap} can be written directly in terms of the shear (Schneider 1996)

$$M_{\text{ap}}(\boldsymbol{\theta}_0) = \int d^2\theta Q(|\boldsymbol{\theta}|) \gamma_t(\boldsymbol{\theta}; \boldsymbol{\theta}_0) , \quad (5.14)$$

where we have defined the *tangential component* γ_t of the shear relative to the point $\boldsymbol{\theta}_0$, and

$$Q(\theta) = \frac{2}{\theta^2} \int_0^\theta d\theta' \theta' U(\theta') - U(\theta) . \quad (5.15)$$

These relations can be derived from (5.7), by rewriting the partial derivatives in polar coordinates and subsequent integration by parts.

We shall now consider a few properties of the aperture mass.

- If U has finite support, then Q has finite support. This implies that the aperture mass can be calculated on a finite data field.
- If $U(\theta) = \text{const.}$ for $0 \leq \theta \leq \theta_{\text{in}}$, then $Q(\theta) = 0$ for the same interval. Therefore, the strong lensing regime (where the shear γ deviates significantly from the reduced shear g) can be avoided by properly choosing U (and Q).
- If $U(\theta) = (\pi\theta_{\text{in}}^2)^{-1}$ for $0 \leq \theta \leq \theta_{\text{in}}$, $U(\theta) = -[\pi(\theta_{\text{out}}^2 - \theta_{\text{in}}^2)]^{-1}$ for $\theta_{\text{in}} < \theta \leq \theta_{\text{out}}$, and $U = 0$ for $\theta > \theta_{\text{out}}$, then $Q(\theta) = \theta_{\text{out}}^2 \theta^{-2} [\pi(\theta_{\text{out}}^2 - \theta_{\text{in}}^2)]^{-1}$ for $\theta_{\text{in}} \leq \theta \leq \theta_{\text{out}}$, and $Q(\theta) = 0$ otherwise. For this special choice of U ,

$$M_{\text{ap}} = \bar{\kappa}(\theta_{\text{in}}) - \bar{\kappa}(\theta_{\text{in}}, \theta_{\text{out}}) , \quad (5.16)$$

the mean mass density inside θ_{in} minus the mean density in the annulus $\theta_{\text{in}} \leq \theta \leq \theta_{\text{out}}$. Since the latter is non-negative, this yields lower limit to $\bar{\kappa}(\theta_{\text{in}})$, and thus to $M(\theta_{\text{in}})$.

6. Cosmic shear – lensing by the LSS

Up to now we have considered the lensing effect of localized mass concentrations, like galaxies and clusters. In addition to that, light bundles propagating through the Universe are continuously deflected and distorted by the gravitational field of the inhomogeneous mass distribution, the large-scale structure (LSS) of the cosmic matter field (the reader is referred to John Peacock's lecture for the definition of cosmological parameters and the theory of structure growth in the Universe). This distortion of light bundles causes shape distortions of images of distant galaxies, and therefore, the statistics of the distortions reflect the statistical properties of the LSS.

Cosmic shear deals with the investigation of this connection, from the measurement of the correlated image distortion to the inference of cosmological information from this distortion statistics. As we shall see, cosmic shear has become a very important tool in observational cosmology. From a technical point-of-view, it is quite challenging, first because the distortions are indeed very weak and therefore difficult to measure, and second, in contrast to ‘ordinary’ lensing, here the light deflection does not occur in a ‘lens plane’ but by a 3-D matter distribution; one therefore needs a different description of the lensing optics. We start by looking at the description of light propagating through the Universe.

6.1. Light propagation in an inhomogeneous Universe

The laws of light propagation follow from Einstein’s General Relativity; according to it, light propagates along the null-geodesics of the space-time metric. As shown in SEF, one can derive from General Relativity that the governing equation for the propagation of thin light bundles through an arbitrary space-time is the equation of geodesic deviation,

$$\frac{d^2\boldsymbol{\xi}}{d\lambda^2} = \mathcal{T}\boldsymbol{\xi}, \quad (6.1)$$

where $\boldsymbol{\xi}$ is the separation vector of two neighboring light rays, λ the affine parameter along the central ray of the bundle, and \mathcal{T} is the *optical tidal matrix* which describes the influence of space-time curvature on the propagation of light. \mathcal{T} can be expressed directly in terms of the Riemann curvature tensor.

For the case of a weakly inhomogeneous Universe, the tidal matrix can be explicitly calculated in terms of the Newtonian potential. For that, we write the slightly perturbed metric of the Universe in the form

$$ds^2 = a^2(\tau) \left[\left(1 + \frac{2\Phi}{c^2} \right) c^2 d\tau^2 - \left(1 - \frac{2\Phi}{c^2} \right) (dw^2 + f_K^2(w)d\omega^2) \right], \quad (6.2)$$

where w is the comoving radial distance, $a = (1+z)^{-1}$ the scale factor, normalized to unity today, τ is the conformal time, related to the cosmic time t through $dt = a d\tau$, $f_K(w)$ is the comoving angular diameter distance, which equals w in a spatially flat model, and Φ denotes the Newtonian peculiar gravitational potential. In this metric, the equation of geodesic deviation yields, for the comoving separation vector $\mathbf{x}(\boldsymbol{\theta}, w)$ between a ray separated by an angle $\boldsymbol{\theta}$ at the observer from a fiducial ray, the evolution equation

$$\frac{d^2\mathbf{x}}{dw^2} + K \mathbf{x} = -\frac{2}{c^2} \left[\nabla_{\perp} \Phi(\mathbf{x}(\boldsymbol{\theta}, w), w) - \nabla_{\perp} \Phi^{(0)}(w) \right], \quad (6.3)$$

where $K = (H_0/c)^2 (\Omega_m + \Omega_\Lambda - 1)$ is the spatial curvature, $\nabla_{\perp} = (\partial/\partial x_1, \partial/\partial x_2)$ is the transverse *comoving* gradient operator, and $\Phi^{(0)}(w)$ is the potential along the fiducial ray. The formal solution of the transport equation is obtained by the method of Green’s function, to yield

$$\mathbf{x}(\boldsymbol{\theta}, w) = f_K(w)\boldsymbol{\theta} - \frac{2}{c^2} \int_0^w dw' f_K(w-w') \left[\nabla_{\perp} \Phi(\mathbf{x}(\boldsymbol{\theta}, w'), w') - \nabla_{\perp} \Phi^{(0)}(w') \right]. \quad (6.4)$$

A source at comoving distance w with comoving separation \mathbf{x} from the fiducial light ray would be seen, in the absence of lensing, at the angular separation $\boldsymbol{\beta} = \mathbf{x}/f_K(w)$ from the fiducial ray (this statement is nothing but the definition of the comoving angular diameter distance). Hence, in analogy with standard lens theory, we define the Jacobian matrix

$$\mathcal{A}(\boldsymbol{\theta}, w) = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = \frac{1}{f_K(w)} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}}, \quad (6.5)$$

and obtain

$$\mathcal{A}_{ij}(\boldsymbol{\theta}, w) = \delta_{ij} - \frac{2}{c^2} \int_0^w dw' \frac{f_K(w-w')f_K(w')}{f_K(w)} \Phi_{,ik}(\mathbf{x}(\boldsymbol{\theta}, w'), w') \mathcal{A}_{kj}(\boldsymbol{\theta}, w'), \quad (6.6)$$

which describes the locally linearized mapping introduced by LSS lensing. This equation still is exact in the limit of validity of the weak-field metric. Next, we expand \mathcal{A} in powers of Φ , and truncate the series after the linear term:

$$\boxed{\mathcal{A}_{ij}(\boldsymbol{\theta}, w) = \delta_{ij} - \frac{2}{c^2} \int_0^w dw' \frac{f_K(w-w')f_K(w')}{f_K(w)} \Phi_{,ij}(f_K(w')\boldsymbol{\theta}, w')} . \quad (6.7)$$

Hence, to linear order, the distortion can be obtained by integrating along the unperturbed ray; this is also called the Born approximation. Corrections to the Born approximation are necessarily of order Φ^2 . If we now define the deflection potential

$$\psi(\boldsymbol{\theta}, w) := \frac{2}{c^2} \int_0^w dw' \frac{f_K(w-w')f_K(w')}{f_K(w)} \Phi(f_K(w')\boldsymbol{\theta}, w') \quad (6.8)$$

then $\mathcal{A}_{ij} = \delta_{ij} - \psi_{,ij}$, just as in ordinary lens theory. *In this approximation, lensing by the 3-D matter distribution can be treated as an equivalent lens plane with deflection potential ψ , mass density $\kappa = \nabla^2\psi/2$, and shear $\gamma = (\psi_{,11} - \psi_{,22})/2 + i\psi_{,12}$.*

6.2. Cosmic shear: the principle

6.2.1. The effective surface mass density

Next, we relate κ to the fractional density contrast δ of matter fluctuations in the Universe; this is done in a number of steps:

- (a) Take the 2-D Laplacian of ψ , and add the term $\Phi_{,33}$ in the integrand; this latter term vanishes in the line-of-sight integration, as can be seen by integration by parts.
- (b) We make use of the 3-D Poisson equation in comoving coordinates

$$\nabla^2\Phi = \frac{3H_0^2\Omega_m}{2a}\delta \quad (6.9)$$

to obtain

$$\boxed{\kappa(\boldsymbol{\theta}, w) = \frac{3H_0^2\Omega_m}{2c^2} \int_0^w dw' \frac{f_K(w')f_K(w-w')}{f_K(w)} \frac{\delta(f_K(w')\boldsymbol{\theta}, w')}{a(w')}} . \quad (6.10)$$

Note that κ is proportional to Ω_m , since lensing is sensitive to $\Delta\rho \propto \Omega_m \delta$, not just to the density contrast $\delta = \Delta\rho/\bar{\rho}$ itself.

- (c) For a redshift distribution of sources with $p_z(z) dz = p_w(w) dw$, the effective surface mass density becomes

$$\boxed{\kappa(\boldsymbol{\theta}) = \int dw p_w(w) \kappa(\boldsymbol{\theta}, w) = \frac{3H_0^2\Omega_m}{2c^2} \int_0^{w_h} dw g(w) f_K(w) \frac{\delta(f_K(w)\boldsymbol{\theta}, w)}{a(w)}} \quad (6.11)$$

with

$$\boxed{g(w) = \int_w^{w_h} dw' p_w(w') \frac{f_K(w'-w)}{f_K(w')}} , \quad (6.12)$$

which is essentially the source-redshift weighted lens efficiency factor D_{ds}/D_s for a density fluctuation at distance w , and w_h is the comoving horizon distance.

6.2.2. Limber's equation

The density field δ is assumed to be a realization of a random field. It is the properties of the random field that cosmologists can predict. In particular, the second-order sta-

tistical properties of the density field are described in terms of the power spectrum. We shall therefore look at the relation between the quantities relevant for lensing and the power spectrum. The basis of this relation is Limber's equation. If δ is a homogeneous and isotropic 3-D random field, then the projections

$$g_i(\boldsymbol{\theta}) = \int dw q_i(w) \delta(f_K(w)\boldsymbol{\theta}, w) \quad (6.13)$$

also are (2-D) homogeneous and isotropic random fields, where the q_i are weight functions. In particular, the correlation function

$$C_{12} = \langle g_1(\boldsymbol{\varphi}_1) g_2(\boldsymbol{\varphi}_2) \rangle \equiv C_{12}(|\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2|) \quad (6.14)$$

depends only on the modulus of the separation vector. The original form of the Limber equation relates C_{12} to the correlation function of δ which is a line-of-sight projection. Alternatively, one can consider the Fourier-space version of this relation: The power spectrum $P_{12}(\ell)$ – the Fourier transform of $C_{12}(\theta)$ – depends linearly on the power spectrum $P_\delta(k)$ of the density fluctuations (Kaiser 1992),

$$\boxed{P_{12}(\ell) = \int dw \frac{q_1(w) q_2(w)}{f_K^2(w)} P_\delta\left(\frac{\ell}{f_K(w)}, w\right)}, \quad (6.15)$$

if the largest-scale structures in δ are much smaller than the effective range Δw of the projection. Hence, we obtain the (very reasonable) result that the power at angular scale $1/\ell$ is obtained from the 3-D power at length scale $f_K(w)$ ($1/\ell$), integrated over w . Comparing (5.8) with (6.15), one sees that $\kappa(\boldsymbol{\theta})$ is such a projection of δ with the weights $q_1(w) = q_2(w) = (3/2)(H_0/c)^2 \Omega_m g(w) f_K(w)/a(w)$, so that

$$\boxed{P_\kappa(\ell) = \frac{9H_0^4\Omega_m^2}{4c^4} \int_0^{w_h} dw \frac{g^2(w)}{a^2(w)} P_\delta\left(\frac{\ell}{f_K(w)}, w\right)}. \quad (6.16)$$

The power spectrum P_κ , if obtained through observations, can therefore be used to constrain the 3-D power spectrum P_δ .

6.3. Second-order cosmic shear measures

As we shall see, all second-order statistics of the cosmic shear yield (filtered) information about P_κ . The most-often used second-order statistics are:

- The two-point correlation function(s) of the shear, $\xi_\pm(\theta)$,
- the shear dispersion in a (circular) aperture, $\langle |\bar{\gamma}|^2 \rangle(\theta)$, and
- the aperture mass dispersion, $\langle M_{\text{ap}}^2 \rangle(\theta)$.

These will be discussed next, and their relation to $P_\kappa(\ell)$ shown. As a preparation, consider the Fourier transform of κ ,

$$\hat{\kappa}(\boldsymbol{\ell}) = \int d^2\theta e^{i\boldsymbol{\ell}\cdot\boldsymbol{\theta}} \kappa(\boldsymbol{\theta}); \quad (6.17)$$

then,

$$\langle \hat{\kappa}(\boldsymbol{\ell}) \hat{\kappa}^*(\boldsymbol{\ell}') \rangle = (2\pi)^2 \delta_D(\boldsymbol{\ell} - \boldsymbol{\ell}') P_\kappa(\ell), \quad (6.18)$$

which provides another definition of the power spectrum P_κ . The Fourier transform of the shear is

$$\hat{\gamma}(\boldsymbol{\ell}) = \left(\frac{\ell_1^2 - \ell_2^2 + 2i\ell_1\ell_2}{|\boldsymbol{\ell}|^2} \right) \hat{\kappa}(\boldsymbol{\ell}) \quad (6.19)$$

which implies that

$$\langle \hat{\gamma}(\ell) \hat{\gamma}^*(\ell') \rangle = (2\pi)^2 \delta_D(\ell - \ell') P_\kappa(\ell). \quad (6.20)$$

Hence, the power spectrum of the shear is the same as that of the convergence.

6.3.1. Shear correlation functions

Consider a pair of points (i.e., galaxy images); their separation direction φ (i.e. the polar angle of the separation vector $\boldsymbol{\theta}$) is used to define the tangential and cross-component of the shear at these positions *for this pair*,

$$\gamma_t = -\mathcal{R}\text{e}(\gamma e^{-2i\varphi}), \quad \gamma_x = -\mathcal{I}\text{m}(\gamma e^{-2i\varphi}). \quad (6.21)$$

Then, the shear correlation functions are defined as

$$\begin{aligned} \xi_{\pm}(\theta) &= \langle \gamma_t \gamma_t \rangle \pm \langle \gamma_x \gamma_x \rangle(\theta), \\ \xi_x(\theta) &= \langle \gamma_t \gamma_x \rangle(\theta). \end{aligned}$$

Due to parity symmetry, $\xi_x(\theta)$ is expected to vanish, since under such a transformation, $\gamma_t \rightarrow \gamma_t$, but $\gamma_x \rightarrow -\gamma_x$. Next we relate the shear correlation functions to the power spectrum P_κ : Using the definition of ξ_{\pm} , replacing γ in terms of $\hat{\gamma}$, and making use of relation between $\hat{\gamma}$ and $\hat{\kappa}$, one finds:

$$\boxed{\xi_+(\theta) = \int_0^\infty \frac{d\ell \ell}{2\pi} J_0(\ell\theta) P_\kappa(\ell); \quad \xi_-(\theta) = \int_0^\infty \frac{d\ell \ell}{2\pi} J_4(\ell\theta) P_\kappa(\ell)} \quad (6.22)$$

ξ_{\pm} can be measured as follows: on a data field, select all pairs of faint galaxies with separation within $\Delta\theta$ of θ and then take the average $\langle \epsilon_{ti} \epsilon_{tj} \rangle$ over all these pairs; since $\epsilon = \epsilon^{(s)} + \gamma(\boldsymbol{\theta})$, the expectation value of $\langle \epsilon_{ti} \epsilon_{tj} \rangle$ is $\langle \gamma_t \gamma_t \rangle(\theta)$, provided source ellipticities are uncorrelated. Similarly, the correlation for the cross-components is obtained.

6.3.2. The shear dispersion

Consider a circular aperture of radius θ ; the mean shear in this aperture is $\bar{\gamma}$. Averaging over many such apertures, one defines the shear dispersion $\langle |\bar{\gamma}|^2 \rangle(\theta)$. It is related to the power spectrum through

$$\boxed{\langle |\bar{\gamma}|^2 \rangle(\theta) = \frac{1}{2\pi} \int d\ell \ell P_\kappa(\ell) W_{\text{TH}}(\ell\theta)}, \quad \text{where } W_{\text{TH}}(\eta) = \frac{4J_1^2(\eta)}{\eta^2} \quad (6.23)$$

is the top-hat filter function. The shear dispersion can be measured by averaging the square of the mean galaxy ellipticities over many independent apertures.

6.3.3. The aperture mass

Consider a circular aperture of radius θ ; for a point inside the aperture, define the tangential and cross-components of the shear relative to center of aperture (as before); then define

$$M_{\text{ap}}(\theta) = \int d^2\vartheta Q(|\vartheta|) \gamma_t(\vartheta), \quad (6.24)$$

where Q is a weight function with support $\vartheta \in [0, \theta]$. In the following we shall use

$$Q(\vartheta) = \frac{6}{\pi\theta^2} \frac{\vartheta^2}{\theta^2} \left(1 - \frac{\vartheta^2}{\theta^2}\right) H(\theta - \vartheta),$$

in which case the dispersion of $M_{\text{ap}}(\theta)$ is related to the power spectrum by

$$\boxed{\langle M_{\text{ap}}^2 \rangle(\theta) = \frac{1}{2\pi} \int_0^\infty d\ell \ell P_\kappa(\ell) W_{\text{ap}}(\ell\theta)}, \quad \text{with } W_{\text{ap}}(\eta) := \frac{576J_4^2(\eta)}{\eta^4}. \quad (6.25)$$

6.3.4. Interrelations

These various 2-point statistics all depend linearly on the power spectrum P_κ ; therefore, one should not be too surprised that they are all related to each other. The surprise perhaps is that these interrelations are quite simple (Crittenden et al. 2002). First, the relations between ξ_\pm and P_κ can be inverted, making use of the orthonormality relation of Bessel functions:

$$P_\kappa(\ell) = 2\pi \int_0^\infty d\theta \theta \xi_+(\theta) J_0(\ell\theta) = 2\pi \int_0^\infty d\theta \theta \xi_-(\theta) J_4(\ell\theta). \quad (6.26)$$

Next, we take one of these and plug it into the relation between the other correlation function and P_κ , to find:

$$\boxed{\xi_+(\theta) = \xi_-(\theta) + \int_\theta^\infty \frac{d\vartheta}{\vartheta} \xi_-(\vartheta) \left(4 - 12 \frac{\vartheta^2}{\theta^2} \right);} \quad (6.27)$$

$$\boxed{\xi_-(\theta) = \xi_+(\theta) + \int_0^\theta \frac{d\vartheta}{\vartheta^2} \xi_+(\vartheta) \left(4 - 12 \frac{\vartheta^2}{\theta^2} \right).} \quad (6.28)$$

Using (6.26) in the expression for the shear dispersion, one finds

$$\boxed{\langle |\bar{\gamma}|^2 \rangle(\theta) = \int_0^{2\theta} \frac{d\vartheta}{\vartheta^2} \xi_+(\vartheta) S_+ \left(\frac{\vartheta}{\theta} \right) = \int_0^\infty \frac{d\vartheta}{\vartheta^2} \xi_-(\vartheta) S_- \left(\frac{\vartheta}{\theta} \right),} \quad (6.29)$$

where the S_\pm are simple functions, given explicitly in Schneider et al. (2002a). Finally, the same procedure for the aperture mass dispersion lets us write

$$\boxed{\langle M_{\text{ap}}^2 \rangle(\theta) = \int_0^{2\theta} \frac{d\vartheta}{\vartheta^2} \xi_+(\vartheta) T_+ \left(\frac{\vartheta}{\theta} \right) = \int_0^\infty \frac{d\vartheta}{\vartheta^2} \xi_-(\vartheta) T_- \left(\frac{\vartheta}{\theta} \right),} \quad (6.30)$$

again with known functions T_\pm (Schneider et al. 2002a). Hence, all these 2-point statistics can be evaluated from the correlation functions $\xi_\pm(\theta)$, which is of particular interest, since they can be measured best: Real data fields contain holes and gaps (like CCD defects, bright stars, nearby galaxies, etc.) which makes the placing of apertures difficult; however, the evaluation of the correlation functions is not affected by gaps, as one uses all pairs of galaxy images with a given angular separation.

6.4. Cosmic shear and cosmology

6.4.1. Why cosmology from cosmic shear?

Before continuing, it is worth to pause for a second and ask the question as to why one tries to investigate cosmological questions by using cosmic shear – since the CMB can measure cosmological parameters with high accuracy. Partial answers to this question are:

- Cosmic shear measures the mass distribution at much lower redshifts ($z \lesssim 1$) and at smaller physical scales [$R \sim 0.3 h^{-1} (\theta/1') \text{ Mpc}$] than the CMB; indeed, it is the only way to map out the dark matter distribution directly without any assumptions about the relation between dark and baryonic matter. The smaller scales probed are very important for constraining the shape of the power spectrum, i.e., the primordial tilt and the shape parameter Γ_{spect} .
- Cosmic shear measures the non-linearly evolved mass distribution and its associated power spectrum $P_\delta(k)$; hence, in combination with the CMB it allows us to study the evolution of the power spectrum and in particular, provides a very powerful test of the gravitational instability paradigm for structure growth.

- It provides a fully independent way to probe the cosmological model. Given the revolutionary claims coming from the CMB, SN Ia, and the LSS of the galaxy distribution, namely that more than 95% of the stuff in the Universe is in a form about whose physical nature we have not the slightest idea, an additional independent verification of these claims is certainly welcome.
- As we shall see shortly, cosmic shear studies provide a new and highly valuable search method for cluster-scale matter concentrations.

6.4.2. *Expectations*

The cosmic shear signal depends on the cosmological model, parametrized by Ω_m , Ω_Λ , and the shape parameter Γ_{spect} of the power spectrum, the normalization of the power spectrum, usually expressed in terms of σ_8 , and the redshift distribution of the sources. By measuring ξ_\pm over a significant range of angular scales one can derive constraints on these parameters.

The accuracy with which ξ_\pm can be measured depends on number density of galaxies (that is, depth and quality of the images), the total solid angle covered by the survey, and its geometric arrangement (compact survey vs. widely separated pointings); it is determined by a combination of intrinsic ellipticity dispersion and the cosmic (or sampling) variance. For angular scales below about 1 degree, the non-linear evolution of the power spectrum becomes important for the cosmic shear signal; because of this, the expected signal is considerably larger than estimated from linear perturbation theory of structure evolution. Furthermore, the signal depends quite strongly on the mean redshift of the source galaxies, which suggests that deep surveys, aiming for higher-redshift galaxies, are best suited for cosmic shear studies.

6.5. *Observation of cosmic shear*

6.5.1. *First detections*

Whereas the theory of cosmic shear was worked out in the early 1990's (Blandford et al. 1991; Miralda-Escudé 1991; Kaiser 1992), it took until the year 2000 before this effect was first discovered. The reason for this time lag must be seen as a combination of instrumental developments, i.e. the wide-field CCD mosaic cameras, and the image analysis software, like IMCAT, with which shapes of galaxies can be corrected for PSF effects. Finally, in March 2000, four groups independently published their first discoveries of cosmic shear (Bacon et al. 2000; Kaiser et al. 2000; van Waerbeke et al. 2000, Wittman et al. 2000). In these surveys, of the order of 10^5 galaxy images have been analyzed, covering about 1 deg². The fact that the results from these four independent teams agreed within the respective error bars immediately gave credence to this new window of observational cosmology. Furthermore, 4 different telescopes, 5 different cameras, independent data reduction tools and at least two different image analysis methods have been used in these studies. Maoli et al. (2001) reported a significant cosmic shear measurement from 50 widely separated FORS1@VLT images, which also agreed with the earlier measurements.

6.5.2. *Deriving constraints*

From the measured correlation functions $\xi_\pm(\theta)$, obtaining constraints on cosmological parameters can proceed through minimizing

$$\chi^2(p) = \sum_{ij} (\xi_i(p) - \xi_i^{\text{obs}}) \text{Cov}_{ij}^{-1} (\xi_j(p) - \xi_j^{\text{obs}}) , \quad (6.31)$$

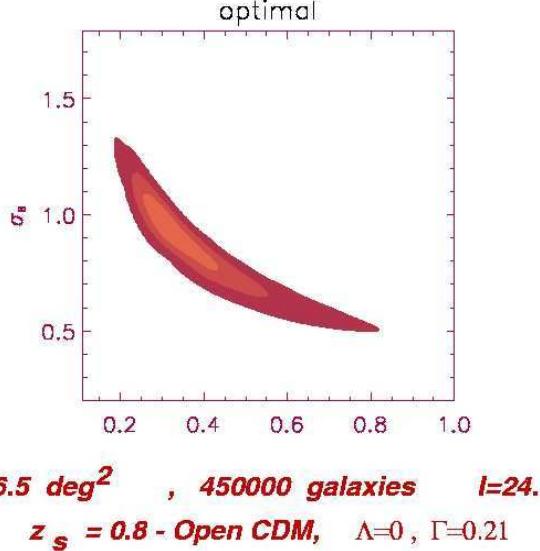


FIGURE 7. Constraints in the $\Omega_m - \sigma_8$ parameter plane, from the VIRIMOS-DESCART survey (van Waerbeke et al. 2001). For this figure, in which the 1, 2, and 3- σ confidence regions are indicated, a zero cosmological constant has been assumed, the redshift distribution of the source galaxies was assumed to be known, as well as the shape parameter $\Gamma_{\text{spect}} = 0.21$.

with $\xi_i = \xi(\theta_i)$ being the binned correlation function(s) (i.e., either ξ_{\pm} , or using both), p is a set of cosmological parameters, and Cov_{ij}^{-1} the inverse covariance matrix. The latter can be determined either from the ξ_{\pm} itself, from simulations, or estimated from the data (see Schneider et al. 2002b). Nevertheless, the calculation of the covariance is fairly cumbersome, and most authors have used approximate methods to derive it, such as the field-to-field variations of the measured correlation. As it turns out, $\xi_+(\theta)$ is strongly correlated across angular bins, much less so for $\xi_-(\theta)$; this is due to the fact that the filter function that describes ξ in terms of the power spectrum P_κ is much broader for ξ_+ (namely J_0) than J_4 which applies for ξ_- . Of course, a corresponding figure-of-merit function can be defined for the other second-order shear statistics, with their respective covariance matrices, but as argued before, the correlation functions should be regarded as the basic observable statistics.

6.5.3. Recent results

Since the first detections of cosmic shear, described above, there have been a large number of measurements over the past three years. Instead of mentioning them all here, we refer the reader to the recent reviews by van Waerbeke & Mellier (2003) and Refregier (2003). State-of-the-art are deep surveys, similar to those with which the first cosmic shear results have been derived, but with significantly larger solid angle (van Waerbeke et al. 2001, 2002), or shallower surveys of much larger area (e.g., Hoekstra et al. 2002c; Jarvis et al. 2003). The results of these surveys, which contain of the order of $\sim 10^6$ galaxies, i.e., an order-of-magnitude more than the discovery surveys mentioned above, can be summarized roughly as follows:

Cosmic shear by itself presently does not provide strong constraints on multi-dimensional cosmological parameter space. Hence, if one does not fix most of the cosmological parameters from external sources, the allowed region in multi-dimensional parameter space

is still quite large. On the other hand, if one considers a restricted set of parameters, cosmic shear results are very powerful. An example of that is given in Fig. 7, where all cosmological parameters have been kept fixed, except Ω_m and the normalization σ_8 . In this case, one finds indeed a well-defined maximum of the corresponding likelihood. Interestingly, the direction of the ‘likelihood valley’ nearly coincides with the constraint obtained from the cluster abundance, i.e., it provides a constraint on $\sigma_8 \Omega_m^{0.6}$. If the other cosmological parameters are not assumed to be known precisely, but are marginalized over a plausible uncertainty range, the likelihood contours widen substantially.

When combined with results from other methods, cosmic shear yields very useful information. For example, as pointed out by Hu & Tegmark (1999), when combined with data from CMB anisotropy, cosmic shear can break degeneracies of model parameters which are present when using the CMB data alone. In the $\Omega_m - \sigma_8$ parameter plane, cosmic shear constraints are nearly perpendicular to those from the CMB (van Waerbeke et al. 2002).

Therefore, at present the best use of cosmic shear results is in constraining the normalization σ_8 of the density perturbations, for a set of other cosmological parameters fixed by other methods, such as the CMB, galaxy redshift surveys, etc. The various cosmic shear surveys have given a range of σ_8 determinations which is about as narrow as current estimates from the abundance of massive clusters (see van Waerbeke & Mellier 2003 for a summary of these results). Given the youth of this field, this indeed is a remarkable achievement already. Furthermore, since the determination of σ_8 from cluster abundance and cosmic shear agree, one learns something important: the cluster abundance depends on the assumed Gaussianity of the primordial density field, whereas the constraint from cosmic shear does not. Hence, the agreement between the two methods supports the idea of an initial Gaussian field. Without doubt, the next generation of cosmic shear surveys will provide highly accurate determinations of this normalization, as well as other (combinations of) cosmological parameters.

6.6. *E-modes, B-modes*

In the derivation of the lensing properties of the LSS, we ended up with an equivalent surface mass density. In particular, this implied that \mathcal{A} is a symmetric matrix, that the shear can be obtained in terms of κ or ψ . Now, the shear is a 2-component quantity, whereas both κ and ψ are scalar fields. This implies that the two shear components are not independent of each other!

Recall that (5.7) yields a relation between the gradient of κ and the first derivatives of the shear components; in particular, (5.7) implies that $\nabla \times \mathbf{u}_\gamma \equiv 0$, yielding a local constraint relation between the shear components. The validity of this constraint equation guarantees that the imaginary part of (5.4) vanishes. This constraint is also present at the level of 2-point statistics, since one expects from (6.26) that

$$\int_0^\infty d\theta \theta \xi_+(\theta) J_0(\theta\ell) = \int_0^\infty d\theta \theta \xi_-(\theta) J_4(\theta\ell). \quad (6.32)$$

Hence, the two correlation functions ξ_\pm are not independent. The observed shear field is not guaranteed to satisfy these relations, due to noise, remaining systematics, or other effects. Therefore, searching for deviations from this relation allows a check for these effects. However, there might also be a ‘shear’ component present that is not due to lensing (by a single equivalent thin matter sheet κ). Shear components which satisfy the foregoing relations are called E-modes; those which don’t are B-modes – these names are exported from the polarization of the CMB, which has the same mathematical properties as the shear field.

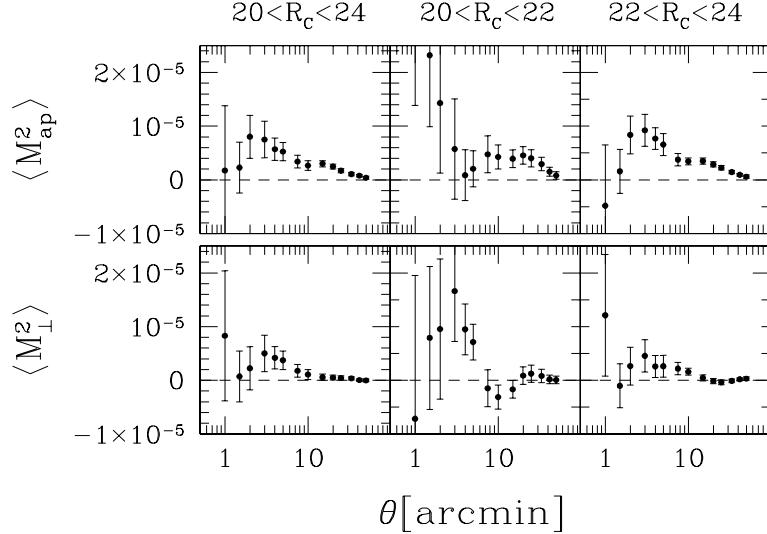


FIGURE 8. Dispersion of the aperture mass $\langle M_{\text{ap}}^2 \rangle$ (upper row) and its analogue $\langle M_{\perp}^2 \rangle$ for the cross-component, as obtained from the Red Cluster Sequence survey (Hoekstra et al. 2002). The left panels show the results for a broad range of galaxy brightnesses, whereas the middle and right rows display the results for the bright and the fainter parts, respectively, of the sample. Clearly, the presence of the B-mode is seen; its strength decreases for the fainter part of the sample; this behavior is expected if the B-mode is due to intrinsic alignments of galaxies. Its relative importance decreases with increasing width of the redshift distribution of galaxies

The best way to separate these modes locally is provided by the aperture measures: $\langle M_{\text{ap}}^2(\theta) \rangle$ is sensitive *only* to E-modes. If one defines in analogy

$$M_{\perp}(\theta) = \int d^2\vartheta Q(|\vartheta|) \gamma_{\times}(\vartheta), \quad (6.33)$$

then $\langle M_{\perp}^2(\theta) \rangle$ is sensitive *only* to B-modes.

Significant B-modes have been discovered in cosmic shear surveys (e.g., van Waerbeke et al. 2002; Hoekstra et al. 2002 – see Fig. 8); the question now is what are they due to? As mentioned before, the noise, which contributes to both E- and B-modes in similar strengths, could be underestimated, there could be remaining systematic effects, or indeed show the real presence of a B-mode on the sky. There are two possibilities known to generate a B-mode through lensing: The first-order in Φ (or ‘Born’) approximation may not be strictly valid, but as shown by ray-tracing simulations through cosmic matter fields (e.g., Jain et al. 2000) the resulting B-modes are expected to be very small. Clustering of sources also yields a finite B-mode (Schneider et al. 2002a), but again, this effect is much smaller than the observed amplitude of the B-modes.

Currently the best guess for the generation of a finite B-mode are intrinsic correlations of galaxy ellipticities. Such intrinsic alignments of galaxy ellipticities can be caused by the tidal gravitational field of the large-scale structure at galaxy formation. Predictions of the alignment of the projected ellipticity of the galaxy mass can be made analytically (e.g. tidal torque theory) or from numerical simulations; however, the predictions from various groups differ by large factors (e.g., Croft & Metzler 2000; Crittenden et al. 2001; Heavens et al. 2000; Jing 2002) which means that the process is not well understood at present. In addition, there remains the question of whether the orientation of the galaxy light (which is the issue of relevance here) is the same as that of the mass.

If intrinsic alignments play a role, then

$$\xi_+ = \langle \epsilon_i \epsilon_j^* \rangle = \left\langle \epsilon_i^{(s)} \epsilon_j^{(s)*} \right\rangle + \xi_+^{\text{lens}}, \quad (6.34)$$

and measured correlations ξ_{\pm} contain both components. Of course, there is no reason why intrinsic correlations should have only an B-mode. If a B-mode contribution is generated through this process, then the measured E-mode is also contaminated by intrinsic alignments. In fact, the various models do not agree on the relative strength of E- and B-modes in the intrinsic alignments of galaxies, but it seems that the E-modes have generally higher amplitude than the B-modes. Given that intrinsic alignments yield ellipticity correlations only for spatially close sources (i.e., close in 3-D, not merely in projection), it is clear that the deeper a cosmic shear survey is, and thus the broader the redshift distribution, the smaller is the relative amplitude of an intrinsic signal. Most of the theoretical predictions on the strength of intrinsic alignments say that the deep cosmic shear surveys (say, with mean source redshifts of $\langle z_s \rangle \sim 1$) are affected at a $\sim 10\%$ level, but that shallow cosmic shear surveys are more strongly affected; for them, the intrinsic alignment can be of the same order as, or larger than the lensing signal.

However, the intrinsic signal can be separated from the lensing signal if redshift information of the sources is available, owing to the fact that $\langle \epsilon_i^{(s)} \epsilon_j^{(s)*} \rangle$ will be non-zero only if the two galaxies are at the same redshift. Hence, if z -information is available (e.g., photometric redshifts), then galaxy pairs which are likely to have similar redshifts are to be avoided in estimating the cosmic shear signal (King & Schneider 2002; Heymans & Heavens 2003). This will change the expectation value of the shear correlation function, but in a controllable way, as the redshifts are assumed to be known. Indeed, using (photometric) redshifts, one can simultaneously determine the intrinsic and the lensing signal, essentially providing a cosmic shear tomography (King & Schneider 2003).

6.7. Higher-order statistics

On the level of second-order statistics, ‘only’ the power spectrum is probed. If the density field was Gaussian, then the power spectrum would fully characterize it; however, in the course of non-linear structure evolution, non-Gaussian features of the density field are generated, which show up correspondingly in the cosmic shear field and which can be probed by higher-order shear statistics. The usefulness of these higher-order measures for cosmic shear has been pointed out in Bernardeau et al. (1997) and van Waerbeke et al. (1999); in particular, the near-degeneracy between σ_8 and Ω_m can be broken. However, these are serious problems with higher-order shear statistics, that shall be illustrated in terms of the third-order statistics. The three-point correlation function has three independent variables (e.g. the sides of a triangle) and 8 components; as was shown in Schneider & Lombardi (2003), none of these eight components vanishes owing to parity invariance. This then implies that the covariance matrix has 6 arguments and 64 components! Of course, this is too difficult to handle efficiently, and therefore one must ask which combinations of the components of the 3-pt correlation function are most useful for studying the dark matter distribution. Unfortunately, this is essentially unknown yet. An additional problem is that the predictions from theory are less well established than for the second-order statistics.

Nevertheless, progress has been made. From ray-tracing simulations through a cosmic matter distribution, the 3-pt correlation function of the shear can be determined (Takada & Jain 2003; see also Zaldarriaga & Scoccimarro 2003); in addition, Schneider & Lombardi (2003) have defined the ‘natural components’ of the 3-pt correlator which are most easily related to the bispectrum of the underlying matter distribution.

Alternatively, aperture measures can be defined to measure the third-order statistics. Schneider et al. (1998) calculated $\langle M_{\text{ap}}^3 \rangle(\theta)$ in the frame of the quasi-linear structure evolution model and showed it to be a strong function of Ω_m . Indeed, $\langle M_{\text{ap}}^3 \rangle$ is sensitive only to the E-modes of the shear field. One might be tempted to use $\langle M_{\perp}^3 \rangle(\theta)$ as a measure for third-order B-mode statistics, but indeed, this quantity vanishes owing to parity invariance. However, $\langle M_{\perp}^2 M_{\text{ap}} \rangle$ is a measure for the B-modes at the third-order statistical level. Bernardeau et al. (2002) measured for the first time a significant 3-rd order shear from the VIRMOS-DESCART survey, employing a suitably filtered integral over the measured 3-pt correlation function. With the upcoming large cosmic shear surveys, the 3-pt function will be measured with high accuracy.

6.8. Weak lensing search for cluster-mass dark halos

As we have seen, the mass distribution of clusters of galaxies can be mapped by weak lensing techniques. In fact, the coherent alignment of background galaxy images clearly shows the presence of a massive matter concentration present at or near the location of the optically or X-ray selected cluster towards which the weak lensing observations were targeted. As pointed out in Schneider (1996), one can use weak lensing to search for clusters: seeing a strong alignment of galaxy images centered onto a point on a wide-field image, one would conclude the presence of the mass concentration there. A very useful way to quantify this is the aperture mass statistics, already introduced. By selecting an appropriate filter function, one can systematically seach for statistically significant peaks of M_{ap} on wide-field images. In fact, the data needed for this investigation is the same as that used in cosmic shear surveys.

Since clusters of galaxies are very important cosmological probes, e.g., to determine the normalization of the power spectrum of the matter inhomogeneities in the Universe, a selection of clusters based on their mass properties only would be extremely useful. Usually, clusters are selected by their optical or X-ray properties; to transform luminosity or X-ray temperature into a mass estimate, and thus to transform a flux-limited cluster sample into a mass-limited sample, which can then be compared to cosmological predictions, one needs to employ a number of approximations and scaling relations. In contrast to this, the shear selection can directly be compared to cosmological predictions, e.g., by calculating the abundance of peaks of M_{ap} directly from N-body simulations of structure formation (e.g., Reblinsky et al. 1999), without reliance on the luminous properties of baryonic matter, nor even for identifying cluster-mass halos in the simulated density fields.

The abundance of peaks above a given threshold M_{ap} , at a given angular scale, can also be used as a cosmic shear measure. In fact, in future large cosmic shear surveys this will become most likely one of the most useful statistics for studying non-Gaussian aspects of the shear field. Within the frame of Press–Schechter theory, Kruse & Schneider (1999) calculated the M_{ap} peak statistics, which was then compared with direct numerical simulation by Reblinsky et al. (1999). White et al. (2002) pointed out that the M_{ap} statistics can be substantially affected by the large-scale structure along the line-of-sight to these mass concentrations; this implies that the relation between M_{ap} and the mass of the clusters is not simple – but again, this method does not require a mass function to be determined, as the M_{ap} -statistics can be obtained directly from LSS simulations.

Several clusters, or cluster candidates, have been found that way. Erben et al. (2000) detected a highly significant shear signal corresponding to a putative mass peak, about $7'$ away from the cluster Abell 1942, seen on two images taken with different filters and different cameras. No obvious concentration of galaxies is seen in this direction, neither in the optical nor near-IR images (Gray et al. 2001), making it a candidate for a ‘dark

clump'; however, before drawing this conclusion, further investigations are needed, such as imaging with the HST to confirm the shear measurements. Umetsu & Futamase (2000) found a significant mass concentration on an HST image, again without an obvious optical counterpart. In contrast to this, Mellier et al. (2000) found a mass peak in one of their 50 VLT fields taken for a cosmic shear survey with FORS, which is clearly associated with a concentration of galaxies. Wittman et al. (2001, 2002) detected two clusters on their wide-field images, and confirmed them spectroscopically. In fact, making use of photometric redshift estimates of the background galaxies, and employing the redshift dependence of the lens strength, they were able to estimate rather precisely the cluster redshifts, which were later confirmed with spectroscopy. Two of the three putative mass concentrations found by Dahle et al. (2002) are also very likely to be associated with luminous clusters. Hence, the shear selection of clusters has already been proven as a very useful concept.

6.9. *Lensing in three dimensions*

Using (photometric) redshift estimates in cosmic shear research is not only useful to remove the potential contribution from intrinsic alignments of galaxy ellipticities. If one defines galaxy populations with different redshift distributions, one can probe different projections of the cosmic density field; see (5.8). This then increases the information one can extract from a cosmic shear survey, and thus the ability to discriminate between different cosmological models (Hu 1999).

More recently, it was pointed out (Taylor 2001; Hu & Keeton 2002) that the use of redshift information can in principle be used to reconstruct the three-dimensional density field δ from the shear measurements. This is based on the possibility to invert (5.8), i.e., to express $\delta(w)$ in terms of $\kappa(w)$. The study of the 3-dimensional mass distribution is particularly interesting for constraining the properties of the dark energy in the Universe (e.g., Heavens 2003).

7. Conclusions

Due to its insensitivity to the nature of matter causing the gravitational potential, gravitational lensing has turned out to be an ideal tool to probe the structure of the (dark) matter distribution in the Universe, from small to large scales. Progress in this field has been very rapid in the past years, and due to the fast pace at which new instruments become available, it is guaranteed to continue its role as an important tool for observational cosmology. For example, the first images from the new camera ACS are breathtaking and will certainly lead to much improved mass models of clusters of galaxies. The new square degree optical cameras will provide cosmic shear surveys covering an appreciable fraction of the sky – as a consequence, statistical uncertainties and cosmic variance will then no longer be the main contribution to the error budget, but systematic effects in estimating shear from CCD images may well take over. The Next Generation Space Telescope will provide a superb tool for studying galaxy-scale lens systems, as well as clusters.

Acknowledgement

I would like to thank John Beckmann for the invitation to this School; being there was a very nice experience. A big thanks to him and his colleagues for the great organization of the School and the associated cultural events. The active participation of the students and their curiosity was very much appreciated. I would like to thank my colleagues and

friends with whom I had the pleasure to discuss the topics of this article over the years, in particular Matthias Bartelmann, Doug Clowe, Thomas Erben, Lindsay King (special thanks to her for carefully reading and commenting this manuscript), Marco Lombardi, Yannick Mellier, Ludovic van Waerbeke, my past and current students, as well as Chris Kochanek and Joachim Wambsganss with whom I had the fun to teach (and learn!) for a week at the 2003 Saas-Fee course. This work was supported by the German Ministry for Science and Education (BMBF) through the DLR under the project 50 OR 0106, by the German Ministry for Science and Education (BMBF) through DESY under the project 05AE2PDA/8, and by the Deutsche Forschungsgemeinschaft under the project SCHN 342/3-1.

REFERENCES

- Bacon, D.J., Refregier, A.R. & Ellis, R.S. 2000, MNRAS, 318, 625
 Bacon, D.J., Refregier, A., Clowe, D. & Ellis, R.S. 2001, MNRAS 325, 1065
 Bartelmann, M. 1995, A&A 299, 11
 Bartelmann, M., Narayan, R., Seitz, S. & Schneider, P. 1996, ApJ 464, L115
 Bartelmann, M. & Schneider, P. 2001, Physics Reports 340, 291
 Bernardeau, F., Mellier, Y. & van Waerbeke, L. 2002, A&A 389, L28
 Bernardeau, F., Van Waerbeke, L. & Mellier, Y. 1997, A&A, 322, 1
 Bernstein, G.M. & Jarvis, M. 2002, AJ 123, 583
 Binney, J. & Tremaine, S. 1987, *Galactic dynamics*, Princeton University Press, Princeton.
 Blandford, R.D., Saust, A.B., Brainerd, T.G. & Villumsen, J.V. 1991, MNRAS, 251, 600
 Bonnet, H. & Mellier, Y. 1995, A&A 303, 331
 Bradac, M., Schneider, P., Steinmetz, M., Lombardi, M., King, L.J. & Porcas, R. 2002, A&A 388, 373
 Bradac, M., Schneider, P., Lombardi, M. et al. 2003, astro-ph/0306238
 Brainerd, T.G., Blandford, R.D. & Smail, I. 1996, ApJ 466, 623
 Broadhurst, T.J., Taylor, A.N. & Peacock, J.A. 1995, ApJ 438, 49
 Burke, W.L. 1981, ApJ 244, L1
 Carlberg, R.G., Yee, H.K.C. & Ellingson, E. 1994, ApJ 437, 63
 Catelan, P., Kamionkowski, M. & Blandford, R.D. 2001, MNRAS, 320, L7
 Clowe, D., Luppino, G.A., Kaiser, N. & Gioia, I.M. 2000 ApJ 539, 540
 Clowe, D. & Schneider, P. 2001, A&A 379, 384
 Clowe, D. & Schneider, P. 2002, A&A 395, 385
 Colley, W.N., Tyson, J.A. & Turner, E.L. 1996, ApJ 461, L83
 Crittenden, R.G., Natarajan, P., Pen, U.-L. & Theuns, T. 2001, ApJ, 559, 552
 Crittenden, R.G., Natarajan, P., Pen, U.-L. & Theuns, T. 2002, ApJ, 568, 20
 Croft, R.A.C. & Metzler, C.A. 2001, ApJ, 545, 561
 Czoske, O., Moore, B., Kneib, J.-P. & Soucail, G. 2002, A&A 386, 31
 Dahle, H., Pedersen, K., Lilje, P.B., Maddox, S.J. & Kaiser, N. 2002, astro-ph/0208050
 Dahle, H., Hannestad, S. & Sommer-Larsen, J. 2003, ApJ 388, L73
 Dalal, N. & Kochanek, C.S. 2002, ApJ 572, 25
 Dye, S., Taylor, A.N., Greve, T.R. et al. 2002, A&A 386, 12
 Ebbels, T., Ellis, R., Kneib, J.-P. et al. 1998, MNRAS 295, 75
 Erben, T., van Waerbeke, L., Mellier, Y., Schneider, P., Cuillandre, J.C., Castander, F.J. & Dantel-Fort, M. 2000, A&A 355, 23

- Erben, T., van Waerbeke, L., Bertin, E., Mellier, Y. & Schneider, P. 2001, A&A 366, 717
- Fahlman, G., Kaiser, N., Squires, G. & Woods, D. 1994, ApJ 437, 56
- Falco, E.E., Gorenstein, M.V. & Shapiro, I.I. 1985, ApJ 289, L1
- Falco, E.E., Impey, C.D., Kochanek, C.S. et al. 1999, ApJ 523, 617
- Fischer, P. 1999, AJ 117, 2024
- Fischer, P., McKay, T.A., Sheldon, E. et al. 2000, AJ 120, 1198
- Fort, B. & Mellier, Y. 1994, A&A Review 5, 239
- Fort, B., Mellier, Y. & Dantel-Fort, M. 1997, A&A 321, 353
- Freedman, W.L., Madore, B.F., Gibson, B.K. et al. 2001, ApJ 553, 47
- Gavazzi, R., Fort, B., Mellier, Y., Pelló, R. & Dantel-Fort, M. 2003, A&A 403, 11
- Gray, M.E., Ellis, R.S., Lewis, J.R., McMahon, R.G. & Firth, Andrew E. 2001, MNRAS 325, 111
- Gray, M.E., Taylor, A.N., Meisenheimer, K., Dye, S., Wolf, C. & Thommes, E. 2002, ApJ 568, 141
- Guzik, J. & Seljak, U. 2002, MNRAS 335, 311
- Heavens, A.F. 2003, astro-ph/0304151
- Heavens, A.F., Refregier, A. & Heymans, C.E.C. 2000, MNRAS, 319, 649
- Heymans, C. & Heavens, A. 2003, MNRAS 339, 711
- Hoekstra, H., Franx, M. & Kuijken, K. 2000, ApJ 532, 88
- Hoekstra, H., Franx, M., Kuijken, K. & van Dokkum, P.G. 2002a, MNRAS 333, 911
- Hoekstra, H., van Waerbeke, L., Gladders, M.D., Mellier, Y. & Yee, H.K.C. 2002b, ApJ 577, 604
- Hoekstra, H., Yee, H.K.C. & Gladders, M.D. 2002c, ApJ 577, 595
- Hu, W. 1999, ApJ 522, L21
- Hu, W. & Keeton, C.R. 2002, astro-ph/0205412
- Hu, W. & Tegmark, M. 1999, ApJ 514, L65
- Jain, B., Seljak, U. & White, S.D.M. 2000, ApJ, 530, 547
- Jarvis, M., Bernstein, G.M., Fischer, P. et al. 2003, AJ 125, 1014
- Jing, Y.P. 2002, MNRAS 335, L89
- Kaiser, N. 1992, ApJ, 388, 272
- Kaiser, N. 1995, ApJ 439, L1
- Kaiser, N. 2000, ApJ 537, 555
- Kaiser, N., Squires, G., 1993, ApJ, 404, 441
- Kaiser, N., Squires, G. & Broadhurst, T. 1995, ApJ 449, 460
- Kaiser, N., Wilson, G., Luppino, G. et al. 1998, astro-ph/9809268
- Kaiser, N., Wilson, G. & Luppino, G. 2000, astro-ph/0003338
- Keeton, C.R., Kochanek, C.S. & Falco, E.E. 1998, ApJ 509, 561
- King, L.J., Jackson, N., Blandford, R.D. et al. 1998, MNRAS 289, 450
- King, L., Clowe, D.I. & Schneider, P. 2002, A&A 383, 118
- King, L. & Schneider, P. 2002, A&A 396, 411
- King, L. & Schneider, P. 2003, A&A 398, 23
- Kneib, J.-P., Mathez, G., Fort, B., Mellier, Y., Soucail, G. & Longaretti, P.-Y. 1994, A&A 286, 701
- Kochanek, C.S. 1995, ApJ 445, 559
- Kochanek, C.S. 2002, ApJ 578, 25
- Kochanek, C.S. 2003, ApJ 583, 49
- Kochanek, C.S. & Dalal, N. 2003, astro-ph/0302036

- Kuijken, K. 1999, A&A 352, 355
- Koopmans, L.V.E. & Treu, T. 2003, ApJ 583, 606
- Kruse, G. & Schneider, P. 1999, MNRAS 302, 821
- Langston, G.I., Conner, S.R., Lehar, J., Burke, B.F. & Weiler, K.W. 1990, Nature 344, 43
- Lombardi, M. & Bertin, G. 1998, A&A 335, 1
- Luppino, G.A. & Kaiser, N. 1997, ApJ 475, 20
- Mao, S. & Schneider, P. 1998, MNRAS 295, 587
- McKay, T.A., Sheldon, E., Racusin, J. et al. 2001, astro-ph/0108013
- Mellier, Y. et al. 2000, ESO Messenger 102, 30
- Mellier, Y. 1999, ARA&A 37, 127
- Miralda-Escudé, J. 1991, ApJ 380, 1
- Maoli, R., van Waerbeke, L., Mellier, Y., et al. 2001, A&A, 368, 766
- Narayan, R. & Bartelmann, M. 1999, in: *Formation of Structure in the Universe*, A. Dekel & J.P. Ostriker (eds.), Cambridge University Press, p.360
- Navarro, J.F., Frenk, C.S. & White, S.D.M. 1997, ApJ 490, 493
- Paczynski, B. 1996, ARA&A 34, 419
- Petters, A.O., Levine, H. & Wambsganss, J. 2001, Singularity Theory and Gravitational Lensing (Boston: Birkhäuser)
- Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T. 1986, Numerical Recipes, (Cambridge: University Press)
- Reblinsky, K., Kruse, G., Jain, B. & Schneider, P. 1999, A&A 351, 815
- Refregier, A. 2003, ARA&A, submitted
- Refregier, A. & Bacon, D. 2003, MNRAS 338, 48
- Roulet, E. & Mollerach, S. 1997, Physics Report 279, 67
- Rusin, D., Kochanek, C.S., Falco, E.E. et al. 2003, ApJ 587, 143
- Sand, D.J., Treu, T. & Ellis, R.S. 2002, ApJ 574, L129
- Schechter, P.L. & Wambsganss, J. 2002, ApJ 580, 685
- Schmidt, R.W., Kundic, T., Pen, U.-L. et al. 2002, A&A 392, 773
- Schneider, P. 1984, A&A 140, 119
- Schneider P., 1996, MNRAS, 283, 83
- Schneider P., 1998, ApJ 498, 43
- Schneider, P., Ehlers, J. & Falco, E.E. 1992, Gravitational Lenses (New York: Springer)
- Schneider, P. & Lombardi, M. 2003, A&A 397, 809
- Schneider, P. & Seitz, C. 1995, A&A 294, 411
- Schneider P., van Waerbeke L., Jain B., Kruse G., 1998, MNRAS, 296, 873
- Schneider, P., van Waerbeke, L. & Mellier, Y. 2002a, A&A 389, 729
- Schneider, P., van Waerbeke, L., Kilbinger, M. & Mellier, Y. 2002b, A&A 396, 1
- Schramm, T. & Kayser, R. 1995, A&A 299, 1
- Seitz, C., Kneib, J.-P., Schneider, P. & Seitz, S. 1996, A&A 314, 707
- Seitz, C. & Schneider, P. 1997, A&A 318, 687
- Seitz, S. & Schneider, P. 1992, A&A 265, 1
- Seitz, S. & Schneider, P. 2001, A&A, 374, 740
- Seitz, S., Schneider, P. & Bartelmann, M. 1998, A&A 337, 325
- Squires, G., Kaiser, N., Babul, A. et al. 1996, ApJ 461, 572
- Takada, M. & Jain, B. 2003, MNRAS 340, 580
- Taylor, A.N. 2001, astro-ph/0111605

- Taylor, A.N., Dye, S., Broadhurst, T.J., Benítez, N. & van Kampen, E. 1998, ApJ 501, 539
- Treu, T. & Koopmans, L.V.E. 2002, MNRAS 337, L6
- Umetsu, K. & Futamase, T. 2000, ApJ 539, L5
- van Waerbeke, L. 1998, A&A 334, 1
- van Waerbeke, L. 2000, MNRAS, 313, 524
- van Waerbeke, L., Bernardeau, F. & Mellier, Y. 1999, A&A, 243, 15
- Van Waerbeke, L., Mellier, Y., Erben, T. et al. 2000, A&A, 358, 30
- van Waerbeke, L. & Mellier, Y. 2003, astro-ph/0305089
- van Waerbeke, L., Mellier, Y., Pelló, R., Pen, U.-L., McCracken, H.J. & Jain, B. 2002, A&A 393, 369
- Wallington, S., Kochanek, C.S. & Narayan, R. 1996, ApJ 465, 64
- Wambsganss, J. 1998, Living Reviews in Relativity, Vol. 1
- Wambsganss, J., Paczyński, B. & Schneider, P. 1990, ApJ 358, L33
- White, M., van Waerbeke, L. & Mackey, J. 2002, ApJ 575, 640
- Williams, R.E., Blacker, B., Dickinson, M. et al. 1996, AJ 112, 1335
- Wilson, G., Kaiser, N. & Luppino, G.A. 2001, ApJ 556, 601
- Wittman, D.M., Tyson, J.A., Kirkman, D., Dell'Antonio, I. & Bernstein, G. 2000, Nat, 405, 143
- Wittman, D., Tyson, J.A., Margoniner, V.E., Cohen, J.G. & Dell'Antonio, I.P. 2001, ApJ 557, L89
- Wittman, D., Margoniner, V.E., Tyson, J.A., Cohen, J.G. & Dell'Antonio, I.P. 2002, astro-ph/0210120
- Wittman, D. 2002, in Lecture Notes in Physics 608, Gravitational Lensing: An Astrophysical Tool, ed. F. Courbin & D. Minniti (Berlin: Springer), 55
- Zaldarriaga, M. & Scoccimarro, R. 2002, astro-ph/0208075

INTRODUCTION

Martin J Rees

Institute of Astronomy, Madingley Road, Cambridge, CB3 0HA

Abstract

It is embarrassing that 95% of the universe is unaccounted for. Galaxies and larger-scale cosmic structures are composed mainly of ‘dark matter’ whose nature is still unknown. Favoured candidates are weakly-interacting particles that have survived from the very early universe, but more exotic options cannot be excluded. (There are strong arguments that the dark matter is not composed of baryons). Intensive experimental searches are being made for the ‘dark’ particles (which pervade our entire galaxy), but we have indirect clues to their nature too. Inferences from galactic dynamics and gravitational lensing allow astronomers to ‘map’ the dark matter distribution; comparison with numerical simulations of galaxy formation can constrain (eg) the particle velocities and collision cross sections. And, of course, progress in understanding the extreme physics of the ultra-early universe could offer clues to what particle might have existed then, and how many would have survived.

The mean cosmic density of dark matter (plus baryons) is now pinned down to be only about 30% of the so-called critical density corresponding to a ‘flat’ universe. However, other recent evidence – microwave background anisotropies, complemented by data on distant supernovae – reveals that our universe actually is ‘flat’, but that its dominant ingredient (about 70% of the total mass-energy) is something quite unexpected — ‘dark energy’ pervading all space, with negative pressure. We now confront two mysteries:

- (i) Why does the universe have three quite distinct basic ingredients – baryons, dark matter and dark energy – in the proportions (roughly) 5%, 25% and 70%?
- (ii) What are the (almost certainly profound) implications of the ‘dark energy’ for fundamental physics?

1 SOME HISTORY

Astronomers have long known that galaxies and clusters would fly apart unless they were held together by the gravitational pull of much more material than we actually see.

The strength of the case built up gradually. The argument that clusters of galaxies would be unbound without dark matter dates back to Zwicky (1937) and others in the 1930s. Kahn and Woltjer (1959) pointed out that the motion of Andromeda towards us implied that there must be dark matter in our Local

Group of galaxies. But the dynamical evidence for massive halos (or ‘coronae’) around individual galaxies firmed up rather later (e.g. Roberts and Rots 1973, Rubin, Thonnard and Ford 1978).

Two 1974 papers were specially influential in the latter context. Here is a quote from each:

The mass of galactic coronas exceeds the mass of populations of known stars by one order of magnitude, as do the effective dimensions. The mass/luminosity ratio rises to $f=100$ for spiral and $f = 120$ for elliptical galaxies. With $H = 50 \text{ km/sec/Mpc}$ this ratio for the Coma cluster is 170 (Einasto, Kaasik and Saar 1974)

Currently-available observations strongly indicate that the mass of spiral galaxies increases almost linearly with radius to nearly 1 Mpc.... and that the ratio of this mass to the light within the Holmberg radios, f , is 200 (M/L_\odot). (Ostriker, Peebles and Yahil, 1974).

The amount of dark matter, and how it is distributed, is now far better established than it was when those papers were written. The immense advances in delineated dark matter in clusters and in individual galaxies are manifest in the programme for this meeting. The rapid current progress stems from the confluence of several new kinds of data within the same few-year interval: optical surveys of large areas and high redshifts, CMB fluctuation measurements, sharp X-ray images, and so forth.

The progress has not been solely observational. Over the last 20 years, a compelling theoretical perspective for the emergence of cosmic structure has been developed. The expanding universe is unstable to the growth of structure, in the sense that regions that start off *very slightly* overdense have their expansion slowed by their excess gravity, and evolve into conspicuous density contrasts. According to this ‘cold dark matter’ (CDM) model, the present-day structure of galaxies and clusters is moulded by the gravitational aggregation of non-baryonic matter, which is an essential ingredient of the early universe (Pagels and Primack 1982, Peebles 1982, Blumenthal *et al.* 1984, Davis *et al.* 1985). These models have been firmed up by vastly improved simulations, rendered possible by burgeoning computer power. And astronomers can now compare these ‘virtual universes’ with the real one, not just at the present era but (by observing very distant objects) can probe back towards the formative stages when the first galaxies emerged.

The following comments are intended to provide a context for the later papers. (For that reason, I do not give detailed references to the topics covered by other speakers – just some citations of historical interest).

2 THE CASE FOR DARK MATTER

2.1 Baryons

The inventory of cosmic baryons is readily compiled. Stars and their remnants, and gas in galaxies, contribute no more than 1% of the critical density (i.e. they give $\Omega_b < 0.01$). However several percent more could be contributed by diffuse material pervading intergalactic space: warm gas (with $kT \simeq 0.1$ keV) in groups of galaxies and loose clusters, and cooler gas pervading intergalactic space that manifests itself via the ‘picket fence’ absorption lines in quasar spectra. (Rich clusters are rare, so their conspicuous gas content, at several KeV, is not directly significant for the total inventory, despite its importance as a probe)

These baryon estimates are concordant with those inferred by matching the He and D abundances at the birth of galaxies with the predicted outcome of nucleosynthesis in the big bang, which is sensitive to the primordial baryon/photon ratio, and thus to Ω_b . The observational estimates have firmed up, with improved measurements of deuterium in high- z absorbing clouds. The best fit occurs for $\Omega_b \simeq 0.02h^{-2}$ where h is the Hubble constant in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$. Observations favour $h \simeq 0.7$.

Ω_b is now pinned down by a variety of argument to be $0.04 - 0.05$. This corresponds to only ~ 0.3 baryons per cubic metre, a value so low that it leaves little scope for dark baryons. (It is therefore unsurprising that the MACHO/OGLE searches should have found that compact objects do not make a substantial contribution to the total mass of our own galactic halo.)

2.2 How much dark matter?

An important recent development is that Ω_{DM} can now be constrained to a value around 0.25 by several independent lines of evidence:

- (i) One of the most ingenious and convincing arguments comes from noting that baryonic matter in clusters – in galaxies, and in intracluster gas – amounts to $0.15 - 0.2$ of the inferred virial mass (White *et al.* 1993). If clusters were a fair sample of the universe, this would then be essentially the same as the cosmic ratio of baryonic to total mass. Such an argument could not be applied to an individual galaxy, because baryons segregate towards the centre. However, there is no such segregation on the much larger scale of clusters: only a small correction is necessary to allow for baryons expelled during the cluster formation process.

(ii) Very distant galaxies appear distorted, owing to gravitational lensing by intervening galaxies and clusters. Detailed modelling of the mass-distributions needed to cause the observed distortions yields a similar estimate. This is a straight measurement of Ω_{DM} which (unlike (i)) does not involve assumptions about Ω_b , though it does depend on having an accurate measure of the clustering amplitude.

(iii) Another argument is based on the way density contrasts grow during the cosmic expansion: in a low density universe, the expansion kinetic energy overwhelms gravity, and the growth of structure saturates at recent epochs. The existence of conspicuous clusters of galaxies with redshifts as large as $z = 1$ is hard to reconcile with the rapid recent growth of structure that would be expected if Ω_{DM} were unity. More generally, numerical simulations based on the cold dark matter (CDM) model are a better fit to the present-day structure for this value of Ω_{DM} (partly because the initial fluctuation spectrum has too little long-wavelength power if Ω_{DM} is unity).

Other methods will soon offer independent estimates. For instance, Ω_{DM} can be estimated from the deviations from the Hubble flow induced by large-scale irregularities in the mass distribution on supercluster scales.

2.3 What could the dark matter be?

The dark matter is not primarily baryonic. The amount of deuterium calculated to emerge from the big bang would be far lower than observed if the average baryon density were ~ 2 (rather than ~ 0.3) per cubic metre. Extra exotic particles that do not participate in nuclear reactions, however, would not scupper the concordance.

Beyond the negative statement that it is non-baryonic, the nature of the dark matter still eludes us. This key question may yield to a three-pronged attack:

1. Direct detection. As described by other contributors to this meeting, several groups are developing cryogenic detectors for supersymmetric particles and axions. This is an exciting quest. Of course, not even optimists can be confident that the actual dark matter particles have parameters within the range that these experiments are yet sensitive to. But the stakes are high: detection of most of the gravitating stuff in the universe, as well as a new class of elementary particle. So it seems well worth committing to these experiments funding that is equivalent to a small fraction of the cost of a major accelerator.

2. Progress in particle physics. Important recent measurements suggest that neutrinos have non-zero masses; this result has crucially important implications for physics beyond the standard model. The inferred neutrino masses seem, how-

ever, too low to be cosmologically important. If the masses and cross-sections of supersymmetric particles were known, it should be possible to predict how many survive, and their contribution to Ω , with the same confidence with which we can compute the nuclear reactions that control primordial nucleosynthesis. Associated with such progress, we might expect a better understanding of how the baryon-antibaryon asymmetry arose, and the consequence for Ω_b . Optimists may hope for progress on still more exotic options.

3. Simulations of galaxy formation and large-scale structure. When and how galaxies form, the way they are clustered, and the density profiles within individual systems, depend on what their gravitationally-dominant constituent is. A combination of better data and better simulations is starting to set generic constraints on the options. The CDM model works well. But there are claimed discrepancies, though many of us suspect these may ease when the galaxy formation process is better understood. For instance the centre of a halo would, according to the simulations, have a ‘cusp’ rather than the measured uniform-density core: this discrepancy has led some authors to explore modifications where the particles are assumed to have significant collision probabilities, or to be moving with non-negligible velocities (i.e. ‘warm’ not cold.). These calculations are in any case offering interesting constraints on the properties of heavy supersymmetric particles. (Also, straight astronomical observations can rule out a contribution to Ω of more than 0.01 from neutrinos – this is compatible with current experimental estimates.)

3 DARK ENERGY

The inference that our universe is dominated by dark matter is in itself a discovery of the first magnitude. But the realisation that even more mass-energy is in some still more mysterious form – dark energy latent in space itself – came as a surprise, and probably has even greater import for fundamental physics.

If this meeting had been taking place 3 years ago, the more open-minded among us would have given equal billing to two options: a hyperbolic universe with Ω of 0.3, (in which it would be a coincidence that the Robertson-Walker curvature radius was comparable with the present Hubble radius), or a flat universe in which something other than CDM makes up the balance, equivalent to Ω of 0.7 (In this case it would be a coincidence that two quite different invisible substances make comparable contributions).

But it is now clear that only the second option remains in the running: there is compelling evidence that the universe is flat. This evidence comes from the slight temperature-differences over the sky in the background radiation, due to density irregularities which are the precursors of cosmic structure. Theory tell

us that the temperature fluctuations should be biggest on a particular length scale that is related to the distance a sound wave can travel in the early universe. The angular scale corresponding to this length depends, however, on the geometry of the universe. If dark matter and baryons were all, we wouldn't be in a flat universe – the geometry would be hyperbolic. Distant objects would look smaller than in a flat universe. In 2001-02, measurements from balloons and from Antarctica pinned down the angular scale of this ‘doppler peak’: the results indicated ‘flatness’ – a result now confirmed with greater precision by the WMAP satellite.

A value of 0.3 for Ω_{DM} would imply (were there no other energy in the universe) an angle smaller by almost a factor of 2 – definitely in conflict with observations. So what’s the other 70%? It is not dark matter but something that does not cluster – some energy latent in space. The simplest form of this idea goes back to 1917 when Einstein introduced the cosmological constant, or lambda. A positive lambda can be interpreted, in the context of the ordinary Friedman equations, as a fixed positive energy density in all space. This leads to a repulsion because, according to Einstein’s equation, gravity depends on pressure as well as density, and vacuum energy has such a large negative pressure – tension – that the net effect is repulsive.

Einstein’s cosmological constant is just one of the options. A class of more general models is being explored (under names such as ‘quintessence’) where the energy is time-dependent. Any form of dark energy must have negative pressure to be compatible with observations – unclustered relativistic particles, for instance, can be ruled out as candidates. The argument is straightforward: at present, dark energy dominates the universe – it amounts to around 70% of the total mass-energy. But had it been equally dominant in the past, it would have inhibited the growth of the density contrasts in cosmic structures, which occurred gravitational instability. This is because the growth timescale for gravitational instability is $\sim (G\rho_c)^{-\frac{1}{2}}$, where ρ_c is the density of the component that participates in the clustering, whereas the expansion timescale scales as $(G\rho_{total})^{-\frac{1}{2}}$ when curvature is unimportant. If ρ_{total} exceeds ρ_c , the expansion is faster, so the growth is impeded. (Meszaros, 1974)

In the standard model, density contrasts in the dark matter grow by nearly 1000 since recombination. If this growth had been suppressed, the existence of present-day clusters would therefore require irregularities that were already of substantial amplitude at the recombination epoch, contrary to the evidence from CMB fluctuations. For the ‘dark energy’ to be less dominant in the past, its density must depend on the scale factor R more slowly than the R^{-3} dependence of pressure-free matter – i.e. its PdV work must be negative. Cosmologists have introduced a parameter w such that $p = w\rho c^2$. A more detailed treatment yields the requirement that $w < -0.5$. This comes from taking account of baryons and dark matter, and requiring that dark energy should not have inhibited the

growth of structure so much that it destroyed the concordance between the CMB fluctuations (which measure the amplitude at recombination) and the present-day inhomogeneity. Note however that unless its value is -1 (the special case of a classical cosmological constant) w will generally be time-dependent. In principle $w(t)$ can be pinned down by measuring the Hubble expansion rate at different redshifts

This line of argument would in itself have led to a prediction of accelerating cosmic expansion. However, as it turned out, studies of the redshift versus the apparent brightness of distant SNIa – strongly suggestive if not yet completely compelling – had already conditioned us to the belief that galaxies are indeed dispersing at an accelerating rate. As often in science, a clear picture gradually builds up, but the order in which the bits of the jigsaw fall into place is a matter of accident or contingency. CMB fluctuations alone can now pin down Ω_{DM} and the curvature independent of all the other measurements.

The ‘modern’ interest in the cosmological constant stems from its interpretation as a vacuum energy. This leads to the reverse problem: Why is lambda at least 120 powers of 10 smaller than its ‘natural’ value, even though the effective vacuum density must have been very high in order to drive inflation. If lambda is fully resurrected, it will be a posthumous ‘coup’ for de Sitter. His model, dating from the 1920s, not only describes inflation, but would then also describe future aeons of our cosmos with increasing accuracy. Only for the 50-odd decades of logarithmic time between the end of inflation and the present would it need modification!. But of course the dark energy could have a more complicated and time-dependent nature – though it must have negative pressure, and it must not participate in gravitational clustering.

4 SUMMARY AND PROSPECTS

Cosmologists can now proclaim with confidence (but with some surprise too) that, in round numbers, our universe consists of 5% baryons, 25% dark matter, and 70% dark energy. It is indeed embarrassing that 95% of the universe is unaccounted for: even the dark matter is of quite uncertain nature, and the dark energy is a complete mystery.

The network of key arguments is summarised in Figure 1. Historically, the supernova evidence came first. But had the order of events been different, one could have predicted an acceleration on the basis of CDM evidence alone; the supernovae would then have offered gratifying corroboration (despite the unease about possible poorly-understood evolutionary effects).

Our universe is flat, but with a strange mix of ingredients. Why should these all

give comparable contributions (within a modest factor) when they could have differed by a hundred powers of ten?

In the coming decade, we can expect advances on several fronts. Physicists may well develop clearer ideas on what determined the favouritism for matter over antimatter in the early universe, and on the particles that make up the dark matter. Understanding the dark energy, and indeed the big bang itself, is perhaps a remoter goal, but ten years from now theorists may well have replaced the boisterous variety of ideas on the ultra-early universe by a firmer best buy. They will do this by discovering internal inconsistencies in some contending theories, and thereby narrowing down the field. Better still, maybe one theory will earn credibility by explaining things we can observe, so that we can apply it confidently even to things we cannot directly observe. In consequence, we may have a better insight into the origin of the fluctuations, the dark energy, and perhaps the big bang itself.

Inflation models have two generic expectations; that the universe should be flat and that the fluctuations should be gaussian and adiabatic (the latter because baryogenesis would occur at a later stage than inflation). But other features of the fluctuations are in principle measurable and would be a diagnostic of the specific physics. One, the ratio of the tensor and scalar amplitudes of the fluctuations, will have to await the next generation of CMB experiments, able to probe the polarization on small angular scales. Another discriminant among different theories is the extent to which the fluctuations deviate from a Harrison-Zeldovich scale-independent format ($n = 1$ in the usual notation); they could follow a different power law (i.e. be tilted), or have a ‘rollover’ so that the spectral slope is itself a function of scale. Such effects are already being constrained by WMAP data, in combination with evidence on smaller scales from present-day clustering, from the statistics of the Lyman alpha absorption-line ‘forest’ in quasar spectra, and from indirect evidence on when the first minihalos collapsed, signalling the formation of the first Population III stars that ended the cosmic dark age.

In parallel, there will be progress in ‘environmental cosmology’. The new generation of 10-metre class ground based telescopes will give more data on the universe at earlier cosmic epochs, as well as better information on gravitational lensing by dark matter. And there will be progress by theorists too. The behaviour of the dark matter, if influenced solely by gravity, can already be simulated with sufficient accuracy. Gas dynamics, including shocks and radiative cooling, can be included too (though of course the resolution isn’t adequate to model turbulence, nor the viscosity in shear layers). Spectacular recent simulations have been able to follow the formation of the first stars. But the later stages of galactic evolution, where feedback is important, cannot be modelled without parametrising such processes in a fashion guided by physical intuition and observations. Fortunately, we can expect rapid improvements, from observations in all wavebands, in our knowledge of galaxies, and the high-redshift

universe.

Via a combination of improved observations, and ever more refined simulations, we can hope to elucidate how our elaborately structured cosmos emerged from a near-homogeneous early universe.

References

- Blumenthal, G, Faber, S, Primack, J.R, and Rees, M.J. 1984 *Nature* 311, 517
- Davis, M, Efstathiou, G.P, Frenk, C.S. and White, S.D.M., 1985 *Astrophys. J.* 292, 371
- Einasto, J , Kaasik, A and Saar, E, 1974 *Nature* 250, 309
- Kahn, F and Woltjer, L, 1959 *Astrophys. J.* 130, 705
- Meszaros, P. 1974 *Astr. Astrophys.* 37, 225
- Ostriker, J. Peebles, P.J.E., and Yahil, A, 1974 *Astrophys.J (Lett)* 193, L1
- Pagels, H. and Primack, J.R. 1982 *Phys. Rev. Lett.* 48, 223
- Peebles, P.J.E. 1982 *Astrophys.J. (Lett)* 263, L1
- Roberts, M.S. and Rots, A.H. 1973 *Astr.Astrophys* 26, 483.
- Rubin, V.C., Thonnard, N., and Ford, W.K., 1978 *Astrophys J. (Lett)* 225 , L107
- White, S.D.M., Navarro, J.F., Evrard, A.E. and Frenk, C.S. 1993, *Nature* 366, 429
- Zwicky, F, 1937 *Astrophys. J* 86, 217

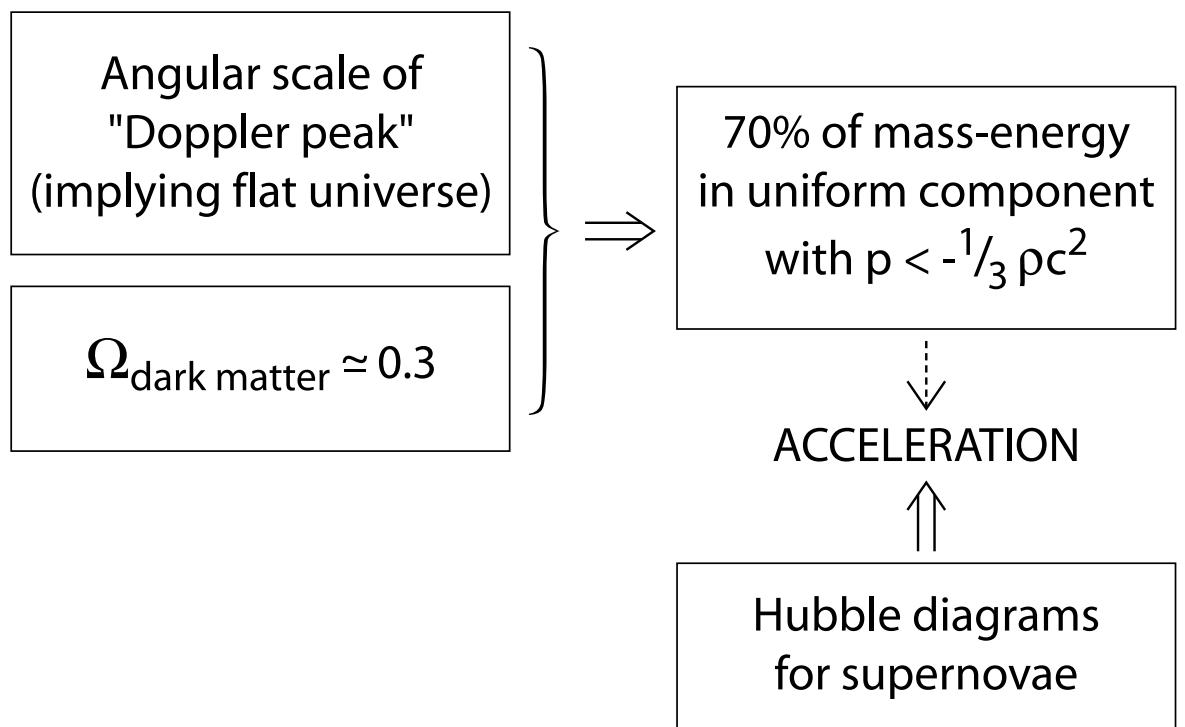


Figure 1. The network of arguments that point towards a flat Universe dominated by ‘dark energy’

Observational Cosmology

R.H. Sanders

Kapteyn Astronomical Institute, Groningen, The Netherlands

Abstract. I discuss the classical cosmological tests—angular size-redshift, flux-redshift, and galaxy number counts—in the light of the cosmology prescribed by the interpretation of the CMB anisotropies. The discussion is somewhat of a primer for physicists, with emphasis upon the possible systematic uncertainties in the observations and their interpretation. Given the curious composition of the Universe inherent in the emerging cosmological model, I stress the value of searching for inconsistencies rather than concordance, and suggest that the prevailing mood of triumphalism in cosmology is premature.

1 Introduction

The traditional cosmological tests appear to have been overshadowed by observations of the anisotropies in the cosmic microwave background (CMB). We are told that these observations accurately measure the geometry of the Universe, its composition, its present expansion rate, and the nature and form of the primordial fluctuations [1]. The resulting values for these basic parameters are very similar to those deduced earlier from a variety of observations—the so-called “concordance model”—with about 30% of the closure density of the Universe comprised of matter (mostly a pressureless, non-baryonic dark matter), the remainder being in negative pressure dark energy [2]. Given the certainty and precision of these assertions, any current discussion of observational cosmology must begin with the question: Is there any room for doubt? Why should we bother with lower precision cosmological tests when we know all of the answers anyway?

While the interpretation of the CMB anisotropies has emerged as the single most important cosmological tool, we must bear in mind that the conclusions drawn do rest upon a number of assumptions, and the results are not altogether as robust as we are, at times, led to believe. One such assumption, for example, is that of adiabatic initial fluctuations—that is, 100% adiabatic. A small admixture of correlated isocurvature fluctuations, an aspect of braneworld scenarios [3], can affect peak amplitudes and thus, the

derived cosmological parameters. A more fundamental assumption is that of the validity of traditional Friedmann-Robertson-Walker (FRW) cosmology in the post-decoupling universe. Is the expansion of the universe described by the Friedmann equation? Even minimal changes to the right-hand-side, such as the equation of state of the dark energy component, can alter the angular size distance to the last scattering surface at $z=1000$ and the luminosity distance to distant supernovae. But even more drastic changes to the Friedmann equation, resulting from modified gravitational physics, have been proposed in attempts to remove the unattractive dark energy [4, 5].

Such suggestions reflect a general unease with the concordance model— a model that presents us with a universe that is strange in its composition. The most abundant form of matter consists of, as yet, undetected non-baryonic particles originally postulated to solve the problems of structure formation and of the missing mass in bound gravitational systems such as galaxies and clusters of galaxies. In this second respect, it is fair to say that it has failed— or, to be generous, not yet succeeded— because the predicted density distribution of dark halos which emerge from cosmic N-body [6] simulations appears to be inconsistent with observations of spiral galaxies [7] or with strong lensing in clusters of galaxies [8].

Even more mysterious is the “dark energy”, the pervasive homogeneous fluid with a negative pressure which may be identified with the cosmological constant, the zero-point energy density of the vacuum. The problem of this unnaturally low energy density, 10^{-122} in Planck units, is well-known, as is the cosmic coincidence problem: why are we observing the Universe at a time when the cosmological constant has, fairly recently, become dynamically important [9]? To put it another way, why are the energy densities of matter and dark energy so comparable at the present epoch? This is strange because the density of matter dilutes with the expanding volume of the Universe while the vacuum energy density does not. It is this problem which has led to the proposal of dynamic dark energy, quintessence— a dark energy, possibly associated with a light scalar field— with an energy density that evolves with cosmic time possibly tracking the matter energy density [10]. Here the difficulty is that the field would generally be expected to have additional observational consequences— such as violations of the equivalence principle at some level, possibly detectable in fifth force experiments [9].

For these reasons, it is even more important to pursue cosmological tests that are independent of the CMB, because one might expect new physics to appear as observations inconsistent with the concordance model. In this sense, discord is more interesting than concord; to take a Hegelian point of view— ideas progress through dialectic, not through concordance. It is with this in mind that I will review observational cosmology with emphasis upon CMB-independent tests.

Below I argue that the evolution of the early, pre-recombination universe is well-understood and tightly constrained by considerations of primordial nucleosynthesis. If one wishes to modify general relativity to give deviations

from Friedmann expansion, then such modifications are strongly constrained at early times, at energies on the order of 1 MeV. However, cosmological evolution is much less constrained in the post-recombination universe where there is room for deviation from standard Friedmann cosmology and where the more classical tests are relevant. I will discuss three of these classical tests: the angular size distance test where I am obliged to refer to its powerful modern application with respect to the CMB anisotropies; the luminosity distance test and its application to observations of distant supernovae; and the incremental volume test as revealed by faint galaxy number counts.

These classical tests yield results that are consistent, to lower precision, with the parameters deduced from the CMB. While one can make minimal changes to standard cosmology, to the equation of state of the dark energy for example, which yield different cosmological parameters, there is no compelling observational reason to do so. It remains the peculiar composition and the extraordinary coincidences embodied by the concordance model that call for deeper insight. Such motivations for questioning a paradigm are not unprecedented; similar worries led to the inflationary scenario which, unquestionably, has had the dominant impact on cosmological thought in the past 25 years and which has found phenomenological support in the recent CMB observations.

I am not going to discuss cosmological tests based upon specific models for structure formation, such as the form of the luminous matter power spectrum [11] or the amplitude of the present mass fluctuations [12]. I do not mean to imply that such such tests are unimportant, it is only that I restrict myself here to more global and model-independent tests. If one is considering a possibility as drastic as a modification of Friedmann expansion due, possibly, to new gravitational physics, then it is tests of the global curvature and expansion history of the Universe that are primary.

I am also going to refrain, in so far as possible, from discussion of theory—of new gravitational physics or of any other sort. The theoretical issues presented by dark matter that can only be detected gravitationally or by an absurdly small but non-zero cosmological constant are essentially not problems for the interpretive astronomer. The primary task is to realistically access the reliability of conclusions drawn from the observations, and that is what I intend to do.

2 Astronomy made simple (for physicists)

I think that it is fair to assume that most of you are physicists, so I begin by defining some of the units and terminology used by astronomers. I do this because much of this terminology is arcane for those not in the field.

First of all there is the peculiar logarithmic scale of flux—magnitudes—whereby a factor of 100 in flux is divided into five equal logarithmic intervals. The system is ancient and has its origin in the logarithmic response of the

human eye. The ratio of the flux of two objects is then given by a difference in magnitudes; i.e.,

$$m_2 - m_1 = -2.5 \log(F_2/F_1) \quad (2.1)$$

where, one will notice, smaller magnitude means larger flux. The zero-point of this logarithmic scale is set by some standard star such as Vega. Because this is related to the flux, and not the luminosity of an object, it is called the “apparent” magnitude. Distant galaxies have apparent magnitudes, in visible light, of greater than 20, and the galaxies in the Hubble Deep Field, go down to magnitudes of 30. The magnitude is typically measured over a specified wavelength range or color band, such as blue (B), visual (V), or infrared (K), and these are designated m_B , m_V , and m_K , or sometimes just B, V, and K. This is made more confusing by the fact that there are several competing photometric systems (or sets of filters) and conversion between them is not always simple.

With a particular photometric system one can measure the color of an astronomical object, expressed as difference in magnitudes in two bands, or color index; e.g.,

$$B - V = 2.5 \log(F_V/F_B) \quad (2.2)$$

Here a larger B-V color index means that an object is relatively redder; a smaller B-V that the object is bluer. Unlike the apparent magnitude, this is an intrinsic property of the object. Or rather, it is intrinsic once the astronomer corrects the magnitudes in the various bands to the zero-redshift ($z = 0$) frame. This is called the “K-correction” and requires a knowledge of the intrinsic spectral energy distribution (SED) of the source, be it a galaxy or a distant supernova.

The luminosity of an object is also an intrinsic property and is usually expressed by astronomers as an “absolute” magnitude. This is the apparent magnitude an object would have if it were placed at a standard distance, taken to be 10 parsecs, i.e. 3×10^{17} m (more on parsecs below). Because this distance is small by extragalactic standards the absolute magnitudes of galaxies turn out to be rather large negative numbers: $M_G \approx -18$ to -21 . The luminosity of a galaxy L_G in units of the solar luminosity L_\odot can be determined from the relation

$$M_G - M_\odot = -2.5 \log(L_G/L_\odot) \quad (2.3)$$

where the absolute magnitude of the sun (in the V band) is 5.5. The luminosities of galaxies typically range from 10^8 to $10^{11} L_\odot$. The peak absolute magnitude of a type I supernova (SNIa) is about -19.5, or comparable to an entire galaxy. This is one reason why these objects are such ideal extragalactic distance probes.

The unit of distance used by astronomers is also archaic: the parsec which is about 3×10^{16} m or about 3 light years. This is the distance to a star with an semi-annual parallax of 1 arc second and is not a bad unit when one

is discussing the very local region of the galaxy. Our galaxy has a diameter between 10 to 20 kiloparsecs, so the kiloparsec is an appropriate unit when discussing galactic structure. The appropriate unit of extragalactic distance, however, is the “megaparsec” or Mpc, with nearby galaxies being those at distances less than 10 Mpc. The nearest large cluster of galaxies, the Virgo cluster, is at a distance of 20 Mpc, and very distant galaxies are those further than 100 Mpc, although here one has to be careful about how distance is operationally defined.

We all know that the Universe is uniformly expanding and the Hubble parameter, H , is the recession velocity of galaxies per unit distance, with H_o being its value in the present Universe. It is typically measured in units of $\text{km s}^{-1}\text{Mpc}^{-1}$ or inverse time. A number of observations point to $H_o \approx 70 \text{ km s}^{-1}\text{Mpc}^{-1}$. The Hubble time is defined as $t_H = H_o^{-1}$ which is about $9.8 \times 10^9 h^{-1}$ years, and this must be comparable to the age of the Universe. The definition $h = H_o/100 \text{ km s}^{-1}\text{Mpc}^{-1}$ is a relic of the recent past when the Hubble parameter was less precisely determined, but I keep using it below because it remains convenient as a unit-less quantity. We can also define a characteristic scale for the universe which is the Hubble radius or $r_H = c/H_o$ and this is $3000 h^{-1}$ Mpc. This would be comparable to the “distance” to the horizon.

Just for interest, one could also define a Hubble acceleration or $a_H = cH_o \approx 7 \times 10^{-10} \text{ m/s}^2$. This modest acceleration of 7 angstroms/second squared is, in effect, the acceleration of the Hubble flow at the horizon if we live in a Universe dominated by a cosmological constant as observations seem to suggest. It is also comparable to the acceleration in the outer parts of galaxies where the need for dark matter first becomes apparent [13]. In some sense, it is remarkable that such a small acceleration has led to a major paradigm shift.

3 Basics of FRW cosmology

The fundamental assumption underlying the construction of cosmological models is that of the cosmological principle: The Universe appears spatially isotropic in all its properties to all observers. The only metric which is consistent with this principle is the Robertson-Walker metric:

$$ds^2 = c^2 dt^2 - \frac{a^2(t) dr^2}{[1 - r^2/R_o^2]} - a^2(t) r^2 (d\theta^2 + \sin^2(\theta) d\phi^2) \quad (3.1)$$

where r is the radial comoving coordinate, $a(t)$ is the dimensionless scale factor by which all distances vary as a function of cosmic time, and R_o^{-2} is a parameter with dimensions of inverse length squared that describes the curvature of the Universe and may be positive, zero, or negative (see [14] for a general discussion).

This is the geometry of the Universe, but dynamics is provided by General Relativity— the Einstein field equations— which yield ordinary differential equations for $a(t)$. The time-time component leads to a second order equation:

$$\ddot{a} = -\frac{4\pi G}{3}a(\rho + 3p/c^2) \quad (3.2)$$

where ρ is the density, p is the pressure and the quantity in parenthesis is the active gravitational mass density. Considering conservation of energy for a perfect fluid

$$d(\rho V) = -pdV/c^2 \quad (3.3)$$

with an equation of state

$$p = w\rho c^2 \quad (3.4)$$

we have $\rho \propto a^{-1(1+w)}$. The equation of state combined with eq. 3.2 tells us that the Universe is accelerating if $w < -1/3$.

The space-space components combined with the time-time component yield the usual first-order Friedmann equation

$$\left(\frac{H}{H_o}\right)^2 - \frac{\Omega_k}{a^2} = \sum_i \Omega_i a^{-3(1+w_i)} \quad (3.5)$$

where $H = \dot{a}/a$ is the running Hubble parameter, the summation is over the various fluids comprising the Universe and

$$\Omega_i = \frac{8\pi G\rho_i}{3H_o^2} \quad (3.6)$$

with $\Omega_k = -(r_H/R_o)^2$. We often see eq. 3.5 written in terms of redshift where $a = (1+z)^{-1}$. Each component has its own equation of state parameter, w_i : $w = 0$ for non-relativistic matter (baryons, CDM); $w = 1/3$ for radiation or other relativistic fluid; $w = -1$ for a cosmological constant; and $-1 < w < -1/3$ for “quintessence”, dynamic dark energy resulting in ultimate acceleration of the universal expansion. I will not consider $w < -1$ which has been termed “phantom” dark energy [15]; here the effective density increases as the Universe expands (this could be realized by a ghost field, a scalar with a kinetic term in the Lagrangian having the wrong sign so it rolls up rather than down a potential hill).

Given a universe composed of radiation, non-relativistic matter, and quintessence, the Friedmann equation takes its familiar form:

$$\left(\frac{H}{H_o}\right)^2 - \frac{\Omega_k}{a^2} = \Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_Q a^{-3(1+w)}. \quad (3.7)$$

Here it is evident that radiation drives the expansion at early times ($a \ll 1$), non-relativistic matter at later times, a non-vanishing curvature ($\Omega_k \neq 0$) at later times still, and, if $w < -1/3$, the vacuum energy density ultimately

dominates. For the purpose of this lecture, I refer to eq. 3.7 with $w = -1$ (the usual cosmological constant) as standard FRW cosmology, while $0 > w \neq -1$ would represent a minimal modification to FRW cosmology. Moreover, when $w = -1$, I replace Ω_Q by Ω_A . I will not consider changes to the Friedmann equation which might result from modified gravitational physics.

Because the subject here is observational cosmology we must discuss the operational definitions of distance in an FRW Universe. If there exists a standard meter stick, an object with a known fixed linear size d which does not evolve with cosmic time, then one could obviously define an angular size distance:

$$D_A = \frac{d}{\theta} \quad (3.8)$$

where θ would be the observed angle subtended by this object. If there exists a standard candle, an object with a known fixed luminosity L which does not vary with cosmic time, then one could also define a luminosity distance:

$$D_L = \left(\frac{L}{4\pi F} \right)^{\frac{1}{2}} \quad (3.9)$$

where F is the measured flux of radiation.

For a RW universe both the angular size distance and the luminosity distance are related to the radial comoving coordinate,

$$r = |R_o| \chi \left[\frac{r_H}{|R_o|} \int_{\tau_o}^{\tau} \frac{d\tau}{a(\tau)} \right] \quad (3.10)$$

where $\tau = tH_o$, $R_o^2 = -r_H^2/\Omega_k$, and

$$\begin{aligned} \chi(x) &= \sin(x) && \text{if } \Omega_k < 0 \\ \chi(x) &= \sinh(x) && \text{if } \Omega_k > 0 \\ \chi(x) &= x && \text{if } \Omega_k = 0. \end{aligned}$$

Then it is the case that

$$D_A = r a(\tau) = r/(1+z) \quad (3.11a)$$

and

$$D_L = r/a(\tau) = r(1+z). \quad (3.11b)$$

It is evident that both the angular size distance and the luminosity distance depend upon the expansion history (through $\int d\tau/a(\tau)$) and the curvature (through $\chi(x)$).

The same is true of a comoving volume element:

$$dV = r^2 dr d\Omega \quad (3.12)$$

where here $d\Omega$ is an incremental solid angle. Therefore, if there exists a class of objects with a non-evolving comoving density, then this leads to another

possible cosmological test: simply count those objects as a function of redshift or flux.

Below, I am going to consider these measures of distance and volume in the form of three classical cosmological tests:

1. Angular size tests which essentially involve the determination of $D_A(z)$. Here one measure θ for objects with a known and (hopefully) standard linear size (such as compact radio sources).
2. Luminosity distance tests which involve the measurement of $F(z)$ for presumably standard candles (such as supernova type Ia, SNIa).
3. dV/dz test which involve the counts of very faint galaxies as a function of flux and redshift.

But before I come to these classic tests, I want to discuss the evidence supporting the validity of the standard hot Big Bang, as an appropriate description of the early pre-recombination Universe.

4 Observational support for the standard model of the early Universe

The discovery 40 years ago of the cosmic microwave background radiation (CMB) ended, for most people, the old debate about Steady-State vs. the Hot Big Bang. Ten years ago, support for the Hot Big Bang was fortified by the COBE satellite which demonstrated that the CMB has a Planck spectrum to extremely high precision; it is, quite literally, the most perfect black body observed in nature [16]. This makes any model in which the CMB is produced by some secondary process, such as thermal re-radiation of starlight by hot dust, seem extremely difficult, if not impossible, to contrive.

Not only does the background radiation have a thermal spectrum, it is now evident that this radiation was hotter in the past than now as expected for adiabatic expansion of the Universe. This is verified by observations of neutral carbon fine structure lines as well as molecular hydrogen rotational transitions in absorption line systems in the spectra of distant quasars. Here, the implied population of different levels, determined primarily by the background radiation field, is an effective thermometer for that radiation field. One example is provided by a quasar with an absorption line system at $z = 3.025$ which demonstrates that the temperature of the CMB at this redshift was $12.1_{-8.2}^{+1.7}$ K, consistent with expectations ($T \propto 1 + z$) [17].

However, the most outstanding success story for the Hot Big Bang is generally considered to be that of Big Bang Nucleosynthesis (BBN) which, for a given number of relativistic particle species, predicts the primordial abundances of the light isotopes with, effectively, one free parameter: the ratio of baryons-to-photons, η [18]. I want to review this success story, and point out that there remains one evident inconsistency which may be entirely observational, but which alternatively may point to new physics.

We saw above in the Friedmann equation (eq. 3.7) that radiation, if present, will always dominate the expansion of the Universe at early enough epochs (roughly at $z \approx 2 \times 10^4 \Omega_m$.) This makes the expansion and thermal history of the Universe particularly simple during this period. The Friedmann equation becomes

$$H^2 = \frac{4\pi G a T^4 N(T)}{3c^2}; \quad (4.1)$$

here a is the radiation constant and $N(T)$ is the number of degrees of freedom in relativistic particles. The scale factor is seen to grow as $t^{1/2}$ which means that the age of the Universe is given by $t = 1/2H$. This implies, from eq. 4.1, an age-temperature relation of the form $t \propto T^{-2}$. Putting in numbers, the precise relation is

$$t = \frac{2.5}{T_{MeV}^2 N(T)^{\frac{1}{2}}} \text{ s} \quad (4.2)$$

where the age is given in seconds and T_{MeV} is the temperature measured in MeV. It is only necessary to count the number of relativistic particle species:

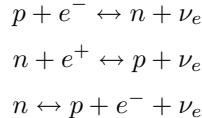
$$N(T) = \sum g_B + \frac{7}{8} \sum g_F \quad (4.3)$$

where the sums are over the number of bosonic degrees of freedom (g_B) and fermionic degrees of freedom (g_F). The factor $7/8$ is due to the difference in Bose-Einstein and Fermi-Dirac statistics. Adding in all the known species—photons, electrons-positrons (when $T_{MeV} > 0.5$), three types of neutrinos and anti-neutrinos— we find

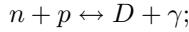
$$t \approx T_{MeV}^{-2} \text{ s} \quad (4.4)$$

for the age-temperature relation in the early Universe.

When the Universe is less than one second old ($T > 1$ MeV) the weak interactions



are rapid enough to establish equilibrium between these various species. But when T falls below 1 MeV, the reaction rates become slower than the expansion rate of the Universe, and neutrons “freeze out”— they fall out of thermal equilibrium, as do the neutrinos. This means the equilibrium ratio of neutrons to protons at $T \approx 1$ MeV is frozen into the expanding soup: $n/p \approx 0.20 - 0.25$. You all know that neutrons outside of an atomic nucleus are unstable particles and decay with a half-life of about 15 minutes. But before that happens there is a possible escape route:



that is to say, a neutron can combine with a proton to make a deuterium nucleus and a photon. However, so long as the mean energy of particles and

photons is greater than the binding energy of deuterium, about 86 Kev, the inverse reaction happens as well; as soon as a deuterium nucleus is formed it is photo-dissociated. This means that it is impossible to build up a significant abundance of deuterium until the temperature of the Universe has fallen below 86 KeV or, looking back at eq. 4.4, until the Universe has become older than about 2.5 minutes. Then all of the remaining neutrons are rapidly processed into deuterium. But the deuterium doesn't stay around for long either.

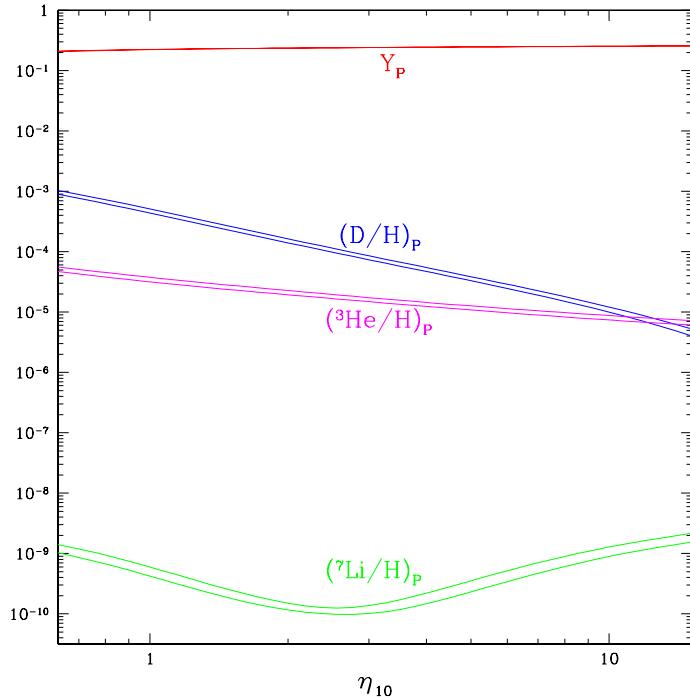


Fig. 1. The predicted abundances of the light isotopes as a function of η [18]. Here Y_p is the predicted mass fraction of helium and is based upon the assumption of three neutrino types. The widths of the bands show the theoretical uncertainty.

Given the temperature and particle densities prevailing at this epoch, there are a series of two-body reactions by which two deuterons combine to make He^4 and trace amounts of lithium and He^3 . These reactions occur at a rate which depends upon the overall abundance of baryons, the ratio of baryons to photons:

$$\eta = n_b/n_\gamma = 274 \Omega_b h^2 \times 10^{-10} \quad (4.5)$$

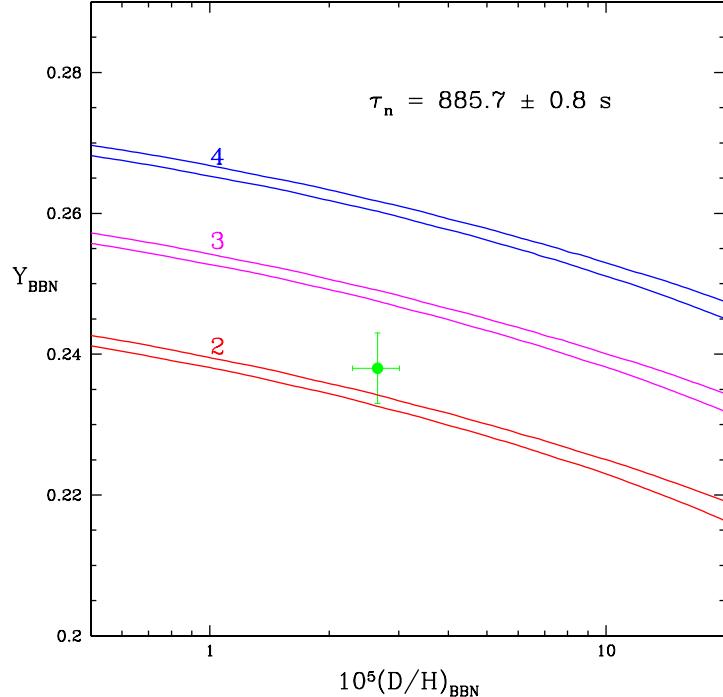


Fig. 2. The predicted abundance, Y_p , of helium (the mass fraction) as a function of the predicted deuterium abundance for two, three, and four neutrino types [18]. The point with error bars is the observed abundances of helium and deuterium.

So essentially all neutrons which survive until $T = 86$ KeV become locked up in He^4 . Therefore, the primordial abundance of helium depends primarily upon the expansion rate of the Universe: the faster the expansion (due, say, to more neutrino types or to a larger constant of gravity) the more helium. The abundance of remaining deuterium, however, depends upon the abundance of baryons, η : the higher η the less deuterium. This is why it is sometimes said [18] that the abundance of primordial helium is a good chronometer (it measures the expansion rate), while the abundance of deuterium is a good baryometer (it measures Ω_b). This is evident in Figs. 1 and 2 where we see first the predicted abundances of various light isotopes as a function of η , and secondly, the predicted abundance of He vs. that of deuterium for two, three and four neutrino types.

The determination of primordial abundances is not a straightforward matter because the abundance of these elements evolves due to processes occurring within stars (“astration”). In general, the abundance of helium increases

(hydrogen is processed to helium providing the primary energy source for stars), while deuterium is destroyed by the same process. This means that astronomers, when trying to estimate primordial abundances of deuterium or helium, must try to find pristine, unprocessed material, in so far as possible. One way to find unprocessed material is to look back at early times, or large redshift, before the baryonic material has been recycled through generations of stars. This can be done with quasar absorption line systems, where several groups of observers have been attempting to identify very shallow absorption lines of deuterium at the same redshift as the much stronger hydrogen Lyman alpha absorption line systems [19, 20, 21, 22]. It is a difficult observation requiring the largest telescopes; the lines identified with deuterium might be misidentified weak hydrogen or metal lines (incidentally, for an astronomer, any element heavier than helium is a metal). Taking the results of various groups at face value, the weighted mean value [18] is $D/H \approx 2.6 \pm 0.3 \times 10^{-5}$. Looking back at Fig. 1, we see that this would correspond to $\eta = 6.1 \pm 0.6 \times 10^{-10}$ or $\Omega_b h^2 = 0.022 \pm 0.003$.

A word of caution is necessary here: the values for the deuterium abundance determined by the different groups scatter by more than a factor of two, which is considerably larger than the quoted statistical errors ($\approx 25\%$). This indicates that significant systematic effects are present. But it is noteworthy that the angular power spectrum of the CMB anisotropies also yields an estimate of the baryon abundance; this is encoded in the ratio of the amplitudes of the second to first peak. The value is $\Omega_b h^2 = 0.024 \pm 0.001$. In other words, the two determinations agree to within their errors. This is quite remarkable considering that the first determination involves nuclear processes occurring within the first three minutes of the Big Bang, and the second involves oscillations of a photon-baryon plasma on an enormous scale when the Universe is about 500,000 years old. If this is a coincidence, it is truly an astounding one.

So much for the baryometer, but what about the chronometer—helium? Again astronomers are obliged to look for unprocessed material in order to estimate the primordial abundance. The technique of looking at quasar absorption line systems doesn't work for helium because the absorption lines from the ground state are far in the ultraviolet—about 600 Å for neutral helium and, more likely, 300 Å from singly ionized helium. This is well beyond the Lyman limit of hydrogen, where the radiation from the background quasar is effectively absorbed [23]. Here the technique is to look for He emission lines from HII regions (ionized gas around hot stars) in nearby galaxies and compare to the hydrogen emission lines. But how does one know that the gas is unprocessed? The clue is in the fact that stars not only process hydrogen into helium, but they also, in the late stages of their evolution, synthesize heavier elements (metals) in their interiors. Therefore the abundance of heavier elements, like silicon, is an indicator of how much nuclear processing the ionized gas has undergone. It is observed that the He abundance is correlated with the metal abundance; so the goal is to find HII regions with as low a metal

abundance as possible, and then extrapolate this empirical correlation to zero metal abundance [24, 25]. The answer turns out to be $\text{He}/\text{H} \approx 0.24$, which is shown by the point with error bars in Fig. 2.

This value is embarrassingly low, given the observed deuterium abundance. It is obviously more consistent with an expansion rate provided by only two neutrino types rather than three, but we know that there are certainly three types. Possible reasons for this apparent anomaly are:

- 1) Bad astronomy: There are unresolved systematic errors in determination of the relative He abundance in HII regions indicated by the fact that the results of different groups differ by more than the quoted statistical errors [18]. The derivation of the helium to hydrogen ratio from the observed He^+/H^+ ratio requires some understanding of the structure of the HII regions. If there are relatively cool ionizing stars ($T < 35000$ K) spatially separated from the hotter stars, there may be relatively less He^+ associated with a given abundance of H^+ . Lines of other elements need to be observed to estimate the excitation temperature; it is a complex problem.
- 2) New neutrino physics: There may be an asymmetry between neutrinos and anti-neutrinos (something like the baryon- antibaryon asymmetry which provides us with the observed Universe). This would manifest itself as a chemical potential in the Boltzmann equation giving different equilibrium ratios of the various neutrino species [26].
- 3) New gravitational physics: any change in the gravitational interaction which is effective at early epochs (braneworld effects?) could have a pronounced effect on nucleosynthesis. For example, a lower effective constant of gravity would yield a lower expansion rate and a lower He abundance. The standard minimal braneworld correction term, proportional to the square of the density [27], goes in the wrong direction.

It is unclear if the low helium abundance is a serious problem for the standard Big Bang. But it is clear that the agreement of the implied baryon abundance with the CMB determination is an impressive success, and strongly supports the assertion that the Hot Big Bang is the correct model for the pre-recombination Universe.

5 The post-recombination Universe: determination of H_o and t_o

Certainly the most basic of the cosmological parameters is the present expansion rate, H_o , because this sets the scale of the Universe. Until a few years ago, there was a factor of two uncertainty in H_o ; with two separate groups claiming two distinct values, one near $50 \text{ km s}^{-1}\text{Mpc}^{-1}$ and the other nearer $100 \text{ km s}^{-1}\text{Mpc}^{-1}$, and the errors quoted by both groups were much smaller than this factor of two difference. This points out a problem which is common in observational cosmology (or indeed, astronomy in general). Often the indicated statistical errors give the impression of great precision, whereas the

true uncertainty is dominated by poorly understood or unknown systematic effects. That was true in the Hubble constant controversy, and there is no less reason to think that this problem is absent in modern results. I will return to this point several times below.

The great leap forward in determination of H_0 came with the Hubble Space Telescope (HST) program on the distance scale. Here a particular kind of variable stars—Cepheid variables—were observed in twenty nearby spiral galaxies. Cepheids exhibit periodic variations in luminosity by a factor of two on timescales of 2–40 days. There is a well-determined empirical correlation between the period of Cepheids and their mean luminosity—the longer the period the higher the luminosity. Of course, this period-luminosity relation must be calibrated by observing Cepheids in some object with a distance known by other techniques and this remains a source of systematic uncertainty. But putting this problem aside, the Hubble Space telescope measured the periods and the apparent magnitudes, without confusion from adjacent bright stars, of a number of Cepheids in each of these relatively nearby galaxies, which yielded a distance determination (eq. 3.9). These galaxies are generally too close (less than 15 Mpc) to sample the pure Hubble flow—the Hubble flow on these scales is contaminated by random motion of the galaxies and systematic cosmic flows—but these determinations do permit a calibration of other secondary distance indicators which reach further out, such as supernovae type Ia (SNIa) and the Tully-Fisher relation (the observed tight correlation between the rotation velocities of a spiral galaxies and their luminosities). After an enormous amount of work by a number of very competent astronomers [28], the answer turned out to be $h = 0.72 \pm .10$

As I mentioned there is the known systematic uncertainty of calibrating the period-luminosity relation, but there are other possible systematic effects that are less well-understood: How can we be certain that the period-luminosity relation for Cepheids is the same in all galaxies? For example, is this relation affected by the concentration of elements heavier than helium (the metallicity)? In view of such potential problems, other more direct physical methods, which by-pass the traditional “distance ladder” are of interest. Chief among these is the Sunyaev-Zeldovich (S-Z) effect which is relevant to clusters of galaxies [29]. The baryonic mass of clusters of galaxies is primarily in the form of hot gas, which typically exceeds the mass in the visible galaxies by more than a factor of two. This gas has a temperature between 10^7 and 10^8 K (i.e., the sound speed is comparable to the one-dimensional velocity dispersion of the galaxies) and is detected by satellite X-ray telescopes with detectors in the range of several KeV. The S-Z effect is a small change in the intensity of the CMB in the direction of such clusters due to Compton scattering of CMB photons by thermal electrons (classical electron scattering would, of course, produce no intensity change). Basically, CMB photons are moved from the Rayleigh-Jeans part of the black body spectrum to the Wien part, so the effect is observable as a spectral distortion of the black body spectrum in the

range of 100 to 300 GHz. It is a small effect (on the order of 0.4 milli Kelvin) but still 5 to 10 times larger than the intrinsic anisotropies in the CMB.

By measuring the amplitude of the S-Z effect one determines an optical depth

$$\tau = \sigma n_e l \quad (5.1)$$

where σ is the frequency dependent cross section, l is the path length, and n_e is the electron density. Because these same clusters emit X-rays via thermal bremsstrahlung, we may also determine, from the observed X-ray intensity, an emission measure:

$$E = n_e^2 l \quad (5.2)$$

Here we have two equations for two unknowns, n_e and l . (This is simplifying the actual calculation because n_e is a function of radius in the cluster.) Knowing l and the angular diameter of the cluster θ we can then calculate the angular size distance to the cluster via eq. 3.8. Hence, the Hubble parameter is given by $H_o = v/D_A$ where v is the observed recession velocity of the cluster. All of this assumes that the clusters have a spherical shape on average, so the method needs to be applied to a number of clusters. Even so biases are possible if clusters have more typically a prolate shape or an oblate shape, or if the X-ray emitting gas is clumpy. Overall, for a number of clusters [30] the answer turns out to be $h = 0.6$ —somewhat smaller than the HST distance ladder method, but the systematic uncertainties remain large.

A second direct method relies on time delays in gravitational lenses [31]. Occasionally, a distant quasar (the source) is lensed by an intervening galaxy (the lens) into multiple images; that is to say, we observe two or more images of the same background object separated typically by one or two seconds of arc. This means that there are two or more distinct null geodesics connecting us to the quasar with two or more different light travel times. Now a number of these quasars are intrinsically variable over time scales of days or months (not periodic but irregular variables). Therefore, in two distinct images we should observe the flux variations track each other with a time delay. This measured delay is proportional to the ratio $D_l D_s / D_{ls}$ where these are the angular size distances to the lens, the source, and the lens to the source. Since this ratio is proportional to H_o^{-1} , the measured time delay, when combined with a mass model for the lens (the main source of uncertainty in the method), provides a determination of the Hubble parameter. This method, applied to several lenses [32, 33], again tends to yield a value of h that is somewhat smaller than the HST value, i.e., ≈ 0.6 . In a recent summary [34] it is claimed that, for four cases where the lens is an isolated galaxy, the result is $h = 0.48 \pm .03$, if the overall mass distribution in each case can be represented by a singular isothermal sphere. On the other hand, in a well-observed lens where the mass distribution is constrained by observations of stellar velocity dispersion [35], the implied value of h is $0.75^{+.07}_{-.06}$. Such supplementary observations are important because the essential uncertainty with this technique is in the adopted mass model of the lens.

It is probably safe to say that $h \approx 0.7$, with an uncertainty of 0.10 and perhaps a slight bias toward lower values, but the story is not over as S-Z and gravitational lens determinations continue to improve. This is of considerable interest because the best fit to the CMB anisotropies observed by WMAP implies that $h = 0.72 \pm .05$ in perfect agreement with the HST result. With the S-Z effect and lenses, there remains the possibility of a contradiction.

With $h = .70$, we find a Hubble time of $t_H = 14$ Gyr. Now in FRW cosmology, the age of the Universe is $t_o = ft_H$ where f is a number depending upon the cosmological model. For an Einstein-de Sitter Universe (i.e., $\Omega_k = 0$, $\Omega_Q = 0$, $\Omega_m = 1$) $f = 2/3$ which means that $t_o = 9.1$ Gyr. For an empty negatively curved Universe, $f = 1$ which means that the age is the Hubble time. Generally, models with a dominant vacuum energy density ($\Omega_Q \approx 1$, $w \approx -1$) are older ($f \geq 1$) and for the concordance model, $f = 0.94$. Therefore, independent determinations of the age of the Universe are an important consistency test of the cosmology.

It is reasonable to expect that the Universe should be older than the oldest stars it contains, so if we can measure the ages of the oldest stars, we have, at least, a lower limit on the age of the Universe. Globular star clusters are old stellar systems in the halo of our own galaxy; these systems are distributed in a roughly spherical region around the galactic disk and have low abundances of heavy elements suggesting they were formed before most of the stars in the disk. If one can measure the luminosity, L_u , of the most luminous un-evolved stars in a globular cluster (that is, stars still burning hydrogen in their cores), then one may estimate the age. That is because this luminosity is correlated with age: a higher L_u means a younger cluster. Up to five years ago, this method yielded globular cluster ages of $t_{gc} \approx 14 \pm 2$ Gyr, which, combined with the Hubble parameter discussed above, would be in direct contradiction with the Einstein-de Sitter $\Omega_m = 1$ Universe. But about ten years ago the Hipparcos satellite began to return accurate parallaxes for thousands of relatively nearby stars which led to a recalibration of the entire distance scale. Distances outside the solar system increased by about 10% (in fact, the entire Universe suddenly grew by this same factor leading to a decrease in the HST value for the Hubble parameter). This meant that the globular clusters were further away, that L_u was 20% larger, and the clusters were correspondingly younger: $t_{gc} \approx 11.5 \pm 1.3$ Gyr. If we assume that the Universe is about 1 Gyr older than the globular clusters, then the age of the Universe becomes 12.5 ± 2 Gyr [36] which is almost consistent with the Einstein-de Sitter Universe. At least there is no longer any compelling time scale argument for a non-zero vacuum energy density, $\Omega_Q > 0$. The value of accurate basic astronomical data (and what is more basic than stellar positions?) should never be underestimated.

A second method for determining the ages of stars is familiar to all physicists, and that is radioactive dating. This has been done recently by observations of a U^{238} line in a metal-poor galactic star (an old star). Although the iron abundance in this star is only 1/800 that of the sun, the abundances

of a group of rare earth metals known as r-process elements are enhanced. The r-process is rapid neutron absorption onto iron nuclei (rapid compared to the timescale for subsequent β decay) which contributes to certain abundance peaks in the periodic table and which occurs in explosive events like supernovae. This means that this old star was formed from gas contaminated by an even older supernova event; i.e. the uranium was deposited at a definite time in the past. Now U^{238} is unstable with a half life of 4.5 Gyr which makes it an ideal probe on cosmological times scales. All we have to do is compare the observed abundance of U^{238} to that of a stable r-process element (in this case osmium), with what is expected from the r-process. The answer for the age of this star (or more accurately, the SN which contaminated the gas out of which the star formed) is 12.5 ± 3 Gyr, which is completely consistent with the globular cluster ages [37].

If we take $0.6 < h < 0.7$, and $9.5 \text{ Gyr} < t_o < 15.5 \text{ Gyr}$ this implies that $0.59 < H_o t_o < 1.1$. This is consistent with a wide range of FRW cosmologies from Einstein-de Sitter to the concordance model. That is to say, independent measurements of H_o and t_o are not yet precise enough to stand as a confirmation or contradiction to the WMAP result.

6 Looking for discordance: the classical tests

6.1 The angular size test

The first of the classical cosmological tests we will consider is the angular size test. Here one measures the angular size of a standard meter stick (hopefully) as a function of redshift; different FRW cosmologies make different predictions, but basically, for all FRW models $\theta(z)$ first decreases as $1/z$ (as would be expected in a Euclidean universe) and then increases with z . This is because the angular size distance is given by $D_A = r/(1+z)$ but the radial comoving coordinate approaches a finite value as $z \rightarrow \infty$. The angular size distance reaches a maximum at a redshift between 1 and 2 and then decreases again.

When giant radio galaxies at large redshift were discovered in the 1960's there was considerable optimism that these could be used as an angular size cosmological probe. Radio galaxies typically have a double-lobe structure with the radio emitting lobes straddling the visible galaxy; these lobes can extend hundreds of kpc beyond the visible object. Such a linear structure may be oriented at any angle to the observer's line-of-sight, so one needs to measure the angular sizes of a number of radio galaxies in a given redshift bin and only consider the largest ones, i.e., those likely to be nearly perpendicular to the line-of-sight.

The result of all this work was disappointing. It appeared that the angular size of radio sources kept decreasing with redshift just as one would expect for a pure Euclidean universe [38]. The obvious problem, that plagues all classical tests, is that of evolution. Very likely, these radio galaxies are not standard

meter sticks at all, but that they were actually smaller at earlier epochs than now. This would be expected, because such objects are thought to result from jets of relativistic particles ejected from the nucleus of the parent galaxy in opposite directions. The jets progress through the surrounding intergalactic medium at a rate determined by the density of that medium, which, of course, was higher at larger redshift.

But there is another class of radio sources that would be less susceptible to such environmental effects: the compact radio sources. These are objects, on a scale of milli-arc-seconds, typically associated with distant quasars, that are observed with radio interferometers having global baselines. The morphology is that of a linear jet with lengths typically less than 30 or 40 pc, so these would presumably be emission from the jets of relativistic particles deep in the galactic nucleus near the central engine producing them. The intergalactic medium, and its cosmological evolution, would be expected to have no effect here [39].

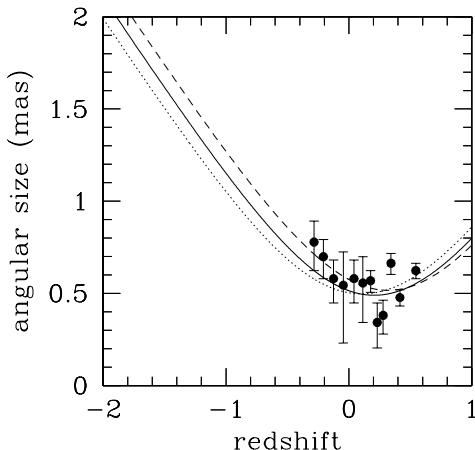


Fig. 3. The median angular size vs. redshift (log-log plots) for 145 compact radio sources in 12 redshift bins. The curves are the three flat cosmological models: dashed, $\Omega_\Lambda = 0.9$; solid, $\Omega_\Lambda = 0.7$ (concordance), dotted, $\Omega_\Lambda = 0.1$. The physical size of the sources (20-40 pc) has been chosen for the best fit

The result of plotting the median angular size of about 150 of these sources as a function of redshift is shown on a log-log plot in Fig. 3 [40]. Also shown are the predicted relations for three flat cosmologies ($\Omega_k = 0$) with $\Omega_m = 0.9$,

0.3, 0.1, the remainder being in a cosmological constant (the middle curve is the concordance model). In each case the linear size of the compact radio sources was chosen to achieve the best fit to the data.

It is evident that the general property of FRW models (that the angular size of a standard meter stick should begin to increase again beyond a redshift of about 1.5) is present in this data. However, no statistical test or maximum likelihood analysis is necessary to see that all three models fit the data equally well. This is basically an imprecise cosmological test and cannot be improved, particularly considering that these objects may also evolve in some unknown way with cosmic time. Looking at the figure, one may notice that measurement of angular sizes for just a few objects at lower redshift might help distinguish between models. However, there are very few such objects at lower redshift, and these have a much lower intrinsic radio power than those near redshift one. It is dangerous to include these objects on such a plot because they are probably of a very different class.

6.2 The modern angular size test: CMB-ology

Although it is not my purpose here to discuss the CMB anisotropies, it is necessary to say a few words on the preferred angular scale of the longest wavelength acoustic oscillations, the “first peak”, because this is now the primary evidence for a flat Universe ($\Omega_k = 0$). In Fig. 4 we see again the now very familiar plot of the angular power spectrum of anisotropies as observed by WMAP [42] (in my opinion, of all the WMAP papers, this reference provides the clearest discussion of the physics behind the peak amplitudes and positions). The solid line is the concordance model— not a fit, but just the predicted angular power spectrum (via CMBFAST [41]) from the $\Omega_m = 0.3$, $\Omega_A = 0.7$ model Universe with an optical depth of $\tau \approx 0.17$ to the surface of last scattering. I must admit that the agreement is impressive.

I remind you that the harmonic index on the horizontal axis is related to angular scale as

$$l \approx \pi/\theta \quad (6.1)$$

so the first peak, at $l \approx 220$, would correspond to an angular scale of about one degree. I also remind you that the first peak corresponds to those density inhomogeneities which entered the horizon sometime before decoupling (at $z = 1000$); enough before so that they have had time to collapse to maximum compression (or expand to maximum rarefaction) just at the moment of hydrogen recombination. Therefore, the linear scale of these inhomogeneities is very nearly given by the sound horizon at decoupling, that is

$$l_h \approx ct_{dec}/\sqrt{3} \quad (6.2)$$

where t_{dec} is the age of the Universe at decoupling.

So one might say, the test is simple: we have a known linear scale l_h which corresponds to an observed angular scale ($\theta \approx 0.014$ rad) so we can determine

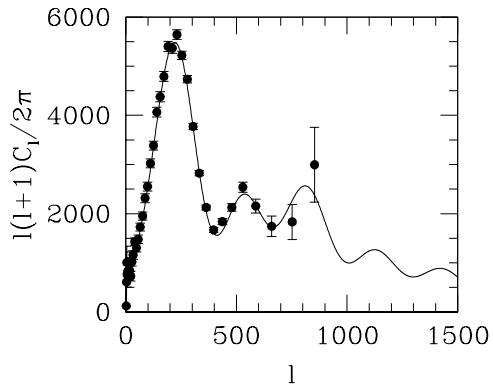


Fig. 4. The angular power spectrum of CMB anisotropies observed by WMAP [42]. The solid line is not a fit but the is the concordance model proposed earlier [2]

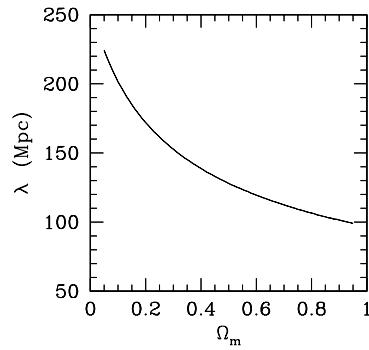


Fig. 5. The comoving linear scale of the perturbation corresponding to the first peak as a function of Ω_m

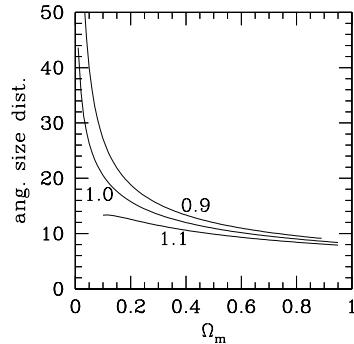


Fig. 6. The angular size distance (Gpc) to the last scattering surface ($z = 1000$) as a function of Ω_m for various values of Ω_{tot}

the geometry of the Universe. It is not quite so simple because the linear scale, l_h depends, via t_{dec} on the matter content of the Universe (Ω_m); basically, the larger Ω_m , the sooner matter dominates the expansion, and the earlier decoupling with a correspondingly smaller l_h . This comoving linear scale is shown in Fig. 5 as a function of Ω_m (Ω_A hardly matters here, because the vacuum energy density which dominates today has no effect at the epoch of decoupling). Another complication is that the angular size distance to the surface of last scattering not only depends upon the geometry, but also upon the expansion history. This is evident in Fig. 6 which shows the comoving angular size distance (in Gpc) to the surface of last scattering as a function of Ω_m for three values of $\Omega_{tot} = \Omega_m + \Omega_A$ (i.e., $\Omega_k = 1 - \Omega_{tot}$). Note that the comoving angular size distance, $D_A(1+z)$, is the same as the radial comoving coordinate r .

We can combine Figs. 5 and 6 to plot the expected angular size (or harmonic index) of the first peak as a function of Ω_m and Ω_{tot} , and this is shown in Fig. 7 with the dashed line giving the observed l of the first peak. We see that a model with $\Omega_{tot} \geq 1.1$ (a closed universe) is clearly ruled out, but it would be possible to have an open model with $\Omega_{tot} = 0.9$ and $\Omega_m = 0.8$ from the position of the first peak alone; the predicted peak amplitude, however, would be about 40% too low. The bottom line of all of this is that the *position* of the first peak does not uniquely define the geometry of the Universe because of a degeneracy with Ω_m (I haven't mentioned the degeneracy with h taken here to be 0.72). To determine whether or not we live in a flat Universe we need an independent handle on Ω_m and that is provided, in WMAP data, by the amplitudes of the first two peaks (the more non-baryonic matter, the deeper the forming potential wells, and the lower the amplitudes). From this

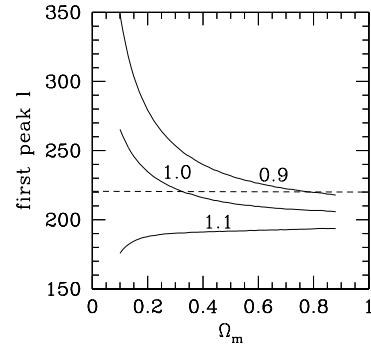


Fig. 7. The harmonic index expected for the first peak as a function of Ω_m for various values of Ω_{tot} .

it is found that $\Omega_m \approx 0.3$, and from Fig. 7 we see that the model Universe should be near flat ($\Omega_{tot} \approx 1.0$). Of course if the Universe is near flat with $\Omega_m = 0.3$ then the rest must be in dark energy; this is the indirect evidence from the CMB anisotropies for dark energy.

I just add here that the observed peak amplitudes (given the optical depth to $z = 1000$ determined from WMAP polarization results [43]), is taken now as definitive evidence for CDM. However, alternative physics which affects the amplitude and positions of peaks (e.g. [3] could weaken this conclusion, as well as affect the derived cosmological parameters. Even taking the peak amplitudes as *prima facie* evidence for the existence of cold dark matter, it is only evidence for CDM at the epoch of recombination ($z = 1000$) and not in the present Universe. To address the cosmic coincidence problem, models have been suggested in which dark matter transmutes into dark energy (e.g. [44]).

Now I turn to the direct evidence for dark energy.

6.3 The flux-redshift test: Supernovae Ia

Type I supernovae are thought to be nuclear explosions of carbon/oxygen white dwarfs in binary systems. The white dwarf (a stellar remnant supported by the degenerate pressure of electrons) accretes matter from an evolving companion and its mass increases toward the Chandrasekhar limit of about $1.4 M_\odot$ (this is the mass above which the degenerate electrons become relativistic and the white dwarf unstable). Near this limit there is a nuclear detonation in the core in which carbon (or oxygen) is converted to iron. A nuclear flame

propagates to the exterior and blows the white dwarf apart (there are alternative models but this is the favored scenario [45]).

These events are seen in both young and old stellar populations; for example, they are observed in the spiral arms of spiral galaxies where there is active star formation at present, as well as in elliptical galaxies where vigorous star formation apparently ceased many Gyr ago. Locally, there appears to be no difference in the properties of SNIa arising in these two different populations, which is important because at large redshift the stellar population is certainly younger.

The peak luminosity of SNIa is about $10^{10} L_\odot$ which is comparable to that of a galaxy. The characteristic decay time is about one month which, in the more distant objects, is seen to be stretched by $1+z$ as expected. The light curve has a characteristic form and the spectra contain no hydrogen lines, so given reasonable photometric and spectroscopic observations, they are easy to identify as SNIa as opposed to type II supernovae; these are thought to be explosions of young massive stars and have a much larger dispersion in peak luminosity [46].

The value of SNIa as cosmological probes arises from the high peak luminosity as well as the observational evidence (locally) that this peak luminosity is the sought-after standard candle. In fact, the absolute magnitude, at peak, varies by about 0.5 magnitudes which corresponds to a 50%-60% variation in luminosity; this, on the face of it, would make them fairly useless as standard candles. However, the peak luminosity appears to be well-correlated with decay time: the larger L_{peak} , the slower the decay. There are various ways of quantifying this effect [46], such as

$$M_B \approx 0.8(\Delta m_{15} - 1.1) - 19.5 \quad (6.3)$$

where M_B is the peak absolute magnitude and Δm_{15} is the observed change in apparent magnitude 15 days after the peak [47]. This is an empirical relationship, and there is no consensus about the theoretical explanation, but, when this correction is applied it appears that $\Delta L_{peak} < 20\%$. If true, this means that SNIa are candles that are standard enough to distinguish between cosmological models at $z \approx 0.5$.

In a given galaxy, supernovae are rare events (on a human time scale, that is), with one or two such explosions per century. But if thousands of galaxies can be surveyed on a regular and frequent basis, then it is possible to observe several events per year over a range of redshift. About 10 years ago two groups began such ambitious programs [48, 49]; the results have been fantastically fruitful and have led to a major paradigm shift.

The most recent results are summarized in [50]: at present, about 230 SNIa have been observed out to $z = 1.2$. The bottom line is that SNIa are 10% to 20% fainter at $z \approx 0.5$ than would be expected in an empty ($\Omega_{tot} = 0$) non-accelerating Universe. But, significantly, at $z \geq 1$ the supernovae appear to become brighter again relative to the non-accelerating case; this should happen in the concordance model at about this redshift because it is here

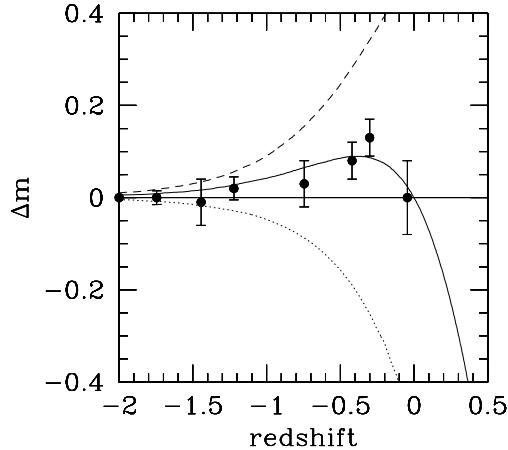


Fig. 8. The Hubble diagram for SNIa normalized to an empty non-accelerating Universe. The points are binned median values for 230 supernovae [50]. The curves show the predictions for three flat ($\Omega_{tot} = 1$) cosmological models: The dashed line is the model dominated by a cosmological constant ($\Omega_\Lambda = 0.9$), the solid curve is the concordance model ($\Omega_\Lambda = 0.7$), and the dotted curve is the matter dominated model ($\Omega_\Lambda = 0.1$).

that the cosmological constant term in the Friedmann equation (eq. 3.7) first begins to dominate over the matter term. This result is shown in Fig. 8 which is a plot of the median Δm , the observed deviation from the non-accelerating case, in various redshift bins as a function of redshift (i.e., the horizontal line at $\Delta m = 0$ corresponds to the empty universe). The solid curves show the prediction for various flat ($\Omega_{tot} = 1$) models with the value of the cosmological term indicated. It is evident that models dominated by a cosmological term or by matter are inconsistent with the observations at extremely high levels of significance, while the concordance model agrees quite well with the observations.

It is also evident from the figure that the significance of the effect is not large, perhaps 3 or 4σ (quite a low level of significance on which to base a paradigm shift). When all the observed supernovae are included on this plot, it is quite a messy looking scatter with a minimum χ^2 per degree of freedom (for flat models) which is greater than one. Moreover the positive result depends entirely upon the empirical peak luminosity-decay rate relationship and, of course, upon the assumption that this relation does not evolve. So, before we

become too enthusiastic we must think about possible systematic effects and how these might affect the conclusions. These effects include:

1) Dust: It might be that supernovae in distant galaxies are more (or less) dimmed by dust than local supernovae. But normal dust, with particle sizes comparable to the wavelength of light, not only dims but also reddens (for the same reason, Rayleigh scattering, that sunsets are red). This is quantified by the so-called color excess. Remember I said that astronomers measure the color of an object by its B-V color index (the logarithm of a flux ratio). The color excess is defined as

$$E(B - V) = (B - V)_{obs} - (B - V)_{int} \quad (6.4)$$

where *obs* means the observed color index and *int* means the intrinsic color index (the color the object would have with no reddening). In our own galaxy it is empirically the case that the magnitudes of absorption is proportional to this color excess, i.e.,

$$A_V = R_V E(B - V) \quad (6.5)$$

where R_V is roughly constant and depends upon average grain properties. So assuming that the dust in distant galaxies is similar to the dust in our own, it should be possible to estimate and correct for the dust obscuration. Significantly [48], it appears that there is no difference between $E(B-V)$ for local and distant supernovae. This implies that the distant events are not more or less obscured than the local ones.

2) Grey dust: It is conceivable (but unlikely) that intergalactic space contains dust particles which are significantly larger than the wavelength of light. Such particles would dim but not redden the distant supernovae and so would be undetectable by the method described above [51]. It is here that the very high redshift supernovae ($z > 1$) play an important role. If this is the cause of the apparent dimming we might expect that the supernovae would not become brighter again at higher redshift.

3) Evolution: It is possible that the properties of these events may have evolved with cosmic time. As I mentioned above, the SN exploding at high redshift come from a systematically younger stellar population than the objects observed locally. Moreover, the abundance of metals was smaller in the earlier Universe than now; this evolving composition, by changing the opacity in the outer layers or the composition of the fuel itself could lead to a systematic evolution in peak luminosity. Here it is important to look for observational differences between local and distant supernovae, and there seem to be no significant differences in most respects, the spectrum or the light curve. There is, however, a suggestion that distant supernovae are intrinsically bluer than nearby objects [46]. If this effect is verified, then it could not only point to a systematic difference in the objects themselves, but could also have lead to an underestimate of the degree of reddening in the distant SN. It is difficult, in general, to eliminate the possibility that the events themselves were different in the past and that this could mimic the effect of a cosmological constant

[52]; a deeper theoretical understanding of the SNIa process is required in order to realistically access this possibility.

4) Sample evolution: The sample of SN selected at large redshift may differ from the nearby sample that is used, for example, to calibrate the peak luminosity-decline rate correlation. There does appear to be an absence, at large redshift, of SN with very slowly declining light curves— which is to say, very luminous SN that are seen locally. Perhaps a class of more luminous objects is missing in the more distant Universe due to the fact that these SN emerge from a systematically younger stellar population. One would hope that the luminosity-decline rate correlation would correct for this effect, assuming, of course, that this relation itself does not evolve.

5) Selection biases: There is a dispersion in the luminosity-decline rate relationship, and in a flux-limited sample, one tends to select the higher luminosity objects. Astronomers call this sort of bias the “Malmquist effect” and it is always present in such observational data. Naively, one would expect such a bias to lead to an underestimate of the true luminosity, and, therefore an underestimate luminosity distance; the bias actually diminishes the apparent acceleration. But there is another effect which is more difficult to access: The most distant supernovae are being observed in the UV of their own rest frame. SNIa are highly non-uniform in the UV, and K-corrections are uncertain. This could introduce systematic errors at the level of a few hundredths of a magnitudes [50].

We see that there are a number of systematic effects that could bias these results. A maximum likelihood analysis over the entire sample [50], confirms earlier results that the confidence contours in Ω_m - Ω_A space are stretched along a line $\Omega_A = 1.4\Omega_m + 0.35$ and that the actual best fit is provided by a model with $\Omega_m \approx 0.7$ and $\Omega_A \approx 1.3$ — not the concordance model. Of course, if we add the condition that $\Omega_{tot} = 1$ (a flat Universe) then the preferred model becomes the concordance model. In [50] it is suggested that this apparent deviation is due to the appearance of one or more of the systematic effects discussed above near $z = 1$ at the level of 0.04 magnitudes.

The result that SNIa are systematically dimmer near $z = 0.5$ than expected in a non-accelerating Universe is robust. At the very least it can be claimed with reasonable certainty that the Universe is not decelerating at present. However, given the probable presence of systematic uncertainties at the level of a few hundredths of a magnitude, it is difficult to constrain the equation of state (w) of the dark energy or its evolution (dw/dt) until these effects are better understood. I will just mention that lines of constant age, $t_o H_o$, are almost parallel to the best fit line in the Ω_m - Ω_A plane mentioned above. This then gives a fairly tight constraint on the age in Hubble times [50]; i.e. $t_o H_o = 0.96 \pm 0.4$, which is consistent with the WMAP result. In a near flat Universe this rules out the dominance of matter and requires a dark energy term.

6.4 Number counts of faint galaxies

The final classical test I will discuss is that of number counts of distant objects— what radio astronomers call the log(N)-log(S) test. Basically one counts the number of galaxies N brighter than a certain flux limit S . If we lived in a static Euclidean universe, then the number of galaxies out to distance R would be $N \propto R^3$ but the flux is related to R as $S \propto R^{-2}$. This implies that $N \propto S^{-3/2}$ or $\log(N) = -3/2 \log(S) + \text{const} = 0.6m + \text{const}$. where m is the magnitude corresponding to the flux S .

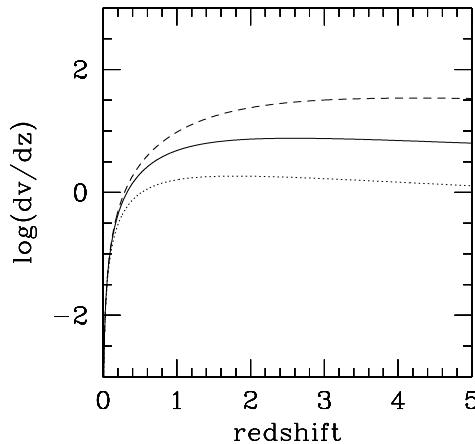


Fig. 9. The log of the incremental volume per incremental redshift (in units of the Hubble volume) as a function of redshift for the three flat cosmological models

But we do not live in a static Euclidean universe; we live in an evolving universe with a non-Euclidean geometry where the differential number counts probe $dV(z)$, the comoving volume as a function of redshift. In Fig. 9 we see $\log(dV/dz)$ as a function of redshift for three different ($\Omega_{tot} = 1$) cosmological models: the matter dominated Universe, the cosmological constant dominated Universe, and the concordance model. For small z , dV/dz increases as z^2 for all models as would be expected in a Euclidean Universe, but by redshift one, the models are obviously diverging, with the models dominated by a cosmological constant having a larger comoving incremental volume. Therefore if we can observe faint galaxies extending out to a redshift of one or two, we might expect number counts to provide a cosmological probe.

There is a long history of counting objects as a function of flux or redshift. Although cosmological conclusions have been drawn (see, e.g. [53]), the overall consensus is that this is not a very good test because the galaxy population evolves strongly with redshift. Galaxies evolve because stars evolve. In the past, the stellar populations were younger and contained relatively more massive, luminous stars. Therefore we expect galaxies to be more luminous at higher redshift. It is also possible that the density of galaxies evolves because of merging, as would be consistent with the preferred model of hierarchical structure formation in the Universe.

The distribution of galaxies by redshift can be used, to some extent, to break this degeneracy between evolution and cosmology. If we can measure the redshifts of galaxies with infrared magnitudes between 23 and 26, for example, that distribution will be skewed toward higher redshift if there is more luminosity evolution.

I have recently reconsidered the number counts of the faint galaxies in the Hubble Deep Fields, north and south [55, 56]. These are two separate small patches of empty sky observed with the Hubble Space Telescope down to a very low flux limit—about $m_I = 30$ (the I band is a far red filter centered around 8000 angstroms). The differential number counts are shown by the solid round points in Fig. 10 where ground based number counts at fainter magnitudes are also shown by the starred points.

For this same sample of galaxies, there are also estimates of the redshifts based upon the galaxy colors—so called photometric redshifts [57]. In order to calculate the expected number counts and redshift distribution one must have some idea of the form of the luminosity function—the distribution of galaxies by redshift. Here, like everyone else, I have assumed that this form is given by the Schechter function [58]:

$$N(L)dL = N_o(L/L_*)^{-\alpha} \exp(-L/L_*)dL \quad (6.6)$$

which is characterized by three parameters: α , a power law at low luminosities, L_* a break-point above which the number of galaxies rapidly decreases, and N_o a normalization. I take this form because the overall galaxy distribution by luminosity at low redshifts is well fit by such a law [59], so I am assuming that at least the form of the luminosity function does not evolve with redshift.

But when I consider faint galaxies at high redshift in a particular band I have to be careful to apply the K-correction mentioned above; that is, I must correct the observed flux in that band to the rest frame. Making this correction [60], but assuming no luminosity or density evolution, I find the differential number counts appropriate to our three flat cosmological models shown by the indicated curves in Fig. 10. We see that the predicted number counts all fall short of the observed counts, but that the cosmological constant dominated model comes closest to matching the observations. However, the distribution by redshift of HDF galaxies between I-band magnitudes of 22 and 26 is shown in Fig. 11 (this is obviously the cumulative distribution). Here we

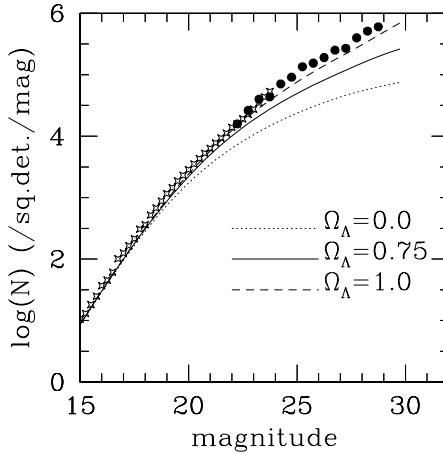


Fig. 10. The solid points are the faint galaxy number counts from the Hubble Deep Fields (north and south [55, 56]) and the star shaped points are the number counts from ground based data. The curves are the no-evolution predictions from three flat cosmological models.

see that all three models seriously fail to match the observed distribution, in the sense that the predicted mean redshift is much too small.

This problem could obviously be solved by evolution. If galaxies are brighter in the past, as expected, then we would expect to shift this distribution toward higher redshifts. One can conceive of very complicated evolution schemes, involving initial bursts of star formation with or without continuing star formation, but it would seem desirable to keep the model as simple as possible; let's take a "minimalist" model for galaxy evolution. A simple one parameter scheme with the luminosity brightening proportional to the look-back time squared, i.e., every galaxy brightens as

$$\Delta M_I = q (H_0 t_{lb})^2 \quad (6.7)$$

where q is the free parameter, can give a reasonable match to evolution models for galaxies [60]. (we also assume that all galaxies are the same— they are not divided into separate morphological classes). I choose the value of q such that the predicted redshift distribution most closely matches the observed distribution for all three models, and the results are shown in Fig. 12.

The required values of q (in magnitudes per t_H^{-2}) for the three cosmological models are: $q = 2.0$ ($\Omega_\Lambda = 1.0$), $q = 3.0$ ($\Omega_\Lambda = 0.7$), and $q = 11.0$ ($\Omega_\Lambda = 0.0$).

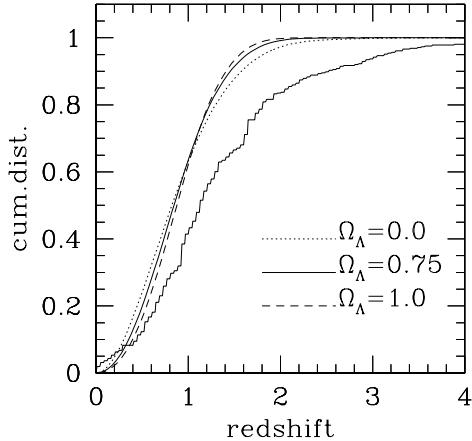


Fig. 11. The cumulative redshift distribution for galaxies between apparent I-band magnitudes of 23 and 26 (photometric redshifts from [57]). The curves are the predicted no-evolution distributions for the three cosmological models.

Obviously, the matter-dominated model requires the most evolution, and with this simple evolution scheme, cannot be made to perfectly match the observed distribution by redshift (this in itself is not definitive because one could always devise more complicated schemes which would work). For the concordance model, the required evolution would be about two magnitudes out to $z = 3$.

For these same evolutionary models, that is, with evolution sufficient to match the number counts, the predicted redshift distributions are shown in Fig. 13. Here we see that the model dominated by a cosmological constant predicts too many low redshift galaxies, the matter dominated model predicts too few, and the model that works perfectly is very close to the concordance model! Preforming this operation for a number of flat models with variable Ω_Λ , I find that $0.59 < \Omega_\Lambda < 0.71$ to 90% confidence.

Now there are too many assumptions and simplifications to make this definitive. The only point I want to make is that faint galaxy number counts and redshift distributions are completely consistent with the concordance model when one considers the simplest minimalist model for pure luminosity evolution. One may certainly conclude that number counts provide no contradiction to the generally accepted cosmological model of the Universe (to my disappointment).

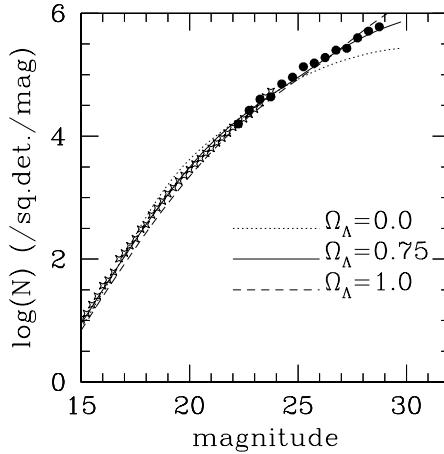


Fig. 12. As in Fig. 10 above the observed galaxy number counts and the predictions for the cosmological models with luminosity evolution sufficient to explain the number counts.

7 Conclusions

In these lectures I have been looking for discord, but have not found it. The classical tests return results for cosmological parameters that are consistent with but considerably less precise than those implied by the CMB anisotropies, given the usual assumptions. It is fair to say that the numbers characterizing the concordance model, $\Omega_m \approx 0.3$, $\Omega_\Lambda \approx 0.7$ are robust *in the context of the framework of FRW cosmology*. It is, in fact, the peculiar composition of the Universe embodied by these numbers which calls that framework into question.

Rather small changes in the assumptions underlying pure FRW cosmology (with only an evolving vacuum energy density in addition to more familiar fluids) can make a difference. For example, allowing $w = -0.6$ brings the number counts and z-distribution of faint galaxies into agreement with a Universe strongly dominated by dark energy ($\Omega_Q = 0.9$). The same also true of the high-z supernovae observations [50]). Allowing a small component of correlated iso-curvature initial perturbations, as expected in braneworld cosmologies, can affect the amplitudes and positions of the peaks in the angular power spectrum of the CMB anisotropies [3], and therefore the derived cosmological parameters.

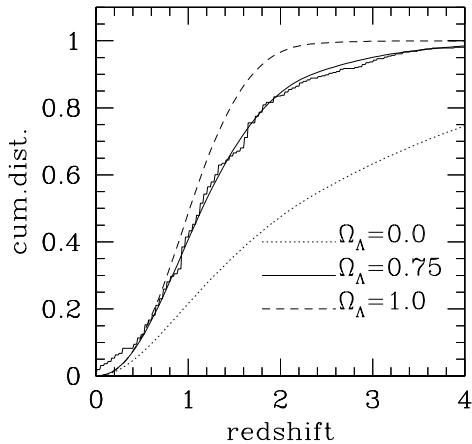


Fig. 13. The cumulative redshift distribution for galaxies between apparent i-band magnitudes of 22 and 26 (photometric redshifts from [57]). The curves are the predicted distributions for the three cosmological models with evolution sufficient to explain the number counts.

But even more drastic changes have been suggested. Certain braneworld scenarios, for example, in which 4-D gravity is induced on the brane [61] imply that gravity is modified at large scale where gravitons begin to leak into the bulk [62]. It is possible that the observed acceleration is due to such modifications and not to dark energy. More ad hoc modifications of General Relativity [5] have also been proposed because of a general unease with dark energy—proposals whereby gravity is modified in the limit of small curvature scalar. My own opinion is that we should also feel uneasy with the mysterious non-baryonic cold dark matter, because the only evidence for its existence, at present, is its gravitational influence; when the theory of gravity is modified to eliminate dark energy, it might also be found that the need for dark matter vanishes.

In general, more attention is being given to so-called infrared modifications of gravity (e.g. [63]), and this is a positive development. High energy modifications, that affect the evolution of the early Universe, are, as we have seen, strongly constrained by considerations of primordial nucleosynthesis (now, in combination with the CMB results). It is more likely that modifications play a role in the late, post-recombination evolution of the Universe, where the peculiarities of the concordance model suggest that they are needed. The fact

that the same rather un-natural values for the comparable densities of dark energy and matter keep emerging in different observational contexts may be calling attention to erroneous underlying assumptions rather than to the actual existence of these “ethers”.

Convergence toward a parameterized cosmology is not, without deeper understanding, sufficient reason for triumphalism. Rather, it should be a motivation to look more carefully at the possible systematic effects in the observations and to question more critically the underlying assumptions of the models.

I thank Rien van de Weygaert, Ole Möller, Moti Milgrom, Art Wolfe, Jacob Bekenstein, and Scott Trager for useful comments on the manuscript. I also thank Gary Steigman, Wendy Freedman, and Luis Ho for permission to use Figs. 1 and 2. I am very grateful to the organizers of the Second Aegean Summer School on the Early Universe, and especially, Lefteris Papantonopoulos, for all their work and for inviting me to the very pleasant island of Syros.

References

1. D.N. Spergel, et al.: *Astrophys.J.Suppl.*, 148, 175 (2003) [astro-ph/0302209]
2. J.P. Ostriker, P.J. Steinhardt, PJ.: *Nature* 377, 600 (1995)
3. R. Maartens: gr-qc/0312059
4. Deffayet C: *Phys.Lett. B* 502, 199 (2001)
5. S.M. Carroll, V. Duvvuri, M. Trodden, M.S. Turner: astro-ph/0306438 (2003)
6. J.F. Navarro, C.S. Frenk, S.D.M. White: *Astrophys.J.* 462: 563 (1996)
7. W.J.G. de Blok, S.S. McGaugh, V.C. Rubin: *Astron.J.*, 122, 2396 (2001)
8. T. Treu, L.V.E.Koopmans, D.J. Sand, G.P. Smith, R.S. Ellis: To appear in the proceedings of IAU Symposium 220 "Dark matter in galaxies", S. Ryder, D.J. Pisano, M. Walker, and K. Freeman eds [astro-ph/0311052].
9. S.M. Carroll: *Living Rev.Rel.* 4, 1 (2001) [astro-ph/0004075]
10. P.J.E. Peebles, B. Ratra': *Rev.Mod.Phys.*, 75, 559 (2003)
11. J.A. Peacock: *Phil.Trans.Roy.Soc.Lond.* A361 2479 (2003) [astro-ph/0309238]
12. H. Hoekstra, H. Yee, M. Gladders: *Astrophys.J.* 577, 595 (2002)
13. M. Milgrom: *Astrophys.J.* , 270, 365 (1983)
14. W. Rindler: *Essential Relativity* (Heidelberg, Springer-Verlag)
15. R.R. Caldwell: *Phys.Lett.* B545, 23 (2002)
16. J.C. Mather et al.: *Astrophys.J.* , 512, 511 (1999)
17. P. Molaro, S.A. Levshakov, M. Dessauges-Zavadsky, S. D'Odorico: *Astron.Astrophys.*, 381, 64 (2002) [astro-ph/0111589]
18. G. Steigman: Carnegie Observatories Astrophysics Series Vol. 2: *Measuring and Modeling the Universe*, W. Freedman ed. (Cambridge, Cambridge Univ. Press) (2003) [astro-ph/ 0307244]
19. S. Burles, D. Tytler: *Astrophys.J.* , 499, 699 (1998)
20. J.M. O'Meara et al.: *Astrophys.J.* , 552, 718 (2001)

21. M. Pettini, D.V. Bowen: *Astrophys.J.*, 560, 41 (2001)
22. D. Kirkman, et al.: *Astrophys.J.Suppl.*, in press (2003), [astro-ph/0302006]
23. P.Jakobsen et al.: *Astrophys.J.*, 417, 528 (1993)
24. K. Olive, G. Steigman: *Astrophys.J.Supp.*, 97, 49 (1995)
25. Y.I. Izotov, T.X. Thuan: *Astrophys.J.*, 500, 188 (1998)
26. V. Barger, J.P. Kneller, P. Langacker, D. Marfatia, G. Steigman: *Phys.Lett.* B569 123 (2003) [hep-ph/0306061]
27. C. Barcelo, M. Visser: *Phys.Lett.* B482, 183 (2000) [hep-th/0004056]
28. W.L. Freedman et al.: *Astrophys.J.*, 553, 47 (2001)
29. R.A. Sunyaev, Ya.B. Zeldovich: *Ann.Rev.Astron. Astrophys.*, 18, 537 (1980)
30. J.E. Carlstrom, G.P. Holder, E.D. Reese: *Ann.Rev.Astron. Astrophys.*, 40, 643 (2002)
31. S. Refsdal: *Mon.Not.RAS*, 128, 307 (1964)
32. L.L.R. Williams, P. Saha: *Astron.J.*, 119, 439 (2000)
33. L.V.E. Koopmans, C.D. Fassnacht: *Astrophys.J.*, 527, 513 (1999)
34. C.S. Kochanek, P.L. Schechter: to appear in Measuring and Modeling the Universe (Carnegie Observatories Astrophysics Series, Vol. 2), ed. W.L. Freedman (2003) [astro-ph/0306040]
35. L.V.E. Koopmans et al.: *Astrophys.J.*, 599, 70 (2003)
36. B. Chaboyer, P. Demarque, P.J. Kernan, L.M. Krauss: *Astrophys.J.*, 494, 96 (1998)
37. R. Cayrel et al.: *Nature*, 409, 691 (2001) [astro-ph/0104357]
38. G.K. Miley: *Mon.Not.RAS*, 152, 477 (1971)
39. K.I. Kellerman: *Nature*, 361, 134 (1993)
40. L.I. Gurvits, K.I. Kellerman, S. Frey: *Astron.Astrophys*, 342, 378 (1999)
41. U. Seljak, M. Zeldarriaga: *Astrophys.J.*, 469, 437 (1996)
42. L. Page et al.: *Astrophys.J.Suppl.*, 148, 233 (2003)
43. A. Kogut et al.: *Astrophys.J.Suppl.* 148, 161 (2003)
44. B.A. Bassett, M. Kunz, D. Parkinson, C. Ungarelli: *Phys.Rev.D* 68, 043504 (2003)
45. D. Branch: *Ann.Rev.Astron.Astrophys.*, 36, 17 (1998)
46. B. Leibundgut: *Ann.Rev.Astron.Astrophys.*, 39, 67 (2001)
47. M. Hamuy, M.M. Phillips, R.A. Schommer, N.B. Suntzeff: *Astron.J.*, 112, 2391 (1996)
48. Perlmutter S, et al. 1999, *Astrophys.J.*, 517, 565
49. P.M. Garnavitch et al.'': *Astrophys.J.*, 493, 53 (1998)
50. J.L. Tonry et al.: *Astrophys.J.*, 594, 1 (2003) [astro-ph/0305008]
51. A. Aguirre, Z. Haiman: *Astrophys.J.*, 532, 28 (2000)
52. P.S. Drell, T.J. Loredo, I. Wasserman: *Astrophys.J.*, 530, 593 (2000)
53. M. Fukugita, F. Takahara, K. Yamashita, Y. Yoshi: *Astrophys.J.*, 361, L1 (1990)
54. Y. Yoshii, F. Takahara: *Astrophys.J.*, 346, 28 (1989)
55. R.E. Williams et al.: *Astron.J.*, 112, 1335 (1996)
56. S. Casertano et al.: *Astron.J.*, 120, 247 (2000)
57. A. Fernández-Soto, K.M. Lanzetta, A. Yahil: *Astrophys.J.*, 513, 34 (1999)
58. P.L. Schechter: *Astrophys.J.*, 203, 297 (1976)
59. M.R. Blanton et al.: *Astron.J.*, 121, 2358 (2001)
60. B.M. Poggianti: *Astron.Astrophys.Suppl.*, 122, 399 (1997)
61. G. Dvali, G. Gabadadze, M. Poratti: *Phys.Lett.*, B485, 208 (2000) [hep-th/0005016]
62. C. Deffayet, G. Dvali, G. Gabadadze: *Phys.Rev.* D65 044023 (2002) [astro-ph/0105068]
63. N. Arkani-Hamed, H-C. Cheng, M.A. Luty, S. Mukohyama: hep-th/0312099

Anisotropies in the Cosmic Microwave Background

Anthony Challinor

Astrophysics Group, Cavendish Laboratory, Madingley Road,
Cambridge, CB3 0HE, U.K. a.d.challinor@mrao.cam.ac.uk

Abstract. The linear anisotropies in the temperature of the cosmic microwave background (CMB) radiation and its polarization provide a clean picture of fluctuations in the universe some 370 kyr after the big bang. Simple physics connects these fluctuations with those present in the ultra-high-energy universe, and this makes the CMB anisotropies a powerful tool for constraining the fundamental physics that was responsible for the generation of structure. Late-time effects also leave their mark, making the CMB temperature and polarization useful probes of dark energy and the astrophysics of reionization. In this review we discuss the simple physics that processes primordial perturbations into the linear temperature and polarization anisotropies. We also describe the role of the CMB in constraining cosmological parameters, and review some of the highlights of the science extracted from recent observations and the implications of this for fundamental physics.

1 Introduction

The cosmic microwave background (CMB) radiation has played an essential role in shaping our current understanding of the large-scale properties of the universe. The discovery of this radiation in 1965 by Penzias and Wilson [1], and its subsequent interpretation as the relic radiation from a hot, dense phase of the universe [2] put the hot big bang model on a firm observational footing. The prediction of angular variations in the temperature of the radiation, due to the propagation of photons through an inhomogeneous universe, followed shortly after [3], but it was not until 1992 that these were finally detected by the Differential Microwave Radiometers (DMR) experiment on the Cosmic Background Explorer (COBE) satellite [4]. The fractional temperature anisotropies are at the level of 10^{-5} , consistent with structure formation in cold dark matter (CDM) models [5, 6], but much smaller than earlier predictions for baryon-dominated universes [3, 7]. Another experiment on COBE, the Far InfraRed Absolute Spectrophotometer (FIRAS), spectacularly confirmed the black-body spectrum of the CMB and determined the (isotropic) temperature to be 2.725 K [8, 9].

In the period since COBE, many experiments have mapped the CMB anisotropies on a range of angular scales from degrees to arcminutes (see [10] for a recent review), culminating in the first-year release of all-sky data from the Wilkinson Microwave Anisotropy Probe (WMAP) satellite in February 2003 [11]. The observed modulation in the amplitude of the anisotropies with angular scale is fully consistent with predictions based on coherent, acoustic oscillations [7], derived from gravitational instability of initially adiabatic density perturbations in a universe with nearly-flat spatial sections. The amplitude and scale of these acoustic features has allowed many of the key cosmological parameters to be determined with unprecedented precision [12], and a strong concordance with other cosmological probes has emerged.

In this review we describe the essential physics of the temperature anisotropies of the CMB, and its recently-detected polarization [13], and discuss how these are used to constrain cosmological models. For reviews that are similar in spirit, but from the pre-WMAP era see e.g. [14, 15]. We begin in Sect. 2 with the fundamentals of CMB physics, presenting the kinetic theory of the CMB in an inhomogeneous universe, and the various physical mechanisms that process initial fluctuations in the distribution of matter and spacetime geometry into temperature anisotropies. Section 3 discusses the effect of cosmological parameters on the power spectrum of the temperature anisotropies, and the limits to parameter determination from the CMB alone. The physics of CMB polarization is reviewed in Sect. 4, and the additional information that polarization brings over temperature anisotropies alone is considered. Finally, in Sect. 5 we describe some of the scientific highlights that have emerged from recent CMB observations, including the detection of CMB polarization, implications for inflation, and the direct signature of dark-energy through correlations between the large-scale anisotropies and tracers of the mass distribution in the local universe. Throughout, we illustrate our discussion with computations based on Λ CDM cosmologies, with baryon density $\Omega_b h^2 = 0.023$ and cold dark matter density $\Omega_c h^2 = 0.111$. For flat models we take the dark-energy density parameter to be $\Omega_\Lambda = 0.75$ giving a Hubble parameter $H_0 = 73 \text{ km s}^{-1} \text{ Mpc}^{-1}$. We adopt units with $c = 1$ throughout, and use a spacetime metric signature $+$ $-$ $-$ $-$.

2 Fundamentals of CMB Physics

In this section we aim to give a reasonably self-contained review of the essential elements of CMB physics.

2.1 Thermal History and Recombination

The high temperature of the early universe maintained a low equilibrium fraction of neutral atoms, and a correspondingly high number density of free electrons. Coulomb scattering between the ions and electrons kept them in local

kinetic equilibrium, and Thomson scattering of photons tended to maintain the isotropy of the CMB in the baryon rest frame. As the universe expanded and cooled, the dominant element hydrogen started to recombine when the temperature fell below $\sim 4000\text{K}$ – a factor of 40 lower than might be anticipated from the 13.6-eV ionization potential of hydrogen, due to the large ratio of the number of photons to baryons. The details of recombination are complicated since the processes that give rise to net recombination occur too slowly to maintain chemical equilibrium between the electrons, protons and atoms during the later stages of recombination [16, 17] (see [18] for recent refinements). The most important quantity for CMB anisotropy formation is the visibility function – the probability that a photon last scattered as a function of time. The visibility function peaks around $\sim 370\text{ kyr}$ after the big bang, and has a width $\sim 115\text{ kyr}$, a small fraction of the current age $\sim 13.5\text{ Gyr}$ [12]. After recombination, photons travelled mostly unimpeded through the inhomogeneous universe, imprinting fluctuations in the radiation temperature, the gravitational potentials, and the bulk velocity of the radiation where they last scattered, as the temperature anisotropies that we observe today. A small fraction of CMB photons (current results from CMB polarization measurements [19] indicate around 20 per cent; see also Sect. 5.1) underwent further scattering once the universe reionized due to the ionizing flux from the first non-linear structures.

2.2 Statistics of CMB Anisotropies

The spectrum of the CMB brightness along any direction $\hat{\mathbf{n}}$ is very nearly thermal with a temperature $T(\hat{\mathbf{n}})$. The temperature depends only weakly on direction, with fluctuations $\Delta T(\hat{\mathbf{n}})$ at the level of 10^{-5} of the average temperature $T = 2.725\text{ K}$. It is convenient to expand the temperature fluctuation in spherical harmonics,

$$\Delta T(\hat{\mathbf{n}})/T = \sum_{lm} a_{lm} Y_{lm}(\hat{\mathbf{n}}), \quad (1)$$

with $a_{lm}^* = (-1)^m a_{l-m}$ since the temperature is a real field. The sum in (1) runs over $l \geq 1$, but the dipole ($l = 1$) is usually removed explicitly when analysing data since it depends linearly on the velocity of the observer. Multipoles at l encode spatial information with characteristic angular scale $\sim \pi/l$.

The statistical properties of the fluctuations in a perturbed cosmology can be expected to respect the symmetries of the background model. In the case of Robertson–Walker models, the rotational symmetry of the background ensures that the multipoles a_{lm} are uncorrelated for different values of l and m :

$$\langle a_{lm} a_{l'm'}^* \rangle = C_l \delta_{ll'} \delta_{mm'}, \quad (2)$$

which defines the power spectrum C_l . The angle brackets in this equation denote the average over an ensemble of realisations of the fluctuations. The

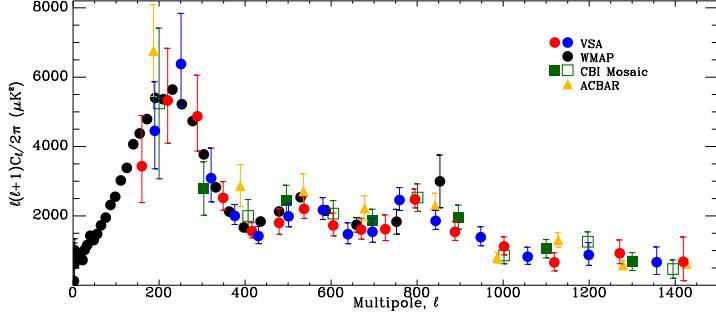


Fig. 1. Compilation of CMB anisotropy measurements (as of February 2004) from WMAP (black filled circles), the Very Small Array (VSA [20]; shaded circles representing two interleaving binning schemes), the Cosmic Background Imager (CBI [21, 22]; open and filled squares for two different binning schemes) and the Arcminute Cosmology Bolometer Array Receiver (ACBAR [23]; triangles). (Figure reproduced, with permission, from [20].)

simplest models of inflation predict that the fluctuations should also be Gaussian at early times, and this is preserved by linear evolution of the small fluctuations. If Gaussian, the a_{lm} s are also independent, and the power spectrum provides the complete statistical description of the temperature anisotropies. For this reason, measuring the anisotropy power spectrum has, so far, been the main goal of observational CMB research. Temperature anisotropies have now been detected up to l of a few thousand; a recent compilation of current data as of February 2004 is given in Fig. 1.

The correlation between the temperature anisotropies along two directions evaluates to

$$\langle \Delta T(\hat{n}_1) \Delta T(\hat{n}_2) \rangle = T^2 \sum_l \frac{2l+1}{4\pi} C_l P_l(\cos \theta) , \quad (3)$$

which depends only on the angular separation θ as required by rotational invariance. Here, $P_l(x)$ are the Legendre polynomials. The mean-square temperature anisotropy is

$$\langle \Delta T^2 \rangle = T^2 \sum_l \frac{2l+1}{4\pi} C_l \approx T^2 \int \frac{l(l+1)}{2\pi} C_l d \ln l , \quad (4)$$

so that the quantity $l(l+1)C_l/2\pi$, which is conventionally plotted, is approximately the power per decade in l of the temperature anisotropies.

2.3 Kinetic Theory

The CMB photons can be described by a one-particle distribution function $f(x^a, p^a)$ that is a function of the spacetime position x^a and four-momentum

p^a of the photon. It is defined such that the number of photons contained in a proper three-volume element $d^3\mathbf{x}$ and with three-momentum in $d^3\mathbf{p}$ is $f d^3\mathbf{x} d^3\mathbf{p}$. The phase-space volume element $d^3\mathbf{x} d^3\mathbf{p}$ is Lorentz-invariant and is conserved along the photon path through phase space (see, e.g. [24]). It follows that f is also frame-invariant, and is conserved in the absence of scattering. To calculate the anisotropies in the CMB temperature, we must evolve the photon distribution function in the perturbed universe.

To avoid over-complicating our discussion, we shall only consider spatially-flat models here, and, for the moment, ignore the effects of polarization. For a more complete discussion, including these complications, see e.g. [25, 26]. Curvature mostly affects the CMB through the geometrical projection of linear scales at last scattering to angular scales on the sky today, but has a negligible impact on pre-recombination physics and hence much of the discussion in this section. The subject of cosmological perturbation theory is rich in methodology, but, for pedagogical reasons, we adopt here the most straightforward approach which is to work directly with the metric perturbations. This is also the most prevalent in the CMB literature. The 1+3-covariant approach [27] is a well-developed alternative that is arguably more physically-transparent than metric-based techniques. It has also been applied extensively in the context of CMB physics [26, 28, 29, 30, 31, 32]. The majority of our discussion will be of scalar perturbations, where all perturbed three-tensors can be derived from the spatial derivatives of scalar functions, although we discuss tensor perturbations briefly in Sect. 2.5.

For scalar perturbations in spatially-flat models we can choose a gauge such that the spacetime metric is [33]

$$ds^2 = a^2(\eta)[(1 + 2\psi)d\eta^2 - (1 - 2\phi)d\mathbf{x}^2], \quad (5)$$

where η is conformal time (related to proper time t by $dt = ad\eta$), a is the scale factor in the background model and, now, \mathbf{x} is comoving position. This gauge, known as the conformal Newtonian or longitudinal gauge, has the property that the congruence of worldlines with constant \mathbf{x} have zero shear. The two scalar potentials ϕ and ψ constitute the scalar perturbation to the metric, with ϕ playing a similar role to the Newtonian gravitational potential. In the absence of anisotropic stress, ϕ and ψ are equal. We parameterise the photon four-momentum with its energy ϵ/a and direction \mathbf{e} (with $\mathbf{e}^2 = 1$), as seen by an observer at constant \mathbf{x} , so that

$$p^\mu = a^{-2}\epsilon[1 - \psi, (1 + \phi)\mathbf{e}]. \quad (6)$$

Free photons move on the geodesics of the perturbed metric, $p^\mu \nabla_\mu p^\nu = 0$, so the energy and direction evolve as

$$\frac{d\epsilon}{d\eta} = -\epsilon d\psi/d\eta + \epsilon(\dot{\phi} + \dot{\psi}), \quad (7)$$

$$d\mathbf{e}/d\eta = -\boldsymbol{\nabla}_\perp(\phi + \psi), \quad (8)$$

where dots denote $\partial/\partial\eta$ and ∇_\perp is the three-gradient projected perpendicular to \mathbf{e} . We see immediately that ϵ is conserved in the absence of perturbations, so that the energy redshifts in proportion to the scale factor in the background model. The change in direction of the photon due to the projected gradient of the potentials in the perturbed universe gives rise to gravitational lensing (see e.g. [34] for a review).

The dominant scattering mechanism to affect CMB anisotropies is classical Thomson scattering off free electrons, since around recombination the average photon energy is small compared to the rest mass of the electron. Furthermore, the thermal distribution of electron velocities can be ignored due to the low temperature. The evolution of the photon distribution function in the presence of Thomson scattering is

$$\begin{aligned} \frac{df}{d\eta} = & -a(1+\psi)n_e\sigma_T f + \frac{3}{16\pi}a(1+\psi)n_e\sigma_T \int f(\epsilon, \mathbf{e}')[(1+(\mathbf{e}\cdot\mathbf{e}')^2] d\mathbf{e}' \\ & - an_e\sigma_T \mathbf{e}\cdot\mathbf{v}_b \epsilon \frac{\partial f}{\partial\epsilon}, \end{aligned} \quad (9)$$

where n_e is the electron (proper) number density, σ_T is the Thomson cross section, and the electron peculiar velocity is $\mathbf{v}_b = d\mathbf{x}/d\eta$. The derivative on the left of (9) is along the photon path in phase space:

$$\frac{df}{d\eta} = \frac{\partial f}{\partial\eta} + \mathbf{e}\cdot\nabla f + (\dot{\phi} - \mathbf{e}\cdot\nabla\psi)\epsilon \frac{\partial f}{\partial\epsilon} \quad (10)$$

to first order, where we have used (7) and (8) and the fact that the anisotropies of f are first order. The first term on the right of (9) describes scattering out of the beam, and the second scattering into the beam. The final term arises from the out-scattering of the additional dipole moment in the distribution function seen by the electrons due to the Doppler effect. In the background model f is isotropic and the net scattering term vanishes, so that f is a function of the conserved ϵ only: $f = \bar{f}(\epsilon)$. Thermal equilibrium ensures that \bar{f} is a Planck function.

The fluctuations in the photon distribution function inherit an energy dependence $\epsilon\partial\bar{f}/\partial\epsilon$ from the source terms in the Boltzmann equation (9). Separating out the background contribution to f , and its energy dependence, we can write

$$f(\eta, \mathbf{x}, \epsilon, \mathbf{e}) = \bar{f}(\epsilon)[1 - \Theta(\eta, \mathbf{x}, \mathbf{e})d\ln\bar{f}/d\ln\epsilon], \quad (11)$$

so that the CMB spectrum is Planckian but with a direction-dependent temperature $\Delta T/T = \Theta$. Using the Lorentz invariance of f , it is not difficult to show that the quadrupole and higher moments of Θ are gauge-invariant. If we now substitute for f in (9), we find the Boltzmann equation for Θ :

$$\begin{aligned} \frac{\partial(\Theta + \psi)}{\partial\eta} + \mathbf{e}\cdot\nabla(\Theta + \psi) = & -an_e\sigma_T\Theta + \frac{3}{16\pi}an_e\sigma_T \int \Theta(\mathbf{e}')[(1+(\mathbf{e}\cdot\mathbf{e}')^2] d\mathbf{e}' \\ & + an_e\sigma_T \mathbf{e}\cdot\mathbf{v}_b + \dot{\phi} + \dot{\psi}. \end{aligned} \quad (12)$$

The formal solution of this equation is an integral along the line of sight $\hat{\mathbf{n}} = -\mathbf{e}$,

$$[\Theta(\hat{\mathbf{n}}) + \psi]_R = e^{-\tau} [\Theta(\hat{\mathbf{n}}) + \psi]_E + \int_E^R e^{-\tau} S d\eta , \quad (13)$$

where R is the reception event, E is the emission event, and $\tau \equiv \int a n_e \sigma_T d\eta$ is the optical depth back from R . The source term S is given by the right-hand side of (12), but with Θ replaced by $-\psi$ in the first term.

We gain useful insight into the physics of anisotropy formation by approximating the last scattering surface as sharp (which is harmless on large angular scales), and ignoring the quadrupole CMB anisotropy at last scattering. In this case (13) reduces to

$$[\Theta(\hat{\mathbf{n}}) + \psi]_R = \Theta_0|_E + \psi|_E - \hat{\mathbf{n}} \cdot \mathbf{v}_b|_E + \int_E^R (\dot{\psi} + \dot{\phi}) d\eta , \quad (14)$$

where Θ_0 is the isotropic part of Θ , and is proportional to the fluctuation in the photon energy density. The various terms in this equation have a simple physical interpretation. The temperature received along direction $\hat{\mathbf{n}}$ is the isotropic temperature of the CMB at the last scattering event on the line of sight, Θ_0 , corrected for the gravitational redshift due to the difference in potential between E and R , and the Doppler shift $\mathbf{e} \cdot \mathbf{v}_b|_E$ resulting from scattering off moving electrons. Finally, there is an additional gravitational redshift contribution arising from evolution of the gravitational potentials [3].

Machinery for an Accurate Calculation

An accurate calculation of the CMB anisotropy on all scales where linear perturbation theory is valid requires a full numerical solution of the Boltzmann equation. The starting point is to expand $\Theta(\theta, \mathbf{x}, \mathbf{e})$ in appropriate basis functions. For scalar perturbations, these are the contraction of the (irreducible) trace-free tensor products $e^{\langle i_1 \dots i_l \rangle}$ (the angle brackets denoting the trace-free part) with trace-free (spatial) tensors derived from derivatives of scalars [28, 31, 35]. Fourier expanding the scalar functions, we end up forming contractions between $e^{\langle i_1 \dots i_l \rangle}$ and $\hat{k}_{\langle i_1} \dots \hat{k}_{i_l \rangle}$ where $\hat{\mathbf{k}}$ is the wavevector. These contractions reduce to Legendre polynomials of $\hat{\mathbf{k}} \cdot \mathbf{e}$, and so the normal-mode expansion of Θ for scalar perturbations takes the form

$$\Theta(\eta, \mathbf{x}, \mathbf{e}) = \sum_{l \geq 0} \int \frac{d^3 k}{(2\pi)^{3/2}} (-i)^l \Theta_l(\eta, \mathbf{k}) P_l(\hat{\mathbf{k}} \cdot \mathbf{e}) e^{i \mathbf{k} \cdot \mathbf{x}} . \quad (15)$$

It is straightforward to show that the implied azimuthal symmetry about the wavevector is consistent with the Boltzmann equation (12). Inserting the expansion of Θ into this equation gives the Boltzmann hierarchy for the moments Θ_l :

$$\dot{\Theta}_l + k \left(\frac{l+1}{2l+3} \Theta_{l+1} - \frac{l}{2l-1} \Theta_{l-1} \right) = a n_e \sigma_T \left[(\delta_{l0} - 1) \Theta_l - \delta_{l1} v_b + \frac{1}{10} \Theta_2 \right] + \delta_{l0} \dot{\phi} + \delta_{l1} k \psi, \quad (16)$$

where $v_b = \int i\hat{k}v_b(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}} d^3k/(2\pi)^{3/2}$, and ϕ and ψ are the Fourier transforms of the potentials. This system of ordinary differential equations can be integrated directly with the linearised Einstein equations for the metric perturbations, and the fluid equations governing perturbations in the other matter components, as in the publically-available COSMICS code [33]. Careful treatment of the truncation of the hierarchy is necessary to avoid unphysical reflection of power back down through the moments.

A faster way to solve the Boltzmann equation numerically is to use the line-of-sight solution (13), as in the widely-used CMBFAST code [36] and its parallelised derivative CAMB [37]. Inserting the expansion (15) gives the integral solution to the hierarchy

$$\Theta_l|_{\eta_0} = (2l+1) \int_0^{\eta_0} d\eta e^{-\tau} \left[(\dot{\phi} + \dot{\psi}) j_l(k\Delta\eta) - \dot{\tau}(\Theta_0 + \psi) j_l(k\Delta\eta) + \dot{\tau} v_b j'_l(k\Delta\eta) - \frac{1}{20} \dot{\tau} \Theta_2 (3j''_l + j_l)(k\Delta\eta) \right], \quad (17)$$

where $\Delta\eta \equiv \eta_0 - \eta$, j_l is a spherical Bessel function, and primes denote derivatives with respect to the argument. Using the integral solution, it is only necessary to evolve the Boltzmann hierarchy to modest l to compute accurately the source terms that appear in the integrand. The integral approach is thus significantly faster than a direct solution of the hierarchy.

The spherical multipoles a_{lm} of the temperature anisotropy can be extracted from (15) as

$$a_{lm} = 4\pi i^l \int \frac{d^3k}{(2\pi)^{3/2}} \frac{\Theta_l}{2l+1} Y_{lm}^*(\hat{\mathbf{k}}) e^{i\mathbf{k}\cdot\mathbf{x}}. \quad (18)$$

Statistical homogeneity and isotropy imply that the equal-time correlator

$$\langle \Theta_l(\eta, \mathbf{k}) \Theta_l^*(\eta, \mathbf{k}') \rangle = \frac{2\pi^2}{k^3} \Theta_l^2(\eta, k) \delta(\mathbf{k} - \mathbf{k}'), \quad (19)$$

so forming the correlation $\langle a_{lm} a_{l'm'}^* \rangle$ gives the power spectrum

$$C_l = \frac{4\pi}{(2l+1)^2} \int \Theta_l^2(k) d\ln k. \quad (20)$$

If we consider (pure) perturbation modes characterised by a single independent stochastic amplitude per Fourier mode (such as the comoving curvature for the adiabatic mode; see Sect. 2.4), the power $\Theta_l^2(k)$ is proportional to the power spectrum of that amplitude. The spherical Bessel functions in (17) peak sharply at $k\Delta\eta = l$ for large l , so that multipoles l are mainly probing

spatial structure with wavenumber $k \sim l/\Delta\eta$ at last scattering. The oscillatory tails of the Bessel functions mean that some power from a given k does also enter larger scale anisotropies. Physically, this arises from Fourier modes that are not aligned with their wavevector perpendicular to the line of sight. As we discuss in the next section, the tightly-coupled system of photons and baryons undergoes acoustic oscillations prior to recombination on scales inside the sound horizon. For the pure perturbation modes, all modes with a given wavenumber reach the maxima or minima of their oscillation at the same time, irrespective of the direction of \mathbf{k} , and so we expect modulation in the C_l s on sub-degree scales. The first three of these acoustic peaks have now been measured definitively; see Fig. 1.

2.4 Photon–Baryon Dynamics

Prior to recombination, the mean free path of CMB photons is $\sim 4.9 \times 10^4 (\Omega_b h^2)^{-1} (1+z)^{-2}$ Mpc. On comoving scales below this length the photons and baryons behave as a tightly-coupled fluid, with the CMB almost isotropic in the baryon frame. In this limit, only the $l=0$ and $l=1$ moments of the distribution function are significant.

The stress-energy tensor of the photons is given in terms of the distribution function by

$$T^{\mu\nu} = a^{-2} \int f(\eta, \mathbf{x}, \epsilon, \mathbf{e}) p^\mu p^\nu \epsilon \mathrm{d}\mathbf{e} , \quad (21)$$

so that the Fourier modes of the fractional over-density of the photons are $\delta_\gamma = 4\Theta_0$ and the photon (bulk) velocity $v_\gamma = -\Theta_1$. The anisotropic stress is proportional to Θ_2 . In terms of these variables, the first two moment equations of the Boltzmann hierarchy become

$$\dot{\delta}_\gamma - \frac{4}{3} k v_\gamma - 4\dot{\phi} = 0 , \quad (22)$$

$$\dot{v}_\gamma + \frac{1}{4} k \delta_\gamma - \frac{2}{5} k \Theta_2 + k\psi = \dot{\tau}(v_\gamma - v_b) . \quad (23)$$

Here, the derivative of the optical depth $\dot{\tau} = -a n_e \sigma_T$ (and so is negative). The momentum exchange between the photons and baryons due to the drag term in (23) gives rise to a similar term in the Euler equation for the baryons:

$$\dot{v}_b + \mathcal{H} v_b + k\psi = R^{-1} \dot{\tau} (v_b - v_\gamma) , \quad (24)$$

where we have ignored baryon pressure. The ratio of the baryon energy density to the photon enthalpy is $R \equiv 3\rho_b/4\rho_\gamma$ and is proportional to the scale factor a , and $\mathcal{H} \equiv \dot{a}/a$ is the conformal Hubble parameter.

In the tightly-coupled limit $|\dot{\tau}| \ll k^{-1}$ and \mathcal{H}^{-1} . In this limit, we can treat the ratios of the mean-free path to the wavelength and the Hubble time as small perturbative parameters. Equations (23) and (24) then imply that

$v_\gamma = v_b$ to first order in the small quantities $k/|\dot{\tau}|$ and $\mathcal{H}/|\dot{\tau}|$. Comparing the continuity equation for the baryons,

$$\dot{\delta}_b - kv_b - 3\dot{\phi} = 0, \quad (25)$$

with that for the photons, we see that $\dot{\delta}_\gamma = 4\dot{\delta}_b/3$, so the *evolution* of the photon–baryon fluid is adiabatic, preserving the local ratio of the number densities of photons to baryons. Combining (23) and (24) to eliminate the scattering terms, and then using $v_\gamma = v_b$, we find the evolution of the photon velocity to leading order in tight coupling:

$$\dot{v}_\gamma + \frac{R}{1+R}\mathcal{H}v_\gamma + \frac{1}{4(1+R)}k\delta_\gamma + k\psi = 0. \quad (26)$$

The $l > 1$ moments of the photon distribution function arise from the balance between isotropisation by scattering and their generation by photons free streaming over a mean free path; these moments are suppressed by factors $(k/|\dot{\tau}|)^{l-1}$. In particular, during tight coupling $\Theta_2 \approx (20/27)k\dot{\tau}^{-1}v_\gamma$ ignoring polarization. (The factor 20/27 rises to 8/9 if we correct for polarization [38].)

Combining (26) with the photon continuity equation (22) shows that the tightly-coupled dynamics of δ_γ is that of a damped, simple-harmonic oscillator driven by gravity [39]:

$$\ddot{\delta}_\gamma + \frac{\mathcal{H}R}{1+R}\dot{\delta}_\gamma + \frac{1}{3(1+R)}k^2\delta_\gamma = 4\ddot{\phi} + \frac{4\mathcal{H}R}{1+R}\dot{\phi} - \frac{4}{3}k^2\psi. \quad (27)$$

The damping term arises from the redshifting of the baryon momentum in an expanding universe, while photon pressure provides the restoring force which is weakly suppressed by the additional inertia of the baryons. The WKB solutions to the homogeneous equation are

$$\delta_\gamma = (1+R)^{-1/4} \cos kr_s, \quad \text{and} \quad \delta_\gamma = (1+R)^{-1/4} \cos kr_s, \quad (28)$$

where the *sound horizon* $r_s \equiv \int_0^\eta d\eta' / \sqrt{3(1+R)}$. Note also that for static potentials, and ignoring the variation of R with time, the mid-point of the oscillation of δ_γ is shifted to $-4(1+R)\psi$. The dependence of this shift on the baryon density produces a baryon-dependent modulation of the height of the acoustic peak in the temperature anisotropy power spectrum; see Section 3.

The driving term in (27) depends on the evolution of the gravitational potentials. If we ignore anisotropic stress, ϕ and ψ are equal, and their Fourier modes evolve as

$$\begin{aligned} \ddot{\phi} + 3\mathcal{H} \left(1 + \frac{\dot{p}}{\dot{\rho}} \right) \dot{\phi} + \left[2\dot{\mathcal{H}} + \left(1 + 3\frac{\dot{p}}{\dot{\rho}} \right) \mathcal{H}^2 \right] \phi + \frac{\dot{p}}{\dot{\rho}} k^2 \phi \\ = \frac{1}{2} \kappa a^2 \left(\delta p - \frac{\dot{p}}{\dot{\rho}} \delta \rho \right) \end{aligned} \quad (29)$$

in a flat universe, which follows from the perturbed Einstein field equations. Here, ρ and p are the total density and pressure in the background model,

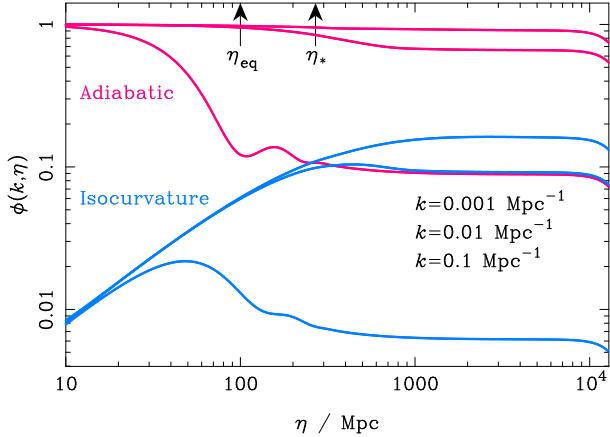


Fig. 2. Evolution of the potential ϕ in adiabatic and CDM-isocurvature models for wavenumbers $k = 0.001, 0.01$ and 0.1 Mpc^{-1} (top to bottom respectively in matter domination). The conformal time at matter–radiation equality η_{eq} and last scattering η_* are marked by arrows.

$\delta\rho$ and δp are the Fourier modes of their perturbations, and $\kappa \equiv 8\pi G$. The source term is gauge-invariant; it vanishes for mixtures of barotropic fluids [$p_i = p_i(\rho_i)$] with $\delta\rho_i/(\rho_i + p_i)$ the same for all components. For *adiabatic* perturbations, this latter condition holds initially and is preserved on super-Hubble scales. It is also preserved in the tightly-coupled photon–baryon fluid as we saw above. For adiabatic perturbations, the potential is constant on scales larger than the sound horizon when p/ρ is constant, but decays during transitions in the equation of state, such as from matter to radiation domination. Above the sound horizon in flat models, it can be shown that the quantity

$$\mathcal{R} \equiv -\dot{\phi} - 2 \frac{\mathcal{H}\dot{\phi} + \mathcal{H}^2}{\kappa a^2(\rho + p)} \quad (30)$$

is conserved even through such transitions. The perturbation to the intrinsic curvature of comoving hypersurfaces (i.e. those perpendicular to the four-velocity of observers who see no momentum density) is given in terms of \mathcal{R} as $4(k^2/a^2)\mathcal{R}$. Using the constancy of \mathcal{R} on large scales, the potential falls by a factor of $9/10$ during the transition from radiation to matter domination. The evolution of the potential is illustrated in Fig. 2 in a flat Λ CDM model with parameters given in Sect. 1. The potential oscillates inside the sound horizon during radiation domination since the photons, which are the dominant component at that time, undergo acoustic oscillations on such scales.

The behaviour of the potentials for *isocurvature* perturbations is quite different on large scales during radiation domination [40], since the source term in (29) is then significant. In isocurvature fluctuations, the initial perturbations in the energy densities of the various components compensate each other

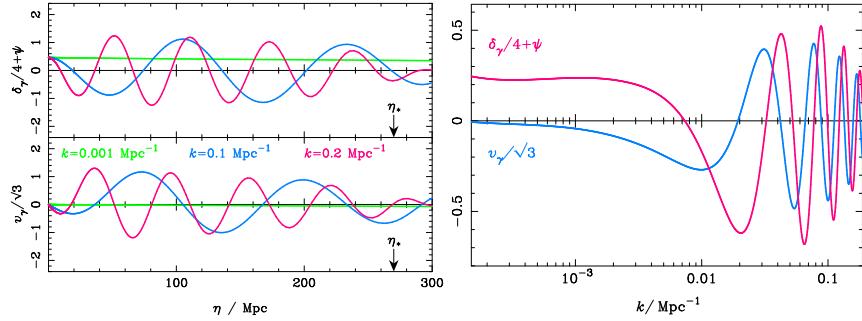


Fig. 3. Evolution of the combination $\delta_\gamma/4 + \psi$ (top left) and the photon velocity v_γ (bottom left) which determine the temperature anisotropies produced at last scattering (denoted by the arrow at η_*). Three modes are shown with wavenumbers $k = 0.001, 0.1$ and 0.2 Mpc^{-1} , and the initial conditions are adiabatic. The fluctuations at the time of last scattering are shown as a function of linear scale in the right-hand plot.

in such a way that the comoving curvature $\mathcal{R} = 0$. Figure 2 shows the evolution of CDM-isocurvature modes, in which there is initially a large fractional perturbation in the dark matter density, with a small compensating fractional perturbation in the radiation. (The full set of possibilities for regular isocurvature modes are discussed in [41].) On large scales in radiation domination the potential grows as a , the scale factor.

Adiabatic Fluctuations

For adiabatic fluctuations, the photons are initially perturbed by $\delta_\gamma(0) = -2\psi(0) = 4\mathcal{R}(0)/3$, i.e. they are over-dense in potential wells, and their velocity vanishes $v_\gamma(0) = 0$. If we consider super-Hubble scales at last scattering, there has been insufficient time for v_γ to grow by gravitational infall and the action of pressure gradients and it remains small. The photon continuity equation (22) then implies that $\delta_\gamma - 4\phi$ remains constant, and the decay of ϕ through the matter–radiation transition leaves $(\delta_\gamma/4 + \psi)(\eta_*) \approx \phi(\eta_*)/3 = -3\mathcal{R}(0)/5$ on large scales ($k < 3 \times 10^{-3} \text{ Mpc}^{-1}$) at last scattering. The combination $\delta_\gamma/4 + \psi = \Theta_0 + \psi$ is the dominant contribution to the large-scale temperature anisotropies produced at last scattering; see (14). The evolution of the photon density and velocity perturbations for adiabatic initial conditions are show in Fig. 3, along with the scale dependence of the fluctuations at last scattering. The plateau in $(\delta_\gamma/4 + \psi)(\eta_*)$ on large scales ensures that a scale-invariant spectrum of curvature perturbations translates into a scale-invariant spectrum of temperature anisotropies, $l(l+1)C_l = \text{constant}$, for small l .

On scales below the sound horizon at last scattering, the photon–baryon fluid has had time to undergo acoustic oscillation. The form of the photon ini-

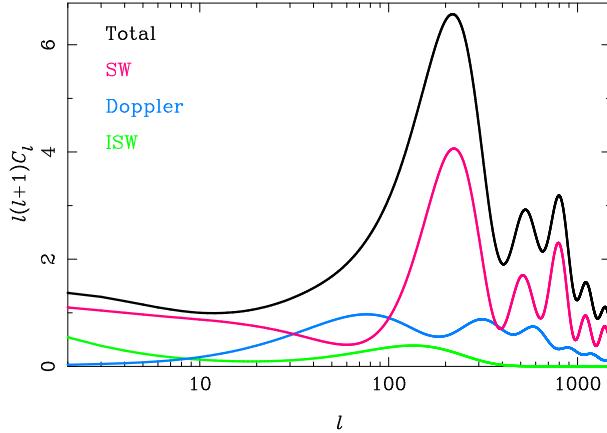


Fig. 4. Contribution of the various terms in (14) to the temperature-anisotropy power spectrum from adiabatic initial conditions. At high l , the contributions are (from top to bottom): total power; $\delta_\gamma/4 + \psi$ (denoted SW for Sachs–Wolfe [3]); Doppler effect from v_b ; and the integrated Sachs–Wolfe effect (ISW) coming from evolution of the potential along the line of sight.

tial condition, and the observation that the driving term in (27) mimics the cosine WKB solution of the homogeneous equation (see Fig. 2), set the oscillation mostly in the $\cos kr_s$ mode. The midpoint of the oscillation is roughly at $\delta_\gamma/4 = -(1 + R)\psi$. This behaviour is illustrated in Fig. 3. Modes with $kr_s(\eta_*) = \pi$ have undergone half an oscillation at last scattering, and are maximally compressed. The large value of $\Theta_0 + \psi$ at this particular scale gives rise to the first acoustic peak in Fig. 1, now measured to be at $l = 220.1 \pm 0.8$ [42]. The subsequent extrema of the acoustic oscillation at $kr_s(\eta_*) = n\pi$ give rise to the further acoustic peaks. The angular spacing of the peaks is almost constant and is set by the sound horizon at last scattering and the angular diameter distance to last scattering. The acoustic part of the anisotropy spectrum thus encodes a wealth of information on the cosmological parameters; see Sect. 3. The photon velocity v_γ oscillates as $\sin kr_s$, so the Doppler term in (14) tends to fill in power between the acoustic peaks. The relative phase of the oscillation of the photon velocity has important implications for the polarization properties of the CMB as discussed in Sect. 4. The contributions of the various terms in (14) to the temperature-anisotropy power spectrum are shown in Fig. 4 for adiabatic perturbations.

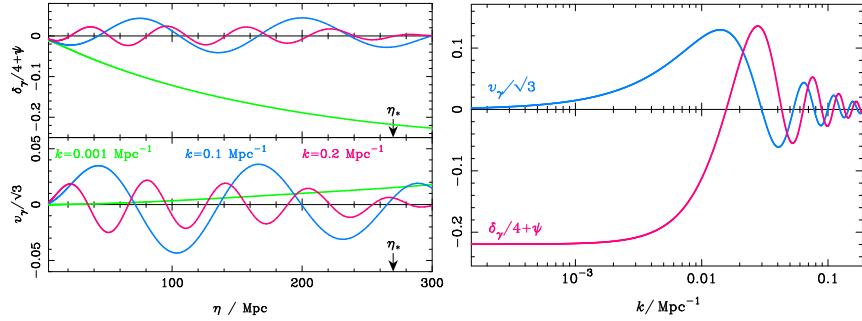


Fig. 5. As Fig. 3 but for CDM-isocurvature initial conditions.

Isocurvature Fluctuations

For the CDM-isocurvature mode¹ the photons are initially unperturbed, as is the geometry: $\delta_\gamma(0) = 0 = \phi(0)$ and $v_\gamma = 0$. On large scales $\delta_\gamma/4 = \phi$ is preserved, so the growth in ϕ during radiation domination is matched by a growth in δ_γ and the photons are under-dense in potential wells. It follows that at last scattering ($\delta_\gamma/4 + \psi)(\eta_*) \approx 2\phi(\eta_*)$ for $k < 3 \times 10^{-3} \text{ Mpc}^{-1}$. Note that the redshift climbing out of a potential well *enhances* the intrinsic temperature fluctuation due to the photon under-density there. The evolution of the photon fluctuations for isocurvature initial conditions are shown in Fig. 5.

The evolution of the potential for isocurvature modes makes the driving term in (27) mimic the sine solution of the homogeneous equation, and so δ_γ follows suit oscillating as $\sin \sim kr_s$ about the equilibrium point $-4(1+R)\psi$. The acoustic peaks are at $kr_s(\eta_*) \sim n\pi/2$, and the photons are under-dense in the potential wells for the odd- n peaks, while over-dense in the even n . The various contributions to the temperature-anisotropy power spectrum for isocurvature initial conditions are shown in Fig. 6. The different peak positions for isocurvature initial conditions allow the CMB to constrain their relative contribution to the total fluctuations. Current constraints are rather dependent on whether one allows for correlations between the adiabatic and isocurvature modes (as are generic in the multi-field inflation models that might have generated the initial conditions), and the extent to which additional cosmological constraints are employed; see [44] for a recent analysis allowing for the most general correlations but a single power-law spectrum.

¹ It is also possible to have the dominant fractional fluctuation in the baryon density rather than the cold dark matter. However, this mode is nearly indistinguishable from the CDM mode since, in the absence of baryon pressure, they differ only by a constant mode in which the radiation and the geometry remain unperturbed, but the CDM and baryon densities have compensating density fluctuations [43].

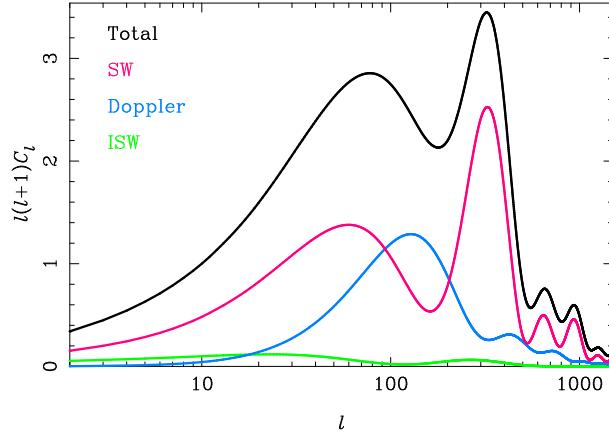


Fig. 6. As Fig. 4 but for CDM-isocurvature initial conditions. The initial spectrum of entropy perturbations is scale-invariant.

Beyond Tight-Coupling

On small scales it is necessary to go beyond tight-coupling of the photon–baryon system since the photon diffusion length can become comparable to the wavelength of the fluctuations. Photons that have had sufficient time to diffuse of the order of a wavelength can leak out of over-densities, thus damping the acoustic oscillations and generating anisotropy [45]. A rough estimate of the comoving scale below which diffusion is important is the square root of the geometric mean of the particle horizon (or conformal age) and the mean-free path of the photons, i.e. $\sqrt{\eta/|\dot{\tau}|}$. Converting this to a comoving wavenumber defines the damping scale

$$k_D^{-2} \sim 0.3(\Omega_m h^2)^{-1/2}(\Omega_b h^2)^{-1}(a/a_*)^{5/2} \text{ Mpc}^2 \quad (31)$$

when the scale factor is a . Here, a_* is the scale factor at last scattering, and the expression is valid well after matter–radiation equality but well before recombination. The effect of diffusion is to damp the photon (and baryon) oscillations exponentially by the time of last scattering on comoving scales smaller than ~ 3 Mpc. The resulting damping effect on the temperature power spectrum has now been measured by several experiments [20, 22, 23].

To describe diffusion damping more quantitatively, we consider scales that were already sub-Hubble during radiation domination. The gravitational potentials will then have been suppressed during their oscillatory phase when the photons (which are undergoing acoustic oscillations themselves) dominated the energy density, and so we can ignore gravitational effects. Furthermore, the dynamical timescale of the acoustic oscillations is then short compared to the expansion time and we can ignore the effects of expansion. In this limit,

the Euler equations for the photons and the baryons can be iterated to give the relative velocity between the photons and baryons to first order in $k/|\dot{\tau}|$:

$$(1 + R^{-1})(v_\gamma - v_b) = \frac{1}{4}k\dot{\tau}^{-1}\delta_\gamma . \quad (32)$$

Using momentum conservation for the total photon–baryon system gives

$$\dot{v}_\gamma + R\dot{v}_b + \frac{1}{4}k\delta_\gamma - \frac{2}{5}k\Theta_2 = 0 , \quad (33)$$

which can be combined with the derivative of (32) to give a new Euler equation for the photons correct to first order in tight coupling:

$$(1 + R)\dot{v}_\gamma \approx -\frac{1}{4}k\delta_\gamma + \frac{R^2}{4(1 + R)}k\dot{\tau}^{-1}\dot{\delta}_\gamma + \frac{16}{45}k^2\dot{\tau}^{-1}v_\gamma . \quad (34)$$

Here, we have used $\Theta_2 \approx 8k\dot{\tau}^{-1}v_\gamma/9$ which includes the correction due to polarization. In the limit of perfect coupling, (34) reduces to (26) on small scales. The continuity equation for the photons, $\dot{\delta}_\gamma = 4kv_\gamma/3$ ($+4\dot{\phi}$), shows that the last two terms on the right of (34) are drag terms, and on differentiating gives

$$\ddot{\delta}_\gamma - \frac{k^2\dot{\tau}^{-1}}{3(1 + R)} \left(\frac{16}{15} + \frac{R^2}{1 + R} \right) \dot{\delta}_\gamma + \frac{k^2}{3(1 + R)}\delta_\gamma = 0 . \quad (35)$$

The WKB solution is

$$\delta_\gamma \propto e^{\pm ikr_s}e^{-k^2/k_D^2} , \quad \text{where} \quad \frac{1}{k_D^2} = \frac{1}{6} \int_0^\eta \frac{|\dot{\tau}^{-1}|}{1 + R} \left(\frac{16}{15} + \frac{R^2}{1 + R} \right) d\eta' \quad (36)$$

is the damping scale.

The finite mean-free path of CMB photons around last scattering has an additional effect on the temperature anisotropies. The visibility function $-\dot{\tau}e^\tau$ has a finite width ~ 80 Mpc and so along a given line of sight photons will be last scattered over this interval. Averaging over scattering events will tend to wash out the anisotropy from wavelengths short compared to the width of the visibility function. This effect is described mathematically by integrating the oscillations in the spherical Bessel functions in (17) against the product of the visibility function and the (damped) perturbations.

Boltzmann codes such as CMBFAST [36] and CAMB [37] use the tight-coupling approximation at early times to avoid the numerical problems associated with integrating the stiff Euler equations in their original forms (23) and (24).

2.5 Other Features of the Temperature-Anisotropy Power Spectrum

We end this section on the fundamentals of the physics of CMB temperature anisotropies by reviewing three additional effects that contribute to the linear anisotropies.

Integrated Sachs–Wolfe Effect

The integrated Sachs–Wolfe (ISW) effect is described by the last term on the right of (14). It is an additional source of anisotropy due to the temporal variation of the gravitational potentials along the line of sight: if a potential well deepens as a CMB photon crosses it then the blueshift due to infall will be smaller than redshift from climbing out of the (now deeper) well. (The combination $\phi + \psi$ has a direct geometric interpretation as the potential for the electric part of the Weyl tensor [46].) The ISW receives contributions from late times as the potentials decay during dark-energy domination, and at early times around last scattering due to the finite time since matter–radiation equality.

The late-time effect contributes mainly on large angular scales since there is little power in the potentials at late times on scales that entered the Hubble radius during radiation domination. The late ISW effect is the only way to probe late-time structure growth (and hence e.g. distinguish between different dark-energy models) with linear CMB anisotropies, but this is hampered by cosmic variance on large angular scales. The late ISW effect produces correlations between the large-scale temperature fluctuations and other tracers of the potential in the local universe, and with the advent of the WMAP data these have now been tentatively detected [47, 48, 49]; see also Sect. 5.

In adiabatic models the early-time ISW effect adds coherently with the contribution $\delta_\gamma/4 + \psi$ to the anisotropies near the first peak, boosting this peak significantly [39]; see Fig. 4. The reason is that the linear scales that contribute here are maximally compressed with $\delta_\gamma/4 + \psi \sim -\psi/2$ which has the same sign as $\dot{\phi}$ for decaying ϕ .

Reionization

Once structure formation had proceeded to produce the first sources of ultra-violet photons, the universe began to reionize. The resulting free electron density could then re-scatter CMB photons, and this tended to isotropise the CMB by averaging the anisotropies from many lines of sight at the scattering event. Approximating the bi-modal visibility function as two delta functions, one at last scattering² η_* and one at reionization η_{re} , if the optical depth through reionization is τ_{re} , the temperature fluctuation at $\mathbf{x} = 0$ at η_0 is

$$\begin{aligned} [\Theta(\hat{\mathbf{n}}) + \psi]_{\eta_0} &\approx (1 - e^{-\tau_{\text{re}}})(\Theta_0 + \psi - \hat{\mathbf{n}} \cdot \mathbf{v}_b)[-\hat{\mathbf{n}}(\eta_0 - \eta_{\text{re}}), \eta_{\text{re}}] \\ &\quad + e^{-\tau_{\text{re}}}(\Theta_0 + \psi - \hat{\mathbf{n}} \cdot \mathbf{v}_b)[-\hat{\mathbf{n}}(\eta_0 - \eta_*), \eta_*]. \end{aligned} \quad (37)$$

Here, we have used (13), neglected the ISW effect, and approximated the scattering as isotropic. The first term on the right describes the effect of blending

² We continue to refer to the last scattering event around recombination as last scattering, even in the presence of re-scattering at reionization.

the anisotropies from different lines of sight (to give Θ_0) and the generation of new anisotropies by re-scattering off moving electrons at reionization; the second term is simply the temperature anisotropy that would be observed with no reionization, weighted by the fraction of photons that do not re-scatter. Since $\Theta_0 + \psi$ at the re-scattering event is the average of $\Theta_0 + \psi - \hat{\mathbf{n}}' \cdot \mathbf{v}_b$ on the electron's last scattering surface, on large scales $k(\eta_{\text{re}} - \eta_*) \ll 1$ it reduces to $\Theta_0 + \psi$ at $[-\hat{\mathbf{n}}(\eta_0 - \eta_*), \eta_*]$, while on small scales it vanishes. It follows that for scales that are super-horizon at reionization, the observed temperature anisotropy becomes

$$\Theta(\hat{\mathbf{n}}) \rightarrow \Theta(\hat{\mathbf{n}}) - (1 - e^{-\tau_{\text{re}}}) \hat{\mathbf{n}} \cdot \Delta \mathbf{v}_b , \quad (38)$$

where $\Delta \mathbf{v}_b$ is the difference between the electron velocity at the reionization event and the preceding last scattering event on the line of sight. On such scales the Doppler terms do not contribute significantly and the temperature anisotropy is unchanged. For scales that are sub-horizon at reionization,

$$\Theta(\hat{\mathbf{n}}) \rightarrow e^{-\tau_{\text{re}}} \Theta(\hat{\mathbf{n}}) - (1 - e^{-\tau_{\text{re}}}) \hat{\mathbf{n}} \cdot \mathbf{v}_b , \quad (39)$$

where the Doppler term is evaluated at reionization. In practice, the visibility function is not perfectly sharp at reionization and the integral through the finite re-scattering distance tends to wash out the Doppler term since only plane waves with their wavevectors near the line of sight contribute significantly to $\hat{\mathbf{n}} \cdot \mathbf{v}_b$. Figure 7 shows the resulting effect $C_l \rightarrow e^{-2\tau_{\text{re}}} C_l$ on the anisotropy power spectrum on small scales. Recent results from WMAP [19] suggest an optical depth through reionization $\tau_{\text{re}} \sim 0.17$. Such early reionization cannot have been an abrupt process since the implied redshift $z_{\text{re}} \sim 15$ is at odds with the detection of traces of smoothly-distributed neutral hydrogen at $z \sim 6$ via Gunn-Peterson troughs in the spectra of high-redshift quasars [50, 51].

Tensor Modes

Tensor modes, describing gravitational waves, represent the transverse trace-free perturbations to the spatial metric:

$$ds^2 = a^2(\eta)[d\eta^2 - (\delta_{ij} + h_{ij})dx^i dx^j] , \quad (40)$$

with $h_i^i = 0$ and $\partial_i h_j^i = 0$. A convenient parameterisation of the photon four-momentum in this case is

$$p^\mu = \frac{\epsilon}{a^2} \left[1, \mathbf{e}^i - \frac{1}{2} h_j^i \mathbf{e}^j \right] , \quad (41)$$

where $\mathbf{e}^2 = 1$ and ϵ is a times the energy of the photon as seen by an observer at constant x . The components of \mathbf{e} are the projections of the photon direction for this observer on an orthonormal spatial triad of vectors $a^{-1}(\partial_i - h_i^j \partial_j / 2)$.

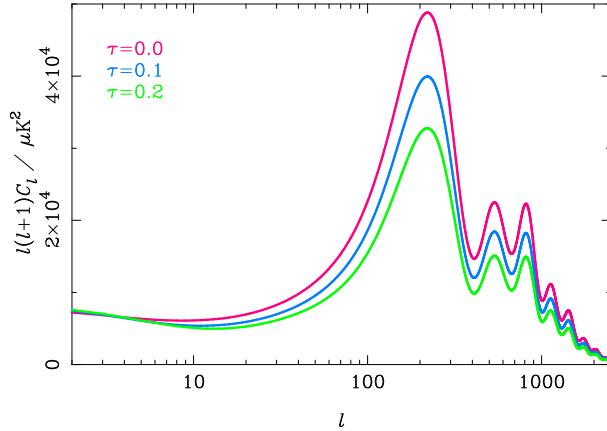


Fig. 7. Effect of reionization on the temperature-anisotropy power spectrum. The spectra are (from top to bottom) for no reionization, $\tau_{\text{re}} = 0.1$ and 0.2 .

In the background $\mathbf{e} = d\mathbf{x}/d\eta$ and is constant. The evolution of the comoving energy ϵ in the perturbed universe is

$$\frac{1}{\epsilon} \frac{d\epsilon}{d\eta} + \frac{1}{2} \dot{h}_{ij} e^i e^j = 0 , \quad (42)$$

and so the Boltzmann equation for $\Theta(\eta, \mathbf{x}, \mathbf{e})$ is

$$\begin{aligned} \frac{\partial \Theta}{\partial \eta} + \mathbf{e} \cdot \nabla \Theta &= -a n_e \sigma_T \Theta + \frac{3}{16\pi} a n_e \sigma_T \int \Theta(e') [(1 + (\mathbf{e} \cdot \mathbf{e}')^2] d\mathbf{e}' \\ &\quad - \frac{1}{2} \dot{h}_{ij} e^i e^j . \end{aligned} \quad (43)$$

Neglecting the anisotropic nature of Thomson scattering, the solution of this equation is an integral along the unperturbed line of sight:

$$\Theta(\hat{\mathbf{n}}) = -\frac{1}{2} \int_0^{\eta_0} e^{-\tau} \dot{h}_{ij} \hat{n}^i \hat{n}^j d\eta . \quad (44)$$

The time derivative \dot{h}_{ij} is the shear induced by the gravitational waves. This quadrupole perturbation to the expansion produces an anisotropic redshifting of the CMB photons and an associated temperature anisotropy.

Figure 8 compares the power spectrum due to gravitational waves with that from scalar perturbations for a tensor-to-scalar ratio $r = 1$ corresponding to an energy scale of inflation 3.3×10^{16} GeV. The constraints on gravitational waves from temperature anisotropies are not very constraining since their effect is limited to large angular scales where cosmic variance from the dominant scalar perturbations is large. Gravitational waves damp as they oscillate inside the horizon, so the only significant anisotropies are from wavelengths that are

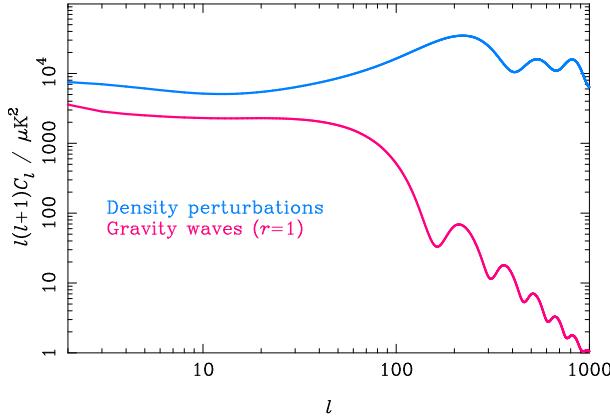


Fig. 8. The temperature-anisotropy power spectrum from scalar perturbations (density perturbations; top) and tensor perturbations (gravity waves; bottom) for a tensor-to-scalar ratio $r = 1$.

super-horizon at last scattering, corresponding to $l \sim 60$. The current 95-per cent upper limit on the tensor-to-scalar ratio is 0.68 [20]. Fortunately, CMB polarization provides an alternative route to detecting the effect of gravitational waves on the CMB which is not limited by cosmic variance [52, 53]; see also Sect. 4.

3 Cosmological Parameters and the CMB

The simple, linear physics of CMB temperature anisotropies, reviewed in the previous section, means that the CMB depends sensitively on many of the key cosmological parameters. For this reason, CMB observations over the past decade have been a significant driving force in the quest for precision determinations of the cosmological parameters. It is not our intention here to give a detailed description of the constraints that have emerged from such analyses, e.g. [10], but rather to provide a brief description of how the key parameters affect the temperature-anisotropy power spectrum. More details can be found in the seminal papers on this subject, e.g. [39, 40, 54] and references therein.

3.1 Matter and Baryons

The curvature of the universe and the properties of the dark energy are largely irrelevant for the pre-recombination physics of the acoustic oscillations. Their main contribution is felt geometrically through the angular diameter distance to last scattering, D_A , which controls the projection of linear scales there to

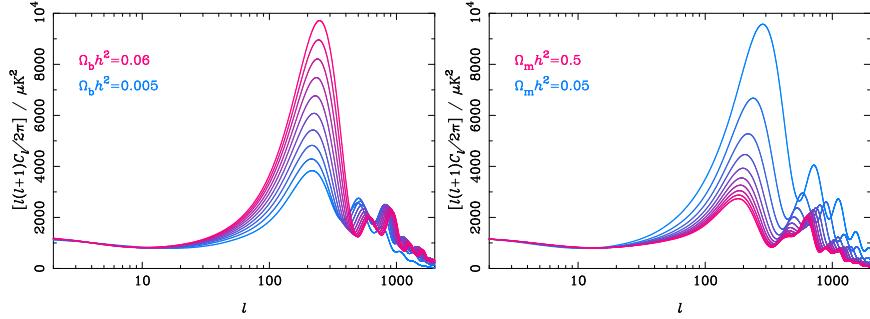


Fig. 9. Dependence of the temperature-anisotropy power spectrum on the physical density in baryons (left) and all non-relativistic matter (right). From top to bottom at the first peak, the baryon densities vary linearly in the range $\Omega_b h^2 = 0.06\text{--}0.005$ (left) and the matter densities in $\Omega_m h^2 = 0.05\text{--}0.5$ (right). The initial conditions are adiabatic.

angular scales on the sky today. In contrast, those parameters that determine the energy content of the universe before recombination, such as the physical densities in (non-relativistic) matter $\Omega_m h^2$, and radiation $\Omega_r h^2$ (determined by the CMB temperature and the physics of neutrinos), play an important role in acoustic physics by determining the expansion rate and hence the behaviour of the perturbations. In addition, the physical density in baryons, $\Omega_b h^2$, affects the acoustic oscillations through baryon inertia and the dependence of the photon mean-free path on the electron density. The effect of variations in the physical densities of the matter and baryon densities on the anisotropy power spectrum is illustrated in Fig. 9 for adiabatic initial conditions.

The linear scales at last scattering that have reached extrema of their oscillation are determined by the initial conditions (i.e. adiabatic or isocurvature) and the sound horizon $r_s(\eta_*)$. Increasing the baryon density holding the total matter density fixed reduces the sound speed while preserving the expansion rate (and moves last scattering to slightly earlier times). The effect is to reduce the sound horizon at last scattering and so the wavelength of those modes that are at extrema of their oscillation, and hence push the acoustic peaks to smaller scales. This effect could be confused with a change in the angular diameter distance D_A , but fortunately baryons have another distinguishing effect. Their inertia shifts the zero point of the acoustic oscillations to $\sim -(1+R)\psi$, and enhances the amplitude of the oscillations. In adiabatic models for modes that enter the sound horizon in matter domination, $\delta_\gamma/4$ starts out at $-2\psi/3$, and so the amplitude of the oscillation is $-\psi(1+3R)/3$. The combination of these two effects is to enhance the amplitude of $\Theta_0 + \psi$ at maximal compression by a factor of $1+6R$ over that at minimal compression. The effect on the power spectrum is to enhance the amplitude of the 1st, 3rd etc. peaks for adiabatic initial conditions, and the 2nd, 4th etc. for isocurvature. Current CMB data gives $\Omega_b h^2 = 0.023 \pm 0.001$ for power-law

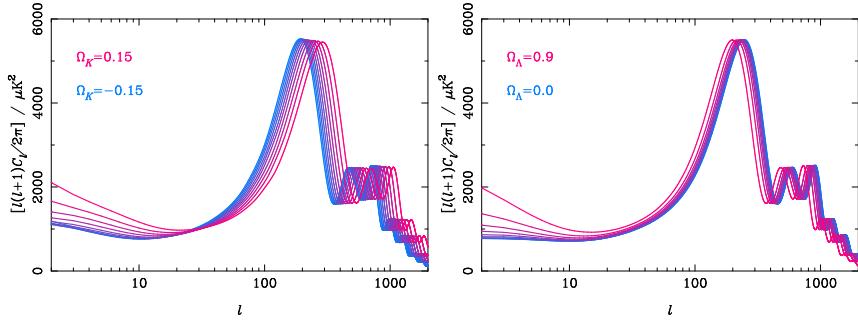


Fig. 10. Dependence of the temperature-anisotropy power spectrum on the curvature Ω_K (left) and cosmological constant Ω_Λ (right) in adiabatic models. In both cases, the physical densities in baryons and matter were held constant, thus preserving the conditions on the last scattering surface. The curvature varies (left to right) in the range -0.15–0.15 and the cosmological constant in the range 0.9–0.0.

Λ CDM models [12], beautifully consistent with determinations from big bang nucleosynthesis. Other effects of baryons are felt in the damping tail of the power spectrum since increasing the baryon density tends to inhibit diffusion giving less damping at a given scale.

The effect of increasing the physical matter density $\Omega_m h^2$ at fixed $\Omega_b h^2$ is also two-fold (see Fig. 9): (i) a shift of the peak positions to larger scales due to the increase in D_A ; and (ii) a scale-dependent reduction in peak height in adiabatic models. Adiabatic modes that enter the sound horizon during radiation domination see the potentials decay as the photon density rises to reach maximal compression. This decay tends to drive the oscillation, increasing the oscillation amplitude. Raising $\Omega_m h^2$ brings matter–radiation equality to earlier times, and reduces the efficiency of the gravitational driving effect for the low-order peaks. Current CMB data gives $\Omega_m h^2 = 0.13 \pm 0.01$ for adiabatic, power-law Λ CDM models [12].

3.2 Curvature, Dark Energy and Degeneracies

The main effect of curvature and dark energy on the linear CMB anisotropies is through the angular diameter distance and the late-time integrated Sachs–Wolfe effect; see Fig. 10 for the case of adiabatic fluctuations in cosmological-constant models. The ISW contribution is limited to large scales where cosmic variance severely limits the precision of power spectrum estimates. There is an additional small effect due to quantisation of the allowed spatial modes in closed models (e.g. [55]), but this is also confined to large scales (i.e. near the angular projection of the curvature scale). Most of the information that the CMB encodes on curvature and dark energy is thus locked in the angular diameter distance to last scattering, D_A .

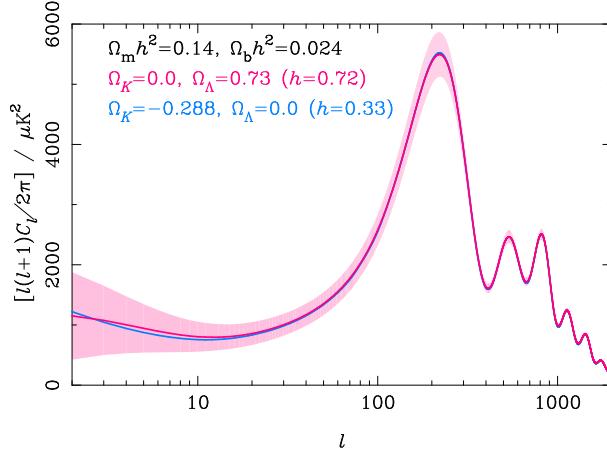


Fig. 11. The geometric degeneracy. A scale-invariant adiabatic Λ CDM model with $\Omega_b h^2 = 0.024$, $\Omega_m h^2 = 0.14$ and $\Omega_\Lambda = 0.73$ and $\Omega_K = 0$ (close to the WMAP best-fit values [12]) produces an almost identical spectrum to a closed model $\Omega_K = -0.288$ with vanishing cosmological constant. However, the Hubble constants are very different – $h = 0.72$ in the flat model and 0.33 in the closed model – and so the latter is easily ruled out by external constraints. The shaded region shows the 1σ cosmic variance errors $\Delta C_l / C_l = \sqrt{2/(2l + 1)}$ on the power spectrum.

With the physical densities $\Omega_b h^2$ and $\Omega_m h^2$ fixed by the acoustic part of the anisotropy spectrum, D_A can be considered a function of Ω_K and the history of the energy density of the dark energy (often modelled through its current density and a constant equation of state). In cosmological constant models D_A is particularly sensitive to the curvature: the 95-per cent interval from WMAP alone (with the weak prior $H_0 > 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$) is $-0.08 < \Omega_K < 0.02$, so the universe is close to being spatially flat. The fact that the impact of curvature and the properties of the dark energy on the CMB is mainly through a single number D_A leads to a geometrical degeneracy in parameter estimation [56], as illustrated in Fig. 11. Fortunately, this is easily broken by including other, complementary cosmological datasets. The constraint on curvature from WMAP improves considerably when supernovae measurements [57, 58], or the measurement of H_0 from the Hubble Space Telescope Key Project [59] are included. Other examples of near-perfect degeneracies for the temperature anisotropies include the addition of gravity waves and a reduction in the amplitude of the initial fluctuations mimicing the effect of reionization. This degeneracy is broken very effectively by the polarization of the CMB.

4 CMB Polarization

The growth in the mean-free path of the CMB photons during recombination allowed anisotropies to start to develop. Subsequent scattering of the radiation generated (partial) linear polarization from the quadrupole anisotropy. This linear polarization signal is expected to have an r.m.s. $\sim 5 \mu\text{K}$, and, for scalar perturbations, to peak around multipoles $l \sim 1000$ corresponding to the angle subtended by the mean-free path around last scattering. The detection of CMB polarization was first announced in 2002 by the Degree Angular Scale Interferometer (DASI) team [13]; WMAP has also detected the polarization indirectly through its correlation with the temperature anisotropies [19]. A direct measurement of the polarization power from two-years of WMAP data is expected shortly. Polarization is only generated by scattering, and so is a sensitive probe of conditions at recombination. In addition, large-angle polarization was generated by subsequent re-scattering as the universe reionized, providing a unique probe of the ionization history at high redshift.

4.1 Polarization Observables

Polarization is conveniently described in terms of Stokes parameters I , Q , U and V , where I is the total intensity discussed at length in the previous section. The parameter V describes circular polarization and is expected to be zero for the CMB since it is not generated by Thomson scattering. The remaining parameters Q and U describe linear polarization. They are the components of the trace-free, (zero-lag) correlation tensor of the electric field in the radiation, so that for a quasi-monochromatic plane wave propagating along the z direction

$$\begin{pmatrix} \langle E_x^2 - E_y^2 \rangle & 2\langle E_x E_y \rangle \\ 2\langle E_x E_y \rangle & -\langle E_x^2 - E_y^2 \rangle \end{pmatrix} = \frac{1}{2} \begin{pmatrix} Q & U \\ U & -Q \end{pmatrix}, \quad (45)$$

where the angle brackets represent an average on timescales long compared to the period of the wave. For diffuse radiation we define the polarization brightness tensor $\mathcal{P}_{ab}(\hat{\mathbf{n}})$ to have components given by (45) for plane waves within a bundle around the line of sight $\hat{\mathbf{n}}$ and around the specified frequency. The polarization tensor is transverse to the line of sight, and, since it inherits its frequency dependence from the the quadrupole of the total intensity, has a spectrum given by the derivative of the Planck function (see equation 11).

The polarization tensor can be decomposed uniquely on the sphere into an electric (or gradient) part and a magnetic (or curl) part [52, 53]:

$$\mathcal{P}_{ab} = \nabla_{(a} \nabla_{b)} P_E - \epsilon_{ab}^c \nabla_c P_B, \quad (46)$$

where angle brackets denote the symmetric, trace-free part, ∇_a is the covariant derivative on the sphere, and ϵ_{ab}^c is the alternating tensor. The divergence $\nabla^a \mathcal{P}_{ab}$ is a pure gradient if the magnetic part $P_B = 0$, and a curl if the electric

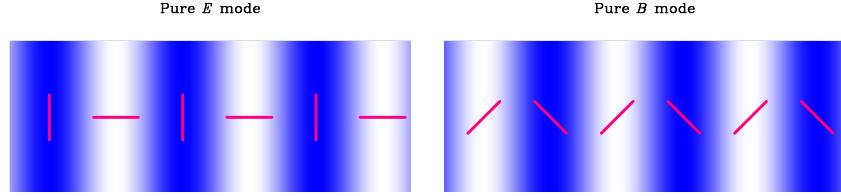


Fig. 12. Polarization patterns for a pure-electric mode (left) and pure-magnetic mode (right) on a small patch of the sky for potentials that are locally Fourier modes. The shading denotes the amplitude of the potential. For the electric pattern the polarization is aligned with or perpendicular to the Fourier wavevector depending on the sign of the potential; for the magnetic pattern the polarization is at 45 degrees.

part $P_E = 0$. The potential P_E is a scalar under parity, but P_B is a pseudo-scalar. For a given potential P , the electric and magnetic patterns it generates (i.e. with $P_E = P$ and $P_B = P$ respectively) are related by locally rotating the polarization directions by 45 degrees. The polarization orientations on a small patch of the sky for potentials that are locally Fourier modes are shown in Fig. 12. The potentials can be expanded in spherical harmonics (only the $l \geq 2$ multipoles contribute to \mathcal{P}_{ab}) as

$$P_E(\hat{\mathbf{n}}) = \sum_{lm} \sqrt{\frac{(l-2)!}{(l+2)!}} E_{lm} Y_{lm}(\hat{\mathbf{n}}), \quad P_B(\hat{\mathbf{n}}) = \sum_{lm} \sqrt{\frac{(l-2)!}{(l+2)!}} B_{lm} Y_{lm}(\hat{\mathbf{n}}). \quad (47)$$

(The normalisation is conventional.) Under parity $E_{lm} \rightarrow (-1)^l E_{lm}$ but $B_{lm} \rightarrow -(-1)^l B_{lm}$. Assuming rotational and parity invariance, B is not correlated with E or the temperature anisotropies T , leaving four non-vanishing power spectra: C_l^T , C_l^E , C_l^B and the cross-correlation C_l^{TE} , where e.g. $\langle E_{lm} T_{lm}^* \rangle = C_l^{TE}$.

4.2 Physics of CMB Polarization

For scalar perturbations, the quadrupole of the temperature anisotropies at leading order in tight coupling is $\Theta_2 \sim k\dot{\tau}^{-1}v_\gamma$. Scattering of this quadrupole into the direction $-\hat{\mathbf{n}}$ generates linear polarization parallel or perpendicular to the projection of the wavevector \mathbf{k} onto the sky, i.e. $\mathcal{P}_{ij} \sim \Theta_2 [\hat{k}_{(i}\hat{k}_{j)}]^{\text{TT}}$, where TT denotes the transverse (to $\hat{\mathbf{n}}$), trace-free part. In a flat universe the polarization tensor is conserved in the absence of scattering; for non-flat models this is still true if the components are defined on an appropriately-propagated basis (e.g. [26]). For a single plane wave perturbation, the polarization on the sky is thus purely electric (see Fig. 12). For tensor perturbations, the polarization $\mathcal{P}_{ij} \sim \dot{\tau}^{-1} [\dot{h}_{ij}]^{\text{TT}}$ since the tightly-coupled quadrupole is proportional to the shear \dot{h}_{ij} . The gravitational wave defines additional directions on the sky when its shear is projected, and the polarization pattern is not purely

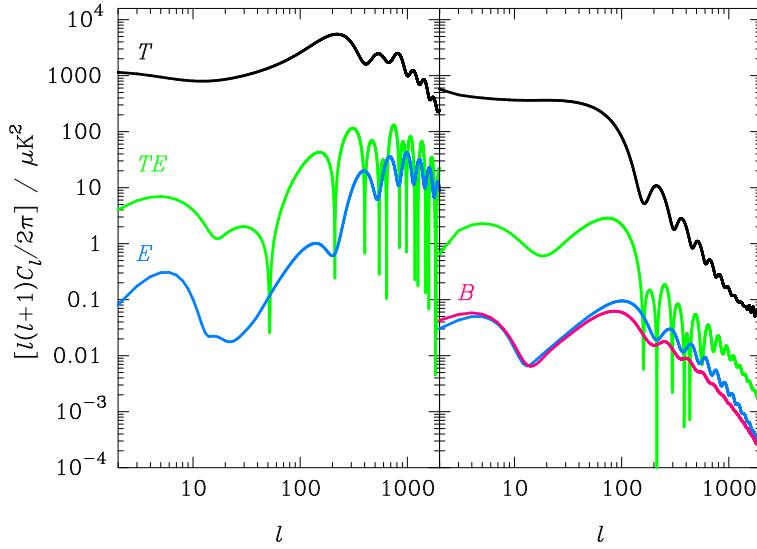


Fig. 13. Power spectra produced by adiabatic scalar perturbations (left) and tensor perturbations (right) for a tensor-to-scalar ratio $r = 1$. On large scales the spectra from scalar perturbations are (from top to bottom) C_l^T , C_l^{TE} and C_l^E . For tensor perturbations, they are C_l^T , C_l^{TE} , C_l^B and C_l^E .

electric. Thus density perturbations do not produce magnetic polarization in linear perturbation theory, while gravitational waves produce both electric and magnetic [52, 53].

The polarization power spectra produced by scalar and tensor perturbations are compared in Fig. 13. The scalar C_l^E spectrum peaks around $l \sim 1000$ since this corresponds to the projection of linear scales at last scattering for which diffusion generates a radiation quadrupole most efficiently. The polarization probes the photon bulk velocity at last scattering, and so C_l^E peaks at the troughs of C_l^T , while C_l^{TE} is zero at the peaks and troughs, and has its extrema in between. For adiabatic perturbations, the large-scale cross-correlation changes sign at $l \sim 50$, and, with the conventions adopted here³ is positive between $l = 50$ and the first acoustic peak in C_l^T . Isocurvature modes produce a negative correlation from $l = 2$ to the first acoustic trough.

Tensor modes produce similar power in electric and magnetic polarization. As gravitational waves damp inside the horizon, the polarization peaks just shortward of the horizon size at last scattering $l \sim 100$ despite these large

³ The sign of E_{lm} for a given polarization field depends on the choice of conventions for the Stokes parameters and their decomposition into electric and magnetic multipoles. We follow [60], which produces the same sign of C_l^{TE} as [25], but note that the Boltzmann codes CMBFAST [36] and CAMB [37] have the opposite sign.

scales being geometrically less efficient at transferring power to the quadrupole during a mean-free time than smaller scales.

For both scalar and tensor perturbations, the polarization would be small on large scales were it not for reionization, since a significant quadrupole is only generated at last scattering when the mean-free path approaches the wavelength of the fluctuations. However, reionization does produce significant large-angle polarization [61] (see Fig. 13). The temperature quadrupole at last scattering peaks on linear scales with $k(\eta_{\text{re}} - \eta_*) \sim 2$, which then re-projects onto angular scales $l \sim 2(\eta_0 - \eta_{\text{re}})/(\eta_{\text{re}} - \eta_*)$. The position of the reionization feature is thus controlled by the epoch of reionization, and the height by the fraction of photons that scatter there i.e. τ_{re} . The measurement of τ_{re} with large-angle polarization allows an accurate determination of the amplitude of scalar fluctuations from the temperature-anisotropy power spectrum. In addition, the fine details of the large-angle polarization power can in principle distinguish different ionization histories with the same optical depth, although this is hampered by the large cosmic variance at low l [62].

5 Highlights of Recent Results

In this section we briefly review some of the highlights from recent observations of the CMB temperature and polarization anisotropies. Analysis of the former have entered a new phase with the release of the first year data from the WMAP satellite [11]; a further three years worth of data are expected from this mission. Detections of CMB polarization are still in their infancy, but here too we can expect significant progress from a number of experiments in the short term.

5.1 Detection of CMB Polarization

The first detection of polarization of the CMB was announced in September 2002 [13]. The measurements were made with DASI, a compact interferometric array operating at 30 GHz, deployed at the South Pole. The DASI team constrained the amplitude of the E and B -mode spectra with assumed spectral shapes derived from a concordant Λ CDM model. They obtained a $\sim 5\sigma$ detection of a non-zero amplitude for E with a central value perfectly consistent with that expected from the amplitude of the temperature anisotropies. DASI also detected the temperature–polarization cross-correlation at 95-per cent significance, but no evidence for B -mode polarization was found. The DASI results of a maximum-likelihood band-power estimation of the E and TE power spectra are given in Fig. 14.

Measurements of C_l^{TE} were also provided in the first-year data release from WMAP, although polarization data itself was not released. These results are also shown in Fig. 14. The existence of a cross-correlation between temperature and polarization on degree angular scales provides evidence for

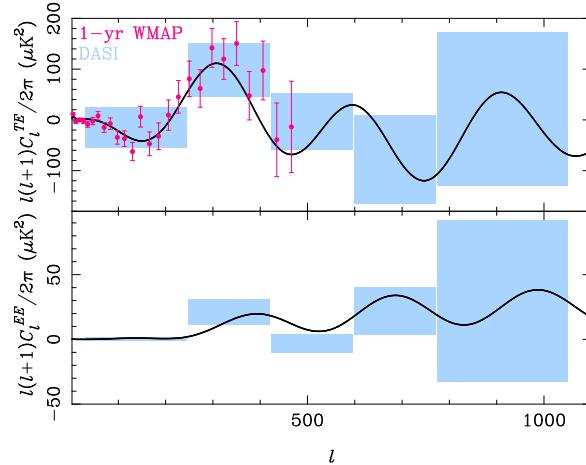


Fig. 14. Current measurements (as of February 2004) of C_l^{TE} (top) and C_l^E (bottom). The points with $1-\sigma$ errors are from the first one-year data release from WMAP [63]. The error boxes are the flat band-power results from DASI [13] centred on the maximum-likelihood band power and spanning the 68-per cent interval. The solid lines are the predicted power from the best-fit model to all the WMAP data.

the existence of super-horizon fluctuations on the last scattering surface at recombination. This is more direct evidence for such fluctuations than from the large-scale temperature anisotropies alone, since the latter could have been generated gravitationally all along the line of sight. The sign of the cross-correlation and the phase of its acoustic peaks relative to those in the temperature-anisotropy spectrum is further strong evidence for adiabatic fluctuations. The one surprise in the WMAP measurement of C_l^{TE} is the behaviour on large scales. A significant excess correlation over that expected if polarization were only generated at recombination is present on large scales ($l < 20$). The implication is that reionization occurred early, $11 < z_{\text{re}} < 30$, giving a significant optical depth for re-scattering: $\tau_{\text{re}} = 0.17 \pm 0.04$ at 68-per cent confidence. As mentioned in Sect. 2.5, reionization at this epoch is earlier than that expected from observations of quasar absorption spectra and suggests a complex ionization history.

5.2 Implications of Recent Results for Inflation

The generic predictions from simple inflation models are that: (i) the universe should be (very nearly) spatially flat; (ii) there should be a nearly scale-invariant spectrum of Gaussian, adiabatic density perturbations giving apparently-super-horizon fluctuations on the last scattering surface; and (iii) there should be a stochastic background of gravitational waves with a nearly scale-invariant (but necessarily not blue) spectrum. The amplitude of the lat-

ter is a direct measure of the Hubble rate during inflation, and hence, in slow-roll models, the energy scale of inflation.

As discussed in Sect. 3.2, the measured positions of the acoustic peaks constrains the universe to be close to flat. The constraint improves further with the inclusion of other cosmological data. There is no evidence for isocurvature modes in the CMB, although the current constraints are rather weak if general, correlated modes are allowed in the analysis [44]. Several of the cosmological parameters for the isocurvature models most favoured by CMB data are violently at odds with other probes, most notably the baryon density which is pushed well above the value inferred from the abundances of the light-elements. There is also no evidence for primordial non-Gaussianity in the CMB (see e.g. [64])⁴.

Within flat Λ CDM models with a power-law spectrum of curvature fluctuations, the spectral index is constrained by the CMB to be close to scale invariant [12], although the inclusion of the latest data from small-scale experiments, such as CBI [70] and VSA [71], tends to pull the best fit from WMAP towards redder power-law spectra: e.g. $n_s = 0.97^{+0.06}_{-0.03}$ at 68-per cent confidence combining WMAP and VSA [71]. Slow-roll inflation predicts that the fluctuation spectrum should be close to a power law, with a run in the spectral index that is second order in slow roll: $dn_s/d\ln k \sim (n_s - 1)^2$. The WMAP team reported weak evidence for a running spectral index by including small-scale data from galaxy redshift surveys and the Lyman- α forest, but modelling uncertainties in the latter have led many to question the reliability of this result (e.g. [72]). New data from CBI and VSA now provide independent evidence for running in flat Λ CDM models at the 2σ level from the CMB alone. This reflects the tension between the spectral index favoured by the low- l CMB data (which is anomalously low for $l < 10$, favouring bluer spectra) and the high- l data from the interferometers. The evidence for running is weakened considerably with the inclusion of external priors from large-scale structure data. The best-fit values for the run in n_s obtained with the CMB alone are uncomfortably large for slow-roll inflation models, and give low power on small scales that is difficult to reconcile with the early reionization implied by the WMAP polarization data. However, a recent analysis [73] argues that the evidence for running depends crucially on the techniques employed to estimate the low- l power from WMAP data, and that the running is strongly suppressed if exact likelihood techniques are adopted. A definitive answer on whether departures from power-law spectra are significant must probably await further data on both large and small scales.

The final prediction of slow-roll inflation – the generation of nearly scale-invariant background of gravitational waves – is yet to be verified. The current

⁴ The WMAP data does appear to harbour some statistically-significant departures from rotational invariance [65, 66, 67, 68, 69]. The origin of these effects, i.e. primordial or systematic due to instrument effects or imperfect foreground subtraction, is as yet unclear.

limits on the tensor-to-scalar ratio are only weak: [71] report $r < 0.68$ at 95-per cent confidence from all CMB data in general, non-flat, adiabatic Λ CDM models. Despite this, observations are beginning to place interesting constraints on specific models of inflation in the r - n_s plane [74, 75]. Already, large-field models with power-law potentials steeper than $V \propto \phi^6$ are ruled out due to their red scalar spectra and comparatively large tensor-to-scalar ratio. Future programmes targeting B -mode polarization may ultimately be able to detect gravitational waves down to an inflationary energy scale of a few $\times 10^{15}$ GeV. Such observations will sharpen constraints in the r - n_s plane considerably, and should allow fine selection amongst the many proposed models of inflation.

5.3 Detection of Late-Time Integrated Sachs-Wolfe Effect

The late-time ISW effect arises from the decay of the gravitational potentials once the universe becomes dark-energy dominated, and so should produce large-angle (positive) correlations between the CMB temperature anisotropies and other tracers of the potential in the local universe. With the advent of the WMAP data, a number of groups have reported the detection of such a correlation. In [47], WMAP data was cross-correlated with data on the hard X-ray background (which is dominated by emission from active galaxies) from the HEAO-1 satellite, and the number density of radio sources from the NVSS catalogue. In each case a positive correlation was detected at significance 3σ and 2.5σ respectively. The correlation with NVSS has also been carried out independently by the WMAP team [48], who also note that the observed positive correlation can be used to rule out the closed, $\Lambda = 0$ model model that is a good fit to the CMB data in isolation (see Fig. 11). Several groups have now also detected the cross-correlation on large scales between the CMB and optical galaxy surveys, e.g. [49].

6 Conclusion

The linear anisotropies of the cosmic microwave background have been studied theoretically for over three decades. The physics, which is now well understood, employs linearised radiative transfer, general relativity, and hydrodynamics to describe the propagation of CMB photons and the evolution of the fluid constituents in a perturbed Friedmann-Robertson-Walker universe. A number of bold predictions have emerged from this theoretical activity, most notably the existence of acoustic peaks in the anisotropy power spectrum due to oscillations in the photon-baryon plasma prior to recombination. Observers have risen to the challenge of verifying these predictions, and their detection is proceeding at a staggering rate. The large-scale Sachs-Wolfe effect, acoustic peak structure, damping tail, late-time integrated Sachs-Wolfe effect, polarization and reionization signature have all been detected, and the first three have been measured in considerable detail. Already, the size and scale of these

effects is allowing cosmological models to be constrained with unprecedented precision. The results are beautifully consistent with almost-scale-invariant adiabatic initial conditions evolving passively in a spatially flat, Λ CDM universe.

Much work still remains to be done to exploit fully the information contained in the CMB anisotropies. The Planck satellite, due for launch in 2007, should provide definitive mapping of the linear CMB anisotropies, and a cosmic-variance limited measurement of the power spectrum up to multipoles $l \sim 2000$. This dataset will be invaluable in assessing many of the issues hinted at in the first-year release of WMAP data, such as the apparent lack of power on large scales and possible violations of rotational (statistical) invariance. Prior to Planck, a number of ground-based programmes should shed further light on the issue of whether departures from a power-law primordial spectrum are required on cosmological scales, and the implications of this for slow-roll inflation. In addition, these small-scale observations will start to explore the rich science of secondary anisotropies, due to e.g. scattering in hot clusters [76] or bulk flows modulated by variations in the electron density in the reionized universe [77, 78], and the weak lensing effect of large-scale structure [79].

Detections of CMB polarization are in their infancy, but we can expect rapid progress on this front too. Accurate measurements of the power spectra of *E*-mode polarization, and its correlation with the temperature anisotropies, can be expected from a number of ground and balloon-borne experiments, as well as from Planck. The ultimate goal for CMB polarimetry is to detect the *B*-mode signal predicted from gravitational waves. This would give a direct measure of the energy scale of inflation, and, when combined with measurements of the spectrum density perturbations, place tight constraints on the dynamics of inflation. Plans are already being made for a new generation of polarimeters with the large numbers of detectors and exquisite control of instrument systematics needed to detect the gravity-wave signal if the energy scale of inflation is around 10^{16} GeV. Ultimately, confusion due to imperfect subtraction of astrophysical foregrounds and the effects of weak lensing on the polarization limit will limit the energy scales that we can probe with CMB polarization; see [80] and references therein.

Acknowledgments

AC acknowledges a Royal Society University Research Fellowship.

References

1. A.A. Penzias, R.W. Wilson: *Astrophys. J.* **142**, 419 (1965)
2. R.H. Dicke et al: *Astrophys. J.* **142**, 414 (1965)
3. R.K. Sachs, A.M. Wolfe: *Astrophys. J.* **147**, 73 (1967)

4. G.F. Smoot et al: *Astrophys. J. Lett.* **396**, 1 (1992)
5. P.J.E. Peebles: *Astrophys. J. Lett.* **263**, 1 (1982)
6. J.R. Bond, G. Efstathiou: *Astrophys. J. Lett.* **285**, 45 (1984)
7. P.J.E. Peebles J.T. Yu: *Astrophys. J.* **162**, 815 (1970)
8. J.C. Mather et al: *Astrophys. J.* **420**, 439 (1994)
9. J.C. Mather et al: *Astrophys. J.* **512**, 511 (1999)
10. J.R. Bond, C.R. Contaldi, D. Pogosyan: *Phil. Trans. Roy. Soc. Lond. A* **361**, 2435 (2003)
11. C.L. Bennett et al: *Astrophys. J. Suppl.* **148**, 1 (2003)
12. D.N. Spergel et al: *Astrophys. J. Suppl.* **148**, 175 (2003)
13. J.M. Kovac et al: *Nature* **420**, 772 (2002)
14. W. Hu, S. Dodelson: *Ann. Rev. Astron. Astrophys.* **40**, 171 (2002)
15. W. Hu: *Ann. Phys.* **303**, 203 (2003)
16. P.J.E. Peebles: *Astrophys. J.* **153**, 1 (1968)
17. Y.B. Zeldovich, V.G. Kurt, R.A. Syunyaev: *Journal of Experimental and Theoretical Physics* **28**, 146 (1969)
18. S. Seager, D.D. Sasselov, D. Scott: *Astrophys. J. Suppl.* **128**, 407 (2000)
19. A. Kogut et al: *Astrophys. J. Suppl.* **148**, 161 (2003)
20. C. Dickinson et al: preprint astro-ph/0402498, (2004).
21. B.S. Mason et al: *Astrophys. J.* **591**, 540 (2003)
22. T.J. Pearson et al: *Astrophys. J.* **591**, 556 (2003)
23. C.L. Kuo et al: *Astrophys. J.* **600**, 32 (2004)
24. C.W. Misner, K.S. Thorne, J.A. Wheeler: *Gravitation*, (W. H. Freeman and Company, San Francisco 1973) pp 583–590
25. W. Hu et al: *Phys. Rev. D* **57**, 3290 (1998)
26. A. Challinor: *Phys. Rev. D* **62**, 043004 (2000)
27. G.F.R. Ellis, J. Hwang, M. Bruni: *Phys. Rev. D* **40**, 1819 (1989)
28. A. Challinor, A. Lasenby: *Astrophys. J.* **513**, 1 (1999)
29. A. Challinor: *Class. Quantum Grav.* **17**, 871 (2000)
30. R. Maartens, T. Gebbie, G.F.R. Ellis: *Phys. Rev. D* **59**, 083506 (1999)
31. T. Gebbie, G.F.R. Ellis: *Ann. Phys.* **282**, 285 (2000)
32. T. Gebbie, P.K.S. Dunsby, G.F.R. Ellis: *Ann. Phys.* **282**, 321 (2000)
33. C. Ma, E. Bertschinger: *Astrophys. J.* **455**, 7 (1995)
34. M. Bartelmann, P. Schneider: *Phys. Rep.* **340**, 291 (2001)
35. M.L. Wilson: *Astrophys. J.* **273**, 2 (1983)
36. U. Seljak, M. Zaldarriaga: *Astrophys. J.* **469**, 437 (1996)
37. A. Lewis, A. Challinor, A. Lasenby: *Astrophys. J.* **538**, 473 (2000)
38. N. Kaiser: *Mon. Not. R. Astron. Soc.* **202**, 1169 (1983)
39. W. Hu, N. Sugiyama: *Astrophys. J.* **444**, 489 (1995)
40. W. Hu, N. Sugiyama: *Phys. Rev. D* **51**, 2599 (1995)
41. M. Bucher, K. Moodley, N. Turok: *Phys. Rev. D* **62**, 083508 (2000)
42. L. Page et al: *Astrophys. J. Suppl.* **148**, 233 (2003)
43. C. Gordon, A. Lewis: *Phys. Rev. D* **67**, 123513 (2003)
44. M. Bucher et al: preprint astro-ph/0401417, (2004)
45. J. Silk: *Astrophys. J.* **151**, 459 (1968)
46. J.M. Stewart: *Class. Quantum Grav.* **7**, 1169 (1990)
47. S. Boughn, R. Crittenden: *Nature* **427**, 45 (2004)
48. M.R. Nolta et al: preprint astro-ph/0305097, (2003)
49. P. Fosalba, E. Gaztañaga and F.J. Castander: *Astrophys. J. Lett.* **597**, 89 (2003)

50. R.H. Becker et al: *Astron. J.* **122**, 2850 (2001)
51. S.G. Djorgovski et al: *Astrophys. J. Lett.* **560**, 5 (2001)
52. U. Seljak, M. Zaldarriaga: *Phys. Rev. Lett.* **78**, 2054 (1997)
53. M. Kamionkowski, A. Kosowsky, A. Stebbins: *Phys. Rev. Lett.* **78**, 2058 (1997)
54. J.R. Bond et al: *Phys. Rev. Lett.* **72**, 13 (1994)
55. L.F. Abbott, R.K. Schaefer: *Astrophys. J.* **308**, 546 (1986)
56. G. Efstathiou, J.R. Bond: *Mon. Not. R. Astron. Soc.* **304**, 75 (1999)
57. A.G. Riess et al: *Astron. J.* **116**, 1009 (1998)
58. S. Perlmutter et al: *Astrophys. J.* **517**, 565 (1999)
59. W.L. Freedman et al: *Astrophys. J.* **553**, 47 (2001)
60. A. Lewis, A. Challinor, N. Turok: *Phys. Rev. D* **65**, 023505 (2002)
61. M. Zaldarriaga: *Phys. Rev. D* **55**, 1822 (1997)
62. G. Holder et al: *Astrophys. J.* **595**, 13 (2003)
63. G. Hinshaw et al: *Astrophys. J. Suppl.* **148**, 135 (2003)
64. E. Komatsu et al: *Astrophys. J. Suppl.* **148**, 119 (2003)
65. A. de Oliveira-Costa et al: preprint astro-ph/0307282, (2003)
66. P. Vielva et al: preprint astro-ph/0310273, (2003)
67. C.J. Copi, D. Huterer, G.D. Starkman: preprint astro-ph/0310511, (2003)
68. H.K. Eriksen et al: preprint astro-ph/0401276, (2004)
69. F.K. Hansen et al: preprint astro-ph/0402396, 2004.
70. A.C.S. Readhead et al: preprint astro-ph/0402359, (2004)
71. R. Rebolo et al: preprint astro-ph/0402466, (2004)
72. U. Seljak, P. McDonald, A. Makarov: *Mon. Not. R. Astron. Soc.* **342**, L79 (2003)
73. A. Slosar, U. Seljak, A. Makarov: preprint astro-ph/0403073, (2004)
74. S.M. Leach, A.R. Liddle: *Mon. Not. R. Astron. Soc.* **341**, 1151 (2003)
75. M. Tegmark et al: preprint astro-ph/0310723, (2003)
76. R.A. Sunyaev, Y.B. Zeldovich: *Comm. Astrophys. Space Phys.* **4**, 173 (1972)
77. R.A. Sunyaev, Y.B. Zeldovich: *Mon. Not. R. Astron. Soc.* **190**, 413 (1980)
78. J.P. Ostriker, E.T. Vishniac: *Astrophys. J. Lett.* **306**, 51 (1986)
79. A. Blanchard, J. Schneider: *Astron. Astrophys.* **184**, 1 (1987)
80. C.M. Hirata, U. Seljak: *Phys. Rev. D* **68**, 083002 (2003)

A Conceptual Tour About the Standard Cosmological Model

Antonio L. Maroto and Juan Ramírez

Dept. Física Teórica I, Universidad Complutense de Madrid, 28040 Madrid, Spain

Acknowledgement

This paper is dedicated to Prof. Alberto Galindo on occasion of his 70th birthday. We are delighted to write on Cosmology, which is one of his favourite subjects. He taught both of us at the Complutense University in Madrid, and we would like to thank him for his outstanding teaching. His insistence on logical reasoning, physical intuition and mathematical comprehension of natural laws followed the best tradition of Galilei: *the book of Nature is written in mathematical language.*

Summary. With the beginning of the XXIst century, a physical model of our Universe, usually called the Standard Cosmological Model (SCM), is reaching an important level of consolidation, based on accurate astrophysical data and also on theoretical developments. In this paper we review the interplay between the basic concepts and observations underlying this model. The SCM is a complex and beautiful building, receiving inputs from many branches of physics. Major topics reviewed are: General Relativity and the cosmological constant, the Cosmological Principle and Friedmann-Robertson-Walker-Lemaître models, Hubble diagrams and dark energy, large scale structure and dark matter, the cosmic microwave background, Big Bang nucleosynthesis, and inflation.

1 Introduction

Cosmology is the science that studies the Universe as a whole. From the dawn of civilization mankind has asked questions about the structure and composition of the Universe and the laws that govern it. Examples of such questions are

- How old is the Universe?
- What is the size and the geometry of the Universe?
- How did the Universe began and how will it end?
- What is the composition of the Universe?
- How did the matter and the structures that we observe in the Universe originate?

A very remarkable historical and philosophical feature of the present epoch, and in particular of the last few years, is the emergence of Cosmology as a mature science in which most of the former questions have a precise answer. For example, within a few percent, we know that the Universe is 13.7 Gyr. old, and that the part of it within our present horizon is flat and has a radius of 14-15 Gpc¹. These definite answers to such fundamental questions are possible in the context of the Standard Cosmological Model (SCM), that arised in the second half of the last century with the discovery of the Cosmic Microwave Background (CMB). The SCM is presently growing towards a level of development and reliability comparable to the Elementary Particles Standard Model (EPSM). This achievement has been possible due to the combined effort of theory based on fundamental physics, and a large number of astronomical observations from a host of scientific satellite and balloon probes, and also from the ground.

The SCM is based on five strong pillars: i) The General Theory of Relativity, introduced by A. Einstein in 1916, which provides the theory of the gravitational field and the basic framework for the cosmological models. ii) The Cosmological Principle, also introduced by Einstein in 1917, that states the homogeneity and isotropy of the Universe. iii) The Hubble law, discovered by E. Hubble in 1929, establishing the expansion of the Universe, and the Hubble diagrams which allow to determine the acceleration/deceleration of the Universe by means of standard candles. iv) The CMB corresponding to blackbody radiation at $T = 2.725$ K, accidentally discovered by A. Penzias and R. Wilson in 1964, whose mean isotropy supports the cosmological principle, and whose small anisotropies in the spatial distribution of the sky temperatures contain a wealth of information on the cosmological parameters. v) The light elements cosmological abundances, namely ^1H , ^2D , ^3He , ^4He , and ^7Li , originated during the primordial Big Bang Nucleosynthesis (BBN) when the Universe was 100 s. old, whose theoretical analysis pioneered by G. Gamow in 1948, reinforces the Big Bang scenario. To this list it should be added the analysis of the Large Scale Structure (LSS) in the Universe based on galaxy catalogs, which has recently received a major boost with the Two degree Field Galaxy Redshift Survey (2dFGRS) and the Sloan Digital Sky Survey (SDSS).

The elucidation of the impact of the former pillars in the model conveys a lot of fundamental physics, from which the EPSM is not the lesser part. From this elucidation emerges the SCM as the Hot Big Bang model with the addendum of a primordial inflationary phase. In the Hot Big Bang model, the Universe expands and cools from a very dense and hot state. The temperature of the plasma in the early phase, and of the CMB photons later on, scale as $T \propto a^{-1}$, with T being the temperature and a the scale factor of the Universe. The history of this expansion and cooling can be followed backwards in time in terms of well known fundamental physics until a time $t \sim 10^{-5}$ s corresponding to a temperature $T \sim 10^{12}$ K, when the Universe was 10^{12} times smaller than it is now. This corresponds to a mean energy in the plasma of 100 MeV, and although particle accelerators have explored the EPSM up to energies of 1 TeV, the accurate control of the physics that is going on is lost at this point because of our present lack of knowledge about the hadron to quark-gluon plasma phase transition, which ought to occur around 170 MeV. However the basic features of the model can be further extrapolated up to the Grand Unification temperature $T_{GU} \sim 10^{29}$ K ($t \sim 10^{-39}$ s) and ultimately to the Planck temperature

¹1pc = 3.2615 light-years = 3.0856×10^{13} km

$T_P \sim 10^{32}\text{K}$ ($t_P \sim 10^{-44}\text{s}$) where our lack of understanding of *quantum gravity?* prevents further extrapolation to earlier times. Classically, the extrapolation to zero time would lead to a state of infinite temperature and density known as the Big Bang singularity. Thus, the name Big Bang has two meanings in Cosmology. The first refers to the very hot and dense plasma composed of protons, neutrons, electrons, positrons, neutrinos and photons existing at $t \sim 10^{-5}\text{s}$ from which our present Universe, containing ourselves, emerges through expansion, cooling and growing of structures. This Big Bang is firmly established through astronomical observations related to CMB, BBN, LSS and Hubble diagrams. The second meaning refers to the classical singularity from which the Universe seems to spring at Planck time $t_P \sim 10^{-45}\text{s}$. A consistent quantum mechanical description of this phase is unfortunately not yet available, but this is a fertile ground for promising theoretical speculations like string theory, M-theory, branes and other TOEs (Theory of Everything).

In 1980 A. Guth and in 1981 A. Linde, inspired in Grand Unified Theories, introduced the hypothesis that around the epoch between Planck time and Grand Unification time, the Universe underwent a period of rapid exponential expansion which augmented its size by a factor between e^{50} and e^{70} . This idea was introduced to solve the so-called problems of the Hot Big Bang model, meaning to find a mechanism which explains why the Universe is so smooth, old and flat. The introduction of the inflationary phase provides a compelling natural explanation, but more importantly, it gives a model for structure formation based on the quantum fluctuations of the inflaton field, whose ground state energy drives the rapid exponential expansion of the Universe.

In this short review we will present a broad view description of the Standard Cosmological Model, and of the theoretical ideas and observational facts that support it. An updated textbook on the Standard Cosmological Model is for example [1].

2 General Relativity, the Cosmological Principle and FRWL Models

2.1 General Relativity and the Cosmological Constant

Despite its weakness with respect to other fundamental forces, the long range and the absence of screening make gravity the driving force of the cosmos. To the extent of the present knowledge, the gravitational force is correctly described by Einstein's General Relativity, which also provides the geometrical framework for cosmological models. In the broad view General Relativity consists of two basic elements: *the Equivalence Principle* and *Einstein's field equations*.

In its Newtonian version the Equivalence Principle amounts to the identification of the gravitational mass entering Newton's gravity law and the inert mass entering the second law of Newtonian mechanics. The first precise experimental test of this equality was performed by R. Eötvös in 1890, who showed that the ratio m_g/m_i does not differ from one substance to another in more than one part in 10^9 . The most accurate test that all bodies fall with the same acceleration in a gravitational field comes from the comparison of the accelerations of the Moon and the Earth as they fall around the Sun by means of lunar laser ranging. These accelerations agree to an accuracy of 1.5×10^{-13} [2].

Equality of the inertial and gravitational masses led Einstein to formulate the relativistic version of the equivalence principle by identifying the gravitational field with the metric tensor $g_{\mu\nu}$ describing the (pseudo) Riemannian geometry of the space-time manifold.

The second element of General Relativity are Einstein's field equations for the gravitational field

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi GT_{\mu\nu} + \Lambda g_{\mu\nu} . \quad (1)$$

These equations derived by D. Hilbert and A. Einstein in 1916, are similar in spirit to Maxwell's equation for the electromagnetic field, and relate the geometry of space-time embodied in the Ricci tensor $R_{\mu\nu}$ and the scalar curvature R , with the stress-energy tensor $T_{\mu\nu}$ as its source.

The second term in the second member, containing the *cosmological constant* Λ , was introduced by Einstein only in 1917, to obtain his famous static cosmological solution. After the discovery of the expansion of the Universe by E. Hubble in 1929, Einstein considered the introduction of the Λ term as the biggest mistake of his life. This case provides a good example of the (dangerous) role that preconceptions can play in Cosmology. The idea of a stationary Universe was the current paradigm, which had reigned for centuries, when Einstein introduced General Relativity. The idea of *evolution*, a common place for Life Sciences since the nineteenth century, was taken seriously in Cosmology only after the discovery of the CMB in 1964.

Ironically enough, after long theoretical efforts to prove that Λ should be equal to zero, the Λ term reappears in two places in modern Cosmology: during the inflationary phase as the vacuum energy of the inflaton field, and as a natural explanation for the observed acceleration of the Universe, discovered in 1998 [3]. In the modern view of Quantum Field Theory, the cosmological constant should receive a contribution from the vacuum energy or zero point energy of the oscillatory modes of the quantum fields. Then, in terms of the energy density $\Lambda/8\pi G$, a natural value would be $M_p^4 \sim 10^{76}$ GeV⁴. This value is even unacceptable for the inflationary phase of the Universe, and it is disparately wrong by 124 orders of magnitude when compared with a cosmological constant driving the present acceleration of the Universe. In addition, each phase transition, like the GUT or the electroweak, shoud give a jump in the vacuum energy density of the order of E^4 , with E being the energy scale of the transition. This rather odd state of affairs with Λ , is undoubtedly one of the most important problems that needs to be settled in physics.

On the other hand, the present value of the cosmological constant, although very important for cosmological dynamics, is too small to have any effect at non-cosmic distances. Indeed Einstein's equations without Λ have been experimentally checked in the weak fields of the Sun and the Earth. In this regime, the simplest Einstein's equations without Λ provide a correct description of gravity. However, the same cannot be guaranteed in strong field regimes or at very short or very long scales. In such cases, modifications of Einstein's equations could be possible. This is the case for instance in string theory and higher derivative theories of gravity. In this sense the equivalence principle, i.e. the description of the gravity field by a metric tensor, seems to be much more fundamental than the precise form of the dynamical equations for the gravitational field.

2.2 The Cosmological Principle

Introduced by Einstein in 1917, the term *Cosmological Principle* was coined by E. Milne in 1933. It states the homogeneity (translation invariance) and isotropy (rotation invariance about each point) of the three-dimensional space-like slices of the Universe at each instant of time. There are very good reasons to formulate this principle. There is very strong observational evidence that the Universe is isotropic about the position occupied by the Earth. The distribution of different backgrounds like galaxies, radio sources, or X-ray background, are pretty isotropic around our position on large scales, but by far the most precise observational test of isotropy comes from CMB. After subtracting the dipole term in the angular distribution of the CMB temperature on the sky, a temperature distribution remains, whose departure from isotropy is only about one part in 10^5 .

From the observation of isotropy around our position, homogeneity can also be inferred through the so-called *Copernican Principle*. Copernicus believed that the Sun, not the Earth was the center of the Universe. Later on, it was discovered that neither the Sun nor the Milky Way occupy a special place in the Universe. In 1960 H. Bondi, coined the term Copernican Principle to mean that we at Earth do not occupy a special place in the Universe. Therefore, since we observe isotropy, it means that isotropy should be observed from any position in the Universe. Finally, geometry tells us that isotropy about all points imply homogeneity. Thus the observed isotropy of CMB from our position, plus the Copernican Principle, imply the homogeneity of the Universe.

On the other hand recent tridimensional galaxy catalogs [4], [5], provide direct observational support for homogeneity. Figure 1 shows the power spectrum, i.e the density contrast $\delta\rho/\rho$ as a function of the averaging scale. 2dF and SDSS maps reach up to 600 Mpc deep in the sky, and from Figure 1 we see that the mass density fluctuations in galaxy distribution diminish to 10% at scales around 400 Mpc. 10% is not a very high precision compared to the 10^{-5} for isotropy, and so the observational check for homogeneity is not so strong as for isotropy. In addition, it has been also argued [6], that the distribution of galaxies could be compatible with a part of a fractal up to the limits of the present catalogs. It should be noticed however, that this does not disprove the Cosmological Principle, nor set reliable limits to the true scale of homogeneity in the geometry. Even if the crossover to homogeneity in luminous matter happens to be at a scale which is not yet reached, it should be taken into account that luminous matter represents less than 1% of the energy density driving the geometry of the Universe. Nevertheless, understanding the relation between the complex galaxy structures and the smooth CMB represents an extremely interesting and important problem at the heart of the theory of structure formation.

Another important physical aspect of the cosmological principle is that homogeneity and isotropy are connected to the fact that the present Universe comes from a previous state of thermal equilibrium: the Hot Big Bang. This is a very satisfactory state of affairs because it makes the Universe comprehensible, since its present state can be understood through causal laws without any reference to particular initial conditions. There is however a caveat. As we shall discuss below, the existence of a suitable inflationary phase guarantees that the region of the Universe that is observable today, comes from a tiny patch which was causally connected -therefore able to be in thermal equilibrium- before inflation happened. Thus, it could well be that we live not very near to the edge of a smooth, homogeneous and isotropic patch

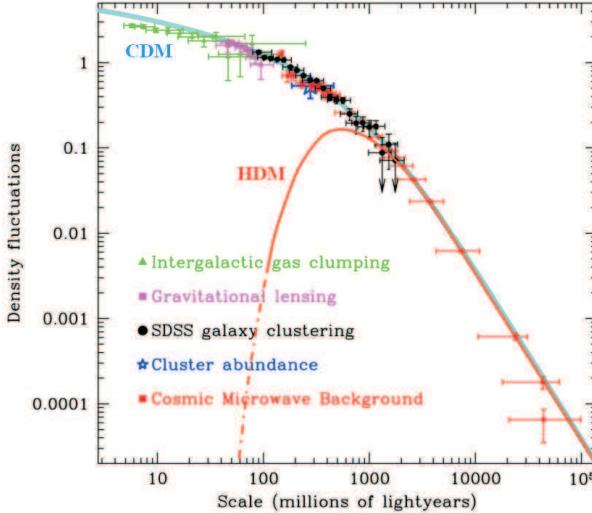


Fig. 1. Scale dependence of density fluctuations from CMB anisotropies and large-scale structure. The continuous line labelled by CDM corresponds to a flat model with cold dark matter ($\Omega_M = 0.28$, $h = 0.72$) and a scale invariant spectrum of fluctuations ($n_s = 1$). The behaviour for $\lambda < \lambda_{eq} \simeq 2.9 \cdot 10^8$ lightyears is approximately logarithmic, whereas it decays as λ^{-2} for $\lambda > \lambda_{eq}$. The line labelled by HDM represents the typical behaviour of a hot dark matter model, with a sharp cut-off for $\lambda < \lambda_{FS}$, (data from SDSS collaboration [5])

caused by inflation, with the Universe as whole being chaotic. So, it is possible that the Cosmological Principle is valid only in a local sense.

The Cosmological Principle together with the Equivalence Principle dictates the geometry of the Universe given by the Robertson-Walker metric

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

In this metric the function $a(t)$ is the scale factor of the Universe, the constant $k = 1, 0, -1$ specifies the sign of the spatial curvature of the Universe, and t is the cosmic time. The cosmic time is the time measured by the fundamental or comoving observers which are at rest with respect to the expansion. They can be characterized as those measuring zero dipole anisotropy in CMB. Since the peculiar velocity of the Sun is roughly 370 km/s with respect to CMB, the time measured by our watches at Earth coincides with cosmic time with an error of approximately 1 part in 10^6 .

The coordinates r, θ, ϕ are (spherical) comoving coordinates meaning that comoving observers remain at rest in these coordinates. Velocities with respect to these coordinate systems are called *peculiar velocities*. Galaxies are nearly comoving, their peculiar velocities being about a few hundred km/s.

As the Universe expands, physical distances between comoving objects and wavelengths scale with $a(t)$. Thus, the wavelength of a freely propagating photon is stretched in proportion to the expansion factor from the epoch of emission to de-

tection

$$1 + z = \frac{\lambda_{\text{obs}}}{\lambda_{\text{em}}} = \frac{a_{\text{obs}}}{a_{\text{em}}} . \quad (3)$$

This expression also defines the redshift parameter z .

The observable region of the Universe at a given time is limited by the physical distance that light can travel since the Big Bang. This distance, called the *particle horizon* is directly related to the scale factor by $d_H(t) = a(t) \int_0^t dt' a^{-1}(t')$.

The Cosmological Principle also restricts the form of the material content of the Universe. Since a perfect fluid can be characterized by its isotropy around observers comoving with the fluid, the stress-energy tensor for the material content of the Universe must have the perfect fluid form

$$T_{\mu\nu} = pg_{\mu\nu} + (p + \rho)u_\mu u_\nu , \quad (4)$$

where p and ρ are the pressure and the energy density measured by a comoving observer, and u^μ is the four velocity of the fluid.

2.3 FRWL Models

The Robertson-Walker (2) metric provides the kinematical framework for cosmological models. It plays a similar role to the metric of the two-dimensional sphere in the study of Geography. Thus the analysis of cosmological observations based only in the RW metric, without any dynamical assumptions, is sometimes called *Cosmography*.

Cosmological dynamics i. e. the study of the time variation of the scale factor and the cosmological densities for the various matter species entering the composition of the Universe, is obtained by the combination of the second basic element of General Relativity: the dynamical equations for the gravitational field, and the Cosmological Principle. This leads to the Friedmann-Lemaître equation

$$H^2 = \frac{8\pi G}{3}\rho + \frac{1}{3}\Lambda - \frac{k}{a^2} , \quad (5)$$

and the energy conservation equation

$$\frac{d}{dt}(\rho a^3) = -p \frac{d}{dt}a^3 , \quad (6)$$

where $H = \dot{a}/a$ is called the *Hubble parameter*. Its present value H_0 is the *Hubble constant*, usually expressed in terms of the adimensional number h in the form $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$. Thus, solving Friedmann-Lemaître equation (5) relates the expansion rate of the Universe given by the Hubble parameter $H(t)$, with the cosmic time t and the red-shift parameter z .

The various species entering the cosmological models are assumed to satisfy linear equations of state of the form $p = w\rho$. This includes in particular the cases of photons in the CMB or an ultrarelativistic plasma ($w = 1/3$), cold non relativistic matter ($w = 0$), and the cosmological constant ($w = -1$). In addition, the density parameters for each species are defined as $\Omega_i = 8\pi G\rho_i/3H^2$, where $3H^2/8\pi G$ is the *critical density* corresponding to a flat Universe ($k = 0$) with $\Lambda = 0$. Moreover Friedmann-Lemaître equation (5) when rewritten as

$$\sum_i \Omega_i + \Omega_\Lambda - \frac{k}{a^2 H^2} = 1 \quad , \quad (7)$$

relates the density parameters to the spatial curvature, including $\Omega_\Lambda = \Lambda/3H^2$ as the density parameter for the cosmological constant. Thus flat universes are those fulfilling the condition $\sum_i \Omega_i + \Omega_\Lambda = 1$. It is important to remark that all universes having a Big Bang, are nearly flat at early times since the density parameter for curvature $\Omega_k = -k/a^2 H^2 \rightarrow 0$ as $t \rightarrow 0$

On the other hand the temporal evolution for the densities is given, in terms of the red-shift parameter z , by the energy conservation equation (6), and the equation of state

$$\Omega_i H^2 (1+z)^{(-3-3w_i)} = \frac{8\pi G}{3} \rho_i (1+z)^{(-3-3w_i)} = \text{const.} \quad (8)$$

As we shall discuss below, observations favour a present Universe which is nearly flat and composed of cold matter and a cosmological constant, with present values of density parameters $\Omega_M \approx 0.27$ and $\Omega_\Lambda \approx 0.73$. As shown in Figure 2 the selection of a region in the Ω_M - Ω_Λ plane centered about these values, arises from the combination of observational information from supernova type Ia Hubble diagrams, CMB anisotropies, and clustering of galaxies in LSS. Taking into account the measured value of the Hubble constant $h \approx 0.72(10\%)$ [7] and solving Friedmann-Lemaître equation for these models, yields an age for the Universe around 13.7 Gyr, and a distance to the particle horizon about 14-15 Gpc.

According to the evolution equation for the densities (8), Ω_Λ decreases and Ω_M increases with increasing red shift. Thus, going backwards in time, for $t \lesssim 1$ Gyr, Ω_Λ becomes negligible and we are left with a critical Universe dominated by cold matter, until radiation begins to dominate. In addition, from the analysis of CMB anisotropies, and from BBN, it follows that only about 15% of this cold matter are baryons, the rest being *cold dark matter*.

Planck's formula for blackbody radiation translates the present CMB temperature $T = 2.725$, into an energy density of CMB photons $\Omega_\gamma h^2 = 2.48 \times 10^{-5}$. Including three massless (or very light) species of neutrinos, the total energy density in radiation would be at present $\Omega_R h^2 = 2.48 \times 10^{-5}$, which is negligible in front of the energy density in cold matter. However as we go backwards in time, the number density of radiation particles scales as a^{-3} , and the wavelength shrinks in proportion to the scale factor a . Therefore, the energy density in radiation scales as a^{-4} while the energy density in cold matter goes with a^{-3} . As a consequence, at sufficient earlier times, radiation dominates the energy density of the Universe. For $h \approx 0.72$ and $\Omega_M \approx 0.27$ the energy densities in matter and radiation become equal at $z_{\text{eq}} \approx 3300$, which corresponds to an approximately 55 kyr old Universe.

3 Hubble Law, Hubble Diagrams and Dark Energy

In the approximation in which galaxies are comoving, the physical distance to a given galaxy scales with $a(t)$, and consequently its recession velocity V is related to its physical distance d at a given time, by

$$V = Hd \quad . \quad (9)$$

This is the *theoretical Hubble law* which is exact, and a direct consequence of the Robertson-Walker form of the cosmic metric. This relation cannot be directly checked because neither the recession velocities nor the physical emission distances to galaxies are empirically measurable.

The Robertson-Walker form of the metric was established in 1936. Earlier, in 1929 E. Hubble found the *empirical Hubble law*

$$z = H_0 d_L , \quad (10)$$

linearly relating the red-shift of galaxies to their luminosity distance. The luminosity distance is defined as $d_L = \sqrt{L/4\pi\mathcal{F}}$, where L is the absolute luminosity of the source and \mathcal{F} its apparent luminosity, i. e. the flux of energy received in the collecting surface of the telescope. So, luminosity distance d_L is defined as such that a source of absolute luminosity L , located in a static Euclidean space, would produce a flux \mathcal{F} at distance d_L . From RW metric, it follows that the relation between d_L and the red-shift parameter z is nonlinear. To second order this relation takes the form

$$H_0 d_L(z) = z + \frac{1}{2}(1 - q_0)z^2 + \dots , \quad (11)$$

where $q = -a\ddot{a}/\dot{a}^2$ is the *deceleration parameter* of the Universe and q_0 its present value. It follows then from RW metric, i. e. from Cosmological Principle, that the empirical Hubble law can be expected to be true only for $z \ll 1$. On the other hand for $z \ll 1$, the approximate equalities $d \approx d_L(z)$ and $V \approx z$ hold. Thus empirical and theoretical Hubble laws coincide in this regime. In this weak sense checking Hubble's law is also a check of the RW metric.

To accurately check Hubble's law, measuring H_0 within the linear approximation to (11), and eventually going deeper in red-shift to determine q_0 , has been a central research program in Cosmology since 1929, called Hubble program. The key observational tool for this endeavour are *standard candles*: luminous sources whose absolute luminosity has been properly calibrated. Once a class of sources has been calibrated, a Hubble diagram can be obtained by representing these sources in a two-dimensional plot of luminosity distances versus red-shifts, the final goal being to extract from the observational points the cosmological parameters H_0 and q_0 . This program received a major boost with the launching of the *Hubble Space Telescope* (HST) in 1990, whose so-called *Key project* was to perform a precise measure of H_0 .

Measuring cosmic distances starts by the trigonometric parallax of nearby stars, due to the annual motion of the Earth around the Sun. From this starting point, a *cosmic distance ladder* is built by means of standard candles. This method works typically by finding precise correlations between the absolute luminosity and another observable for a definite class of objects. The first and basic step is provided by Cepheid variable stars, whose absolute luminosity is tightly related to its period. HST has been able to resolve thousands of Cepheid variables in galaxies up to 20 Mpc. Once the distances to these nearby galaxies are fixed, five different methods are used to go up to 400 Mpc. Three of them are based on global properties of spiral and elliptical galaxies: Tully-Fisher relation that links rotation velocities of spiral galaxies to their luminosity, relation between star velocities dispersion and luminosity in ellipticals, and fluctuations in galaxies surface brightness. The other two are based on the use of supernovae type Ia (SNe Ia), and supernovae type II

(SNe II) as standard candles. The combination of all these methods yields for the Hubble constant a weighted average $H_0 = 72 \pm 8 \text{ km s}^{-1} \text{ Mpc}^{-1}$ [7]. It is most remarkable that this value for H_0 agrees with the value obtained by the analysis of the CMB anisotropy map provided by WMAP: $H_0 = 71 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (5%).

Due to their big absolute luminosity, SNe Ia are the most far reaching standard candles in red-shift. This makes these objects ideal tools to investigate the deceleration parameter q_0 . In 1998 two groups: the *High z Supernovae Search Team* (HZT) and the *Supernova Cosmology Project* (SCP) using SNe Ia, reported the discovery of the accelerated expansion of the Universe ($q_0 < 0$) [3]. The result appears through an exceeding faintness of supernovae as they would have in a decelerating universe. Notice that from eq. (11), an accelerating universe ($q_0 < 0$) results in bigger luminosity distances -hence, fainter objects- for the same red-shift than a decelerating one ($q_0 > 0$). Thus, once other astrophysical effects like dust or evolutionary effects have been discarded, the exceeding faintness of supernovae should be interpreted as due to the acceleration of the Universe expansion.

Friedmann's equations (5), (6) imply that an accelerating Universe should contain part of its energy density in a substance with equation of state parameter $w < -1/3$. This kind of substance goes under the name of *dark energy*, and the most obvious candidate is a cosmological constant ($w = -1$). For a model composed of dark matter plus a cosmological constant, the deceleration parameter $q = \frac{1}{2} \Omega_M - \Omega_\Lambda$, and as shown in Figure 2, the supernovae data select a linearly shaped maximum likelihood region in the $\Omega_M - \Omega_\Lambda$ plane. So, supernovae data alone, suffice to establish the acceleration of the Universe with a very high confidence level. When supplemented with the data coming from CMB and LSS, a best fit is obtained for $\Omega_M \approx 0.28$ and $\Omega_\Lambda \approx 0.72$, corresponding to a nearly flat Universe.

Although the LSS models favored the so-called Λ CDM (cosmological constant plus cold dark matter) scenario, the discovery of the acceleration of the Universe in 1998 came as a surprise, since a universe filled with cold matter ($w = 0$) decelerates ($q = \frac{1}{2} \Omega_M > 0$). However as explained above, when we go backwards in time, Ω_M grows and Ω_Λ decreases. So, at earlier times there should have been a decelerating period of the Universe. In fact, very recently [8], using the HST, 16 new type Ia supernovae has been found at very high red-shift, up to $z \approx 1.7$, which give conclusive evidence for this decelerating period. These newly discovered SNe Ia, together with the 170 previously reported confirm the concordance model with $\Omega_M \approx 0.28$ and $\Omega_\Lambda \approx 0.72$, and give a value for the red-shift of the transition between the accelerating and decelerating epochs $z = 0.46 \pm 0.13$. It is the most remarkable that the confirmation of the effect of deceleration in supernovae, almost fully rule out alternative astrophysical explanations like dust or evolutionary effects for the luminosity distance versus red-shift distribution, since it is very unlikely for these effects to exactly mimick the deceleration/acceleration transition.

In addition to the cosmological constant, other forms of dark energy could be possible with a different or even a variable equation of state w . An interesting class of models are *quintessence* models in which dark energy is the energy density of an evolving scalar field, much the same way as the inflaton during the inflationary phase of the Universe. For a constant equation of state, SNe Ia data yield $w = -1.02_{-0.19}^{+0.13}$ and $w < -0.76$ to 95% confidence level [8]. An alternative proposal to *dark energy* as an explanation for the deceleration/acceleration of the Universe could be the weakening of gravity in our 3+1 dimensions by leaking into the extra dimen-

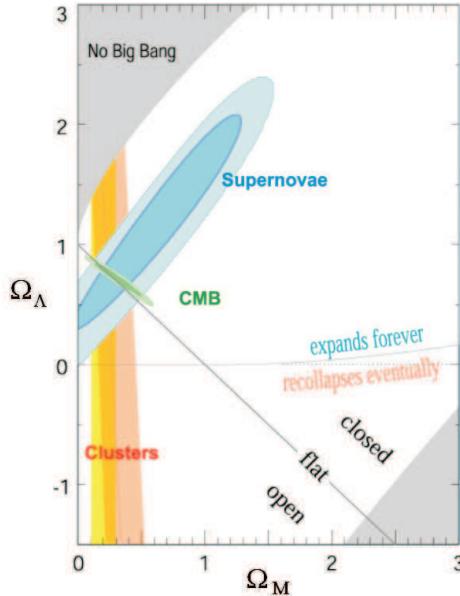


Fig. 2. Matter density Ω_M vs. vacuum energy density Ω_Λ 68% and 95% C.L. contours for supernovae, cluster and CMB data, (from Supernova/Acceleration Probe: SNAP collaboration [9])

sions as suggested by string theories. A major observational effort will be needed to discriminate among these competing models.

4 Large Scale Structure and Dark Matter

Far from a smooth distribution, matter exhibits a complex clustering pattern in the Universe. Thus, there are regions in which matter is strongly clumped forming galaxies, clusters and even larger structures, whereas at the same time, we can also find almost empty regions with very low densities. In fact, strong inhomogeneities can be found at galactic scales (~ 10 kpc), where the density contrast can be as large as $\delta\rho/\rho \sim 10^2$. Galaxies, which can be considered as the elementary building blocks of structures, are not homogeneously distributed either, but grouped hierarchically into groups, clusters and superclusters, the latter ones extending over distances of tens of Mpc. Filament-like chains of galaxies connect different superclusters in a network with scales of around 100 Mpc. Most of the matter distributes on the walls of this cell-like structure with large voids in between. In Figure 1 we can see a data plot showing the scale dependence of the density contrast. Each point represents the density fluctuation measured at a given scale λ , which is obtained by comparing the average density within a sphere of radius λ , as it is placed at different spatial positions. As we see from the data, the density contrast declines as we take larger and larger spheres, in agreement with the Cosmological Principle. Thus, we can

conclude that the Universe can be considered as approximately homogeneous only on very large scales (~ 1000 Mpc).

The SCM allows us to understand the growth of these structures from seeds of primordial density fluctuations. However, the origin of such seeds is left unspecified, this being one of the most important limitations of the classical standard cosmology. In Section 7 we will see a possible generation mechanism based on the idea of inflation. Leaving aside the issue of the origin, the growth mainly takes place during the matter dominated era as more and more matter is attracted towards the initially overdense regions. When the structure is sufficiently large (larger than the so-called Jeans scale), its gravitational self-attraction is able to decouple it from the Hubble expansion, forming a bound object. For small density fluctuations (linear regime), the growth rate is linear with the scale factor. This means that the total growth from the matter-radiation decoupling time ($a_{dec} \sim 10^{-3}$) until present ($a_0 = 1$) would be a factor 10^3 . However, the amplitude of CMB fluctuations at that time, measured by the COsmic Background Explorer (COBE) satellite, was only $\delta\rho/\rho \sim \delta T/T \sim 10^{-5}$. This obviously poses a problem since fluctuations have not had enough time to reach the non-linear regime $\delta\rho/\rho \gg 1$. Notice that in the previous reasoning it is assumed that matter fluctuations are comparable to temperature fluctuations at decoupling. This is indeed the case for baryons, which were coupled to photons until decoupling time, and implies that a universe dominated by baryons at the time of decoupling is not consistent with galaxy formation.

However, if there existed a new type of weakly coupled matter, which did not interact with radiation, its density fluctuations could have started growing much before, thus explaining the apparent mismatch. This is one of the strongest arguments in favor of the existence of dark matter. Once, baryons and radiation decouple, baryons will fall in the potential wells created by dark matter, their density fluctuations acquiring the same amplitude as that of dark matter.

The nature of dark matter determines the final density distribution at different scales. At present there are important projects which aim to collect information about the distribution of galaxies in the Universe. They are galaxy redshifts catalogues such as the 2dFGRS [4] and the SDSS [5]. The first one, which has been completed recently, has measured the redshift of 221000 galaxies over a five years period. The SDSS is in progress and is expected to measure the position and the absolute brightness of 100 million celestial objects. The information obtained from these catalogues combined with that coming from CMB anisotropies and high-redshift supernovae observations is allowing us to determine the cosmological parameters with unprecedented accuracy (see Fig. 2), and to shed light on the nature of dark matter.

Let us see how dark matter affects structure formation. For that purpose it is convenient to differentiate between the so-called hot (HDM) and cold dark matter (CDM) scenarios. In the hot case, dark matter is made out of light particles, which were still ultrarelativistic at the beginning of the structure formation period, with the typical candidate being a light neutrino. Since hot dark matter particles propagate close to the speed of light, they can escape from overdense regions into underdense ones, erasing the density fluctuations on scales smaller than the free-streaming scale λ_{FS} . This is the maximum distance that a particle can travel from the initial time until matter-radiation equality. Typical values are $\lambda_{FS} \sim 40$ Mpc, corresponding to the size of a large cluster, for a neutrino (or any other light particle) mass around $m_\nu \sim 30$ eV. This means that, in this scenario, galaxies cannot grow directly from primordial fluctuations. Superclusters should form first, reach the non-linear regime

and then, by fragmentation, give rise to small clusters and galaxies. As a consequence the power spectrum of density fluctuations should be strongly peaked around the λ_{FS} scale (see Fig. 1). However, recent CMB data together with 2dFGRS or SDSS information on the matter power spectrum, constrain the above effect. The results show that hot dark matter cannot be the dominant dark matter component and $\Omega_\nu h^2 \leq 0.0076$ (2dF+WMAP at 95 % C.L.). This bound can be translated into a very strict limit on the sum of the neutrino masses $\sum_i m_{\nu_i} \leq 0.71$ eV, which improves by several orders of magnitude the laboratory limits.

In the cold dark matter case, dark matter particles are already non-relativistic at matter-radiation equality. Thus, free-streaming damping is not a problem. As shown in Fig. 1, an enormous variety of observations at very large scales ($\gtrsim 1$ Mpc), from cosmic microwave background anisotropies, galaxy surveys, cluster abundances or Ly- α forest are successfully explained within the CDM framework. Despite its success at large scales, the model exhibits certain difficulties at sub-galactic scales. In particular, high resolution N-body simulations of dark halos show cuspy density profiles which contradict observations from low surface brightness galaxies and dwarfs which indicate flatter density profiles. In addition, CDM also predicts too many small subhalos within simulated larger systems, in contradiction with observations of the number of satellite galaxies in the Local Group. In any case a spatially flat Universe with cosmological constant and cold dark matter (Λ CDM) is generally accepted at present as the Standard Model in Cosmology.

4.1 Nature of Dark Matter

Apart from the difficulties of a baryon dominated universe to explain galaxy formation, there are additional evidences that the luminous mass of the Universe is only a small fraction of the total matter density [10]. This deficit is present in two different contexts: first at galactic scales, where dark matter is believed to form spherical halos several times bigger than the galactic disks; and second at cosmological scales. Many different kinds of observations such as rotation curves of galaxies, weak gravitational lensing, cluster abundance, virial motions in clusters, matter power spectrum, CMB anisotropies, ..., agree in a value for the total matter $\Omega_M = 0.27 \pm 0.04$. This value should be compared to the baryon density $\Omega_B = 0.044 \pm 0.004$ and to the luminous mass density $\Omega_{lum} = 0.006 \pm 0.003$, i.e we find $\Omega_{lum} < \Omega_B < \Omega_M$. We thus have two dark matter problems, namely, there are missing baryons which do not contribute to the luminous matter and there is non-baryonic dark matter which make up most of the matter of the Universe.

Concerning the baryonic dark matter problem, it is difficult to find dark baryons in the galactic halos. They could be present in the form of hot or cold clouds of hydrogen, although an entire halo made of gas would conflict with observations of absorption or emission of radiation. They could form massive compact halo objects (MACHOs) similar to big planets. However the current limits from EROS and MACHO microlensing observations show that less than 25% of standard halos can be composed of MACHOs with masses between $10^{-7} - 1$ M_\odot .

On the other hand, the nature of non-baryonic dark matter is even a greater mystery. Different possible explanations include: massive neutrinos, modifications of gravity at large distances or the existence of a background of new weakly interacting massive particles (WIMPs). A natural dark matter particle should be neutral, stable, massive and weakly interacting, so that its relic density could contribute in an

important way to the matter density. Accordingly a massive neutrino would be the most economical solution. However, different limits prevent neutrinos from being a viable candidate. Thus, apart from the strong limit coming from 2dF + WMAP mentioned above, simply imposing that relic neutrinos do not overclose the Universe, i.e. $\Omega_\nu \lesssim 0.3$, then their masses should be either smaller than ~ 20 eV or larger than ~ 20 GeV.

Let us study in more detail each possibility. In the case of light (hot) neutrinos, if they are required to make up the galactic halos, their mass density should be $\rho_{halo} \simeq 0.3$ GeV cm $^{-3}$. However, their number density cannot exceed the limit imposed by the Pauli exclusion principle, so that their masses should be sufficiently high. In particular we get $m_\nu \gtrsim 20$ eV for spiral galaxies and $m_\nu \gtrsim 100$ eV for dwarf galaxies (Tremaine-Gunn limit), in contradiction with the overclose limit. In the case of heavy (cold) neutrinos, the overclose limit is much larger than the laboratory limits on the three known neutrino species, but still there is the possibility of the existence of a stable heavy fourth generation of Dirac or Majorana neutrinos. Current direct detection experiments have enough sensitivity to detect halo particles with cross-sections typical of weak interactions and masses above 20 GeV. However, at present there is no compelling evidence of the detection of such particles, so that a fourth generation of neutrinos is excluded. (DAMA experiment claims the detection of an annual modulation in its dark matter signal. However, such a result seems to be incompatible with other direct detection experiments as CDMS).

The absence of cold dark matter candidates within the known particles is one of the most pressing arguments for the existence of new physics beyond the Standard Model, either as new particles or as modifications of the gravitational interaction at large distances. Among the proposed new particle candidates, we find, on one hand, the axion which is the Goldstone boson associated to the spontaneous breaking of the Peccei-Quinn symmetry postulated to solve the strong CP problem of QCD. The production of axions in the early Universe mainly takes place through the so-called misalignment mechanism in which the Θ angle is initially displaced from its equilibrium value $\Theta = 0$, and oscillates coherently. Such oscillations can be interpreted as a zero-momentum Bose-Einstein condensate which essentially behaves as a non-relativistic matter fluid. Despite the fact that axions are light particles, this non-thermal mechanism produces cosmologically important energy densities. On the other hand we have the thermal relics, produced by the well-known freeze-out mechanism in an expanding Universe. They are typically weakly interacting massive particles (WIMPs) such as the neutralino in supersymmetric theories (for a recent review see [10]). In addition to their weak interactions with ordinary particles included in the EPSM, these candidates usually have also very weak self-interactions (collisionless).

At present there are several kinds of experiments (ground based and satellite borne) which aim to detect cold dark matter halo particles, either directly or indirectly. Direct detection experiments are based on the possibility of measuring the recoil energy that a target nucleus acquires in the elastic collision with a DM particle. Some of the experiments in progress are: DAMA, CRESST and GENIUS, at Gran Sasso Laboratory or CDMS at Soudan mine. The indirect experiments are based on the possibility of detecting the annihilation products of halo DM particles. Typically they include: gamma ray telescopes such as MAGIC (ground based) or GLAST (satellite) which could be sensitive to annihilations into pairs of photons; antimatter detectors such as AMS which can detect positrons produced in e^+e^- annihilations;

and finally high-energy neutrino telescopes such as ANTARES or AMANDA which will be sensitive to neutrino-antineutrino annihilations. The projected sensitivity of these experiments covers a part of the parameter regions (masses and interaction cross-sections) which will be explored by future particle accelerators such as LHC or Tevatron II. However their relative low cost make them very promising alternatives for finding new physics.

5 The Cosmic Microwave Background

The existence of a Cosmic Microwave Background with present temperature around 5K, was theoretically predicted in 1948 by G. Gamow, R. Alpher and R. Hermann, as a necessary relic of a hot phase of the Universe in which the light elements should have been cooked through nuclear reactions starting with primordial protons. This work largely ignored during almost two decades, can be regarded with hindsight as the foundational paper of the Hot Big Bang model. On the other hand, in year 1964, A. Penzias and W. Wilson who where working at Bell Telephone Laboratories to fit an antenna for satellite communications operating in the microwave range, found to his annoyance an “excess” radio noise isotropically distributed and corresponding to a temperature of about 3K, whose origin they did not know and did not hypothesized about. In the same year the group of theoreticians of Princeton University: B. Dicke, P. Peebles, P. Roll, and D. Wilkinson, who did not know or had forgotten about the work of Gamow et al, where following a similar line of reasoning. In fact they where thinking about building a radiometer to detect the fossil radiation left out from a primitive dense and hot phase of the Universe, when they knew about Penzias and Wilson observations and correctly interpreted them as the cosmic background radiation they were looking for. Similar considerations were also done by Y. Zeldovich and his group in Moscow around the same time. Thus CMB was accidentally discovered in 1964, and Penzias and Wilson (but not Gamow et al. nor Dicke et al.) were awarded the Nobel prize in 1978.

Since the temperature of the photons in CMB scales according to $T \propto a^{-1}$, the existence of CMB together with the expansion of the Universe imply a hot early phase, which becomes hotter and denser as we go backwards in time. In this primitive epoch, the Universe is a plasma containing more and more species of particles as we go backwards in time and new channels are opened for pair production to the increasingly energetic photons. In this way, when the photons are cold enough, $T \sim 3000$ K, the baryons and electrons *recombine* to form neutral hydrogen and helium atoms, and the photons are free to propagate: the Universe becomes transparent. This happens for a red-shift parameter $z_{\text{rec}} \sim 1100$, which corresponds to an age of about 380 Kyr. The recombination red-shift z_{rec} defines a *last scattering surface* for charged particles and photons, or cosmic photosphere, where CMB is coming from. Indeed, this is not a mathematical surface but it has a thickness which can be modeled by a Gaussian visibility function with a width $\delta z \simeq 80$.

CMB radiation is therefore a relic from $z_{\text{rec}} \sim 1100$, beyond which the Universe is optically thick in almost all wave bands, and it carries vital information about the processes and features of the early Universe [11]. So, since its discovery and specially in the last fifteen years there has been an intense observational effort on CMB. The COBE satellite launched in November 1989 made a major breakthrough. With the FIRAS (far infrared absolute spectrophotometer) instrument on

board, it was established the almost perfect blackbody spectrum of CMB, which witnesses the perfect thermal equilibrium state from which our Universe is coming from. It is most remarkable that such a pure blackbody spectrum has never been observed in laboratory experiments. In addition, with the help of the DMR (differential microwave radiometer) instrument, a map of CMB temperature anisotropies was obtained for the first time. After COBE a series of ground and balloon based measurements: ARCHEOPS, BOOMERANG, DASI, MAXIMA, VSA and for smaller scales CBI and ACBAR, have been carried out to improve the quality of temperature anisotropies distribution data. The most important recent advance has been the first year of operation results from NASA's WMAP (Wilkinson Microwave Anisotropy Probe) [11, 12]. Launched in June 2001, the first data release in February 2003, corresponds to a twofold full coverage of the sky and provides a much more precise anisotropies map than COBE's (see Fig. 3). WMAP is funded to operate for at least another three years, completing and eightfold covering of the sky and increasing statistical accuracy. After that, ESA's satellite Planck [13], scheduled for launch in 2007, will take over. As WMAP, Planck will be stationed at Lagrange L2 point of the Sun-Earth system, and will cover the full sky reaching an angular resolution up to one tenth of degree.

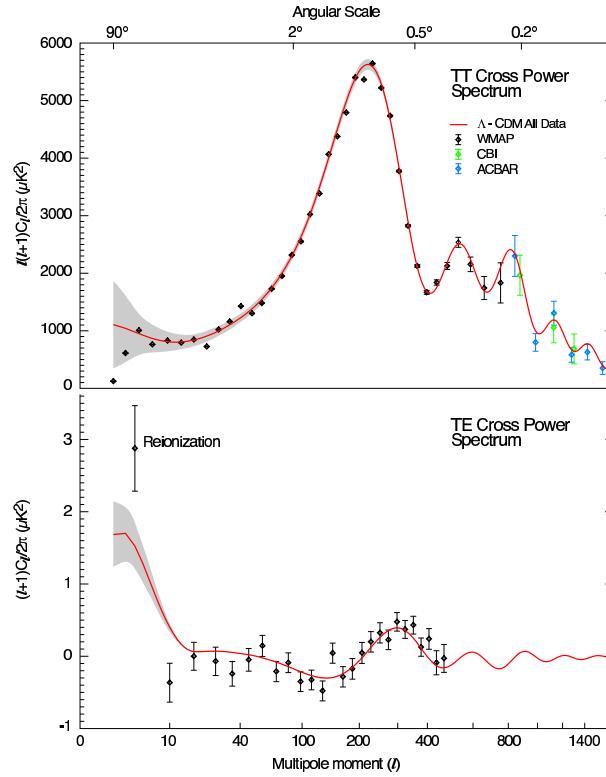


Fig. 3. TT and TE angular power spectra from WMAP one-year data, [11]

The CMB temperature distribution in the sky, being a function defined on a sphere is most naturally analyzed through an spherical harmonics expansion

$$T(\theta, \phi) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\theta, \phi) \quad (12)$$

The monopole component gives the mean temperature of CMB $T = 2.725 \pm 0.001$ K, which according to Planck's law for blackbody radiation, corresponds to a photon number density $n_\gamma = 411 \text{ cm}^{-3}$, and an energy density $\rho_\gamma = 0.260 \text{ eV cm}^{-3}$. The largest anisotropy is the $\ell = 1$ dipole term with amplitude 3.346 ± 0.017 mK interpreted as the result of the Doppler shift caused by the Solar system motion relative to the CMB. The implied velocity for the Sun is $v = 368 \pm 2 \text{ km s}^{-1}$, directed towards Hydra-Centaurus (more precisely, right ascension = 166 ± 3 and declination = -27.1 ± 3). This interpretation is reinforced by the yearly modulation of the anisotropy due to Earth's motion around the Sun, and also by measurements of the velocity field of local galaxies. This rather impressive finding about what could be called “our home's absolute velocity” embodies a delicious irony: the negative result of the Michelson-Morley experiment in 1887 in detecting Earth's motion with respect to aether (the hypothetical mechanical medium supporting the propagation of light), motivated the introduction of Special Relativity. On these foundations General Relativity was built and Cosmology developed during the XXth century. Finally, towards the end of the last century, the peculiar velocity of the Earth was measured with light (CMB) playing the role of aether, and without contradicting Relativity. Simply, CMB photons materialize a convenient comoving coordinate system.

Once the monopole and the dipole have been removed from the expansion (12), we are left with the CMB intrinsic anisotropies which are of the order of, or below 10^{-5} in all angular scales, and contain the imprints of the early Universe physics at radiation-matter decoupling. Most of the cosmological information is contained in the two point temperature-temperature (TT) correlation function. This quantity is defined by averaging the product of the fractional temperature deviations in directions \mathbf{n} and \mathbf{n}' over the sky, and expanding the result in Legendre polynomials

$$C(\theta) \equiv \langle \frac{\Delta T(\mathbf{n})}{T} \frac{\Delta T(\mathbf{n}')}{T} \rangle = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} C_l P_l(\cos \theta) \quad (13)$$

The expansion coefficients C_l , when represented as function of ℓ (more suitably $\log \ell$) give the so-called *angular power spectrum* which is the key function in comparing theory and observations. The cosmological parameters affect the form of this function and this is the way in which they can be deduced from CMB observations. Several physical mechanisms contribute to the angular power spectrum at different angular scales $\theta \sim \pi/\ell$. An important milestone is the size of the comoving Hubble radius at decoupling. This is half the size of the comoving particle horizon at that time if inflation had not happened before, and far from covering the whole sky, it subtends an angular opening $\theta \sim 0.86\sqrt{\Omega}$, with Ω being the present total density parameter of the Universe. This is the origin of the horizon problem that inflation solves as we will discuss below.

The description of the physics contained in C_l , can be separated into three main regions. i) *The Sachs-Wolfe plateau* for $\ell \leq 100$: This region corresponds to angular scales bigger than the Hubble radius at decoupling and it is dominated by the so-called Sachs-Wolfe effect: the photons coming from denser regions have

to climb out of deeper gravitational potential wells and become redder. A nearly scale invariant spectrum of density perturbations, as predicted by inflation, agrees with a plateau for this effect. Also, the integrated Sachs-Wolfe effect due to the time variation of the gravitational potential along CMB photons world lines, and gravity waves too, are expected to contribute in this large angular scale region, although these effects are buried in the *cosmic variance*. This means that doing an statistical analysis of fluctuations for a given set of cosmological parameters, but having only one realization of these fluctuations (our Universe) forces to introduce a kind of ergodic hypothesis: averages of patterns over parts of the sky, extended to the full sky by periodicity, would be equal to averages over different full sky realizations of perturbations for the same set of parameters. Indeed this method becomes increasingly uncertain at large angular scales and this is the meaning of the term cosmic variance.

ii) *The acoustic peaks* for $100 \leq \ell \leq 1000$. Before decoupling, photons, electrons and baryons form a tightly coupled fluid with the photons providing the pressure and the baryons the inertia. The fluid supports acoustic waves whose fundamental halfperiod is determined by the Hubble length at decoupling, and in turn its wavelength is obtained from the known value of the sound speed in the plasma. In this oscillating plasma denser regions correspond also to hotter regions (notice the opposite sign of this effect as compared to Sachs-Wolfe) according to Planck's law. Therefore a series of peaks in C_l are expected. The first one corresponding to the fundamental acoustic mode, and subsequent ones for higher harmonics. These peaks were theoretically predicted by P. Peebles, and Y. Zeldovich and collaborators already in 1970, and the empirical proof of their existence is a major success of modern cosmology. The position and height of the acoustic peaks encode information about the cosmological parameters. Indeed the angular scale subtended today by the fundamental acoustic wavelength depends on the underlying geometry. This is how the position of the first peak, established by WMAP to be around $\ell \sim 220$, results in a flat Universe with $\Omega = 1.02 \pm 0.02$. This way of establishing the flatness of our Universe is in fact very similar in essence to the method attributed to Gauss, who supposedly measured the three angles formed by three peaks in the Harz mountains, in order to check space geometry. In this case the triangle is the one formed by the acoustic fundamental halfwavelenght sitting on the last scattering surface, and two lines of sight from its extremes to us. Thus in a Λ CDM model, the position of the first acoustic peak determines $\Omega_M + \Omega_\Lambda$, while the difference $\Omega_M - \Omega_\Lambda$ can be extracted from SNe Ia Hubble diagrams as explained above. The second peak is not as high as the first one because the baryons feel the radiation pressure but cold dark matter does not, and as a consequence the relative height of the first and second peak gives the amount of baryons in the Universe. Combined results from WMAP, CBI and ACBAR yield $\Omega_B h^2 = 0.023 \pm 0.001$. In turn this result, together with the photon number density, fixes a very important cosmic number, namely the baryon/photon ratio or specific entropy of the Universe, only from CMB data. The resulting value is $\eta_{10} \equiv 10^{10} \eta = 6.14 \pm 0.25$. It is very remarkable, and a strong indication of the maturity of the SCM that this value is consistent with the determination of η via the physics of BBN when the Universe was three minutes old. In addition, the relative height of the first three peaks provides a determination of Ω_M . The Hubble parameter H_0 can be also extracted from the angular power spectrum, although the dependence of its shape with respect to H_0 is more involved. WMAP gives $h = 0.71(5\%)$, also in good agreement with Hubble diagrams.

iii) *The damping tail* for $\ell \geq 1000$. As stated above the transition to transparency is not instantaneous and the last scattering surface has a thickness. This leads to the so-called *Silk damping* of the anisotropies for angular scales smaller than this thickness. In addition, gravitational lensing by non-linear structures at low red-shift, like clusters of galaxies, also deform and smooth the primordial angular spectrum at small scales. As a consequence there are not much primordial anisotropies to observe below 5' of arc.

Electron-photon Thomson scattering at the last scattering surface transforms anisotropies into CMB photons polarization. The analysis of polarization leads to four new non-vanishing two sky points correlators, with their corresponding angular spectra. The theoretical and observational analysis of these spectra lies at the present frontier of CMB research. For example, Planck satellite is expected to do a significant advance in this respect. In particular the gravity wave contribution to CMB anisotropies could be observed. If the gravity wave perturbations were produced by inflation, these observations would determine the energy scale at which inflation happened. Also the WMAP results concerning polarization measurements have recently uncovered an earlier than expected reionization of the Universe at red-shift z around 20 (see the TE power spectrum in Figure 3). This means that the first stars in the Universe were formed as early as a few hundred million years after Big Bang. Such an early star formation could be challenging for the theory of inflation.

6 Big Bang Nucleosynthesis

While the CMB map of anisotropies can be considered as a photograph of the Universe when it was 380 kyr old, the observed abundances of the light nuclides ^1H , ^2D , ^3He , ^4He , and ^7Li represent the most ancient archaeological document about the history of the Universe [14]. These light elements were cooked in nuclear reactions when the Universe was about 3 minutes old. The remaining elements is the work of the stars with a little help of cosmic rays spallation. The first to propose a Big Bang nucleosynthesis was G. Gamow in 1946. In fact, a glance at the helium abundance: 24% in weight against 76% for hydrogen, tells us that so much helium can not have been produced by stars. For example, assuming that the age of the Milky Way is 10^{10} yr and that it has been radiating all the time at its current power $L_{\text{MW}} = 4 \times 10^{36}$ W, with all this power coming from the burning of hydrogen into helium, this will account only for less than 1% helium abundance.

The very early Universe is a too hostile environment for nuclei. When the temperature stays above a few MeV -the typical nucleon binding energy- the photons will immediately destroy any existing nuclei. So, nucleosynthesis has to wait until the Universe has cooled down enough. How much is enough? The first step for nucleosynthesis is the formation of deuterium through the reaction $n + p \rightarrow d + \gamma$, and the binding energy of deuterium is 2.22 MeV. However until the temperature does not reach below 0.1 MeV, formation of deuterium is not possible due to the high specific entropy of the Universe. Put in other words: since the photons outnumber the baryons by a factor 10^9 , even well below the binding energy of deuterium, there are enough hard photons in the high energy tail of the Planck distribution to destroy the deuterons as fast as they are produced. Once the deuterons are able to survive they almost instantaneously transform into helium through the reaction

$d + d \rightarrow {}^4\text{He} + \gamma$ and through other fusion reactions involving ${}^3\text{H}$ and ${}^3\text{He}$ as intermediate steps. Production of helium is very much favoured by its comparatively high binding energy 28.3 MeV. So, from energetic considerations only, it could have happened earlier, but it has to wait until deuterium is formed. This effect is called the *deuterium bottleneck*. Finally some ${}^7\text{Li}$ seven is also formed in collisions of ${}^4\text{He}$ with ${}^3\text{He}$ and ${}^3\text{H}$ nuclei. Why not higher nuclei? The reason is that there are not stable nuclei with $A = 5$ and $A = 8$, and only minute quantities of the nuclei with $A = 2, 3$ are formed in the synthesis of helium. In addition, the Universe is cooling down very fast, and Coulomb barriers which are higher for higher nuclei suppress nuclear reactions. Therefore BBN stops at this point. So, how do the stars manage to form the rest of the elements like C , N , O , which we the observers are made of? The answer was given by F. Hoyle. There exists a metastable resonance of two ${}^4\text{He}$ nuclei. Then, if a third ${}^4\text{He}$ nucleus meets the resonance, a ${}^{12}\text{C}$ nucleus is formed by a two steps chain of two particle collisions. This is possible if the temperature and the density are both very high, but in the early Universe, the density and temperature continually drop, and by the time helium has been synthesized, it is too late for this way of producing carbon. However in the interior of stars temperature and density steadily rise as the star evolves, and eventually, the physical conditions for the transformation of helium into carbon -and then into heavier elements- are attained.

The nuclear and elementary particle physics needed to study BBN is well known, and the temporal dependence of the density and temperature can be derived from Friedmann equations. Therefore, the light elements cosmic abundances can be theoretically calculated and compared to the observed ones. The synthesis of light elements is sensitive to physical conditions for temperatures $T \lesssim 1$ MeV, corresponding to an age $t \gtrsim 1$ s. Above this temperature neutrons and protons are in thermal equilibrium through weak reactions like $e^- + p \leftrightarrow \nu_e + n$ and $\bar{\nu}_e + p \leftrightarrow e^+ + n$, and also similar ones for the other neutrino and lepton families. Thus the neutron abundance of protons and neutrons is fixed by the Boltzmann factor $e^{-Q/T}$, where $Q = 1.293$ MeV is the neutron-proton mass difference. Thus, as long as thermal equilibrium is maintained, the Universe is running out of neutrons very fast as it cools down. Indeed, if thermal equilibrium held until the deuterium bottleneck is surpassed, very few neutrons would survive. However this is not so because before that, weak interactions “freeze-out” and neutrinos go out of thermal equilibrium. This happens because weak interaction cross sections scale with temperature as T^2 and non-relativistic particle densities as T^3 , while Hubble parameter H scales as T^2 . Therefore the weak reaction rates Γ_w over the expansion rate H scale as $\Gamma_w/H \propto T^3$. Thus, eventually at some temperature, which detailed calculations show to be $T \simeq 1$ MeV, the neutron-proton interconversion go out of thermal equilibrium. At this point, the neutron to proton ratio is about 1/6. From this point on, occasional weak interactions with the tails of the lepton and nucleon Fermi distributions, and neutron beta decay still lower (although much more slowly) the neutron to proton ratio. When the deuterium bottleneck is surpassed at $T = 0.1$ MeV, the Universe is about 2 min old, and the neutron to proton ratio has fallen down to $\simeq 1/7$. Therefore, since practically all deuterium transforms into helium, the helium abundance today should be about 25% in weight.

The observed cosmological abundance for primordial ${}^4\text{He}$ are in the range 23-24%, in good agreement with the theoretical calculation based in BBN. The observed primordial abundances for deuterium and ${}^7\text{Li}$ can be estimated to be in the ranges

$D/H = 1 - 7 \times 10^{-5}$, and ${}^7\text{Li}/H = 0.59 - 4.1 \times 10^{-10}$, while for ${}^3\text{He}$ there is not a good estimation, which renders ${}^3\text{He}$ unsuitable as a cosmological probe. In BBN, the light elements abundances depend on the balance between the expansion rate H of the Universe during nucleosynthesis, and the nuclear reactions rates, which in turn depend on the baryon density. All light elements abundances can be explained with a baryon density given by a baryon to photon ratio η_{10} in the range 3.4-6.9 (95% CL). This value agrees remarkably well with the value for the same parameter obtained from CMB acoustic peaks, and provides another key confirmation of the BBN theory and of the SCM.

Since the reactions building ${}^4\text{He}$ are so rapid, ${}^4\text{He}$ primordial abundance depends mainly on the neutron availability when the required temperature to surpass the deuterium bottleneck is attained, and it is rather insensitive to baryon density. In turn, neutron availability at deuterium bottleneck depends on how long it takes to reach this point, i. e. on the expansion rate of the Universe. Therefore, primordial ${}^4\text{He}$ act as a *chronometer*. On the other hand, the other light elements relic abundances depend mainly on the nucleon density and act as a *baryometer*. In particular, primordial deuterium abundance is best known and depends sensitively on η . Therefore, deuterium abundance is the baryometer of choice.

The impressive agreement in the determination of the baryon content of the Universe by means of two totally independent sources of information: the acoustic peaks in the angular spectrum of CMB anisotropies, and BBN, is, beyond all doubts, a sign of the maturity of the SCM. However, a very important question remains unanswered: Why the photon to baryon ratio is about 10^9 , or why are there any baryons at all? We briefly address this question in the following subsection.

6.1 Baryogenesis

Since there is a nucleon for approximately each 10^9 photons, but almost no antinucleons, the observable Universe seems to have a net baryonic number (number of baryons minus the number of antibaryons). In addition, when the temperature of the Universe falls below 1 MeV, electrons and positrons annihilate into photons but not wholly. A small fraction of electrons in excess survive to exactly balance the charge of the protons. Thus, there is a matter-antimatter asymmetry in the Universe, which is very tiny but very important for us (otherwise we would not exist). The exact origin of this asymmetry is still unknown. A. Sakharov formulated three necessary conditions that must be fulfilled in order to generate the matter-antimatter asymmetry. i) There must exist *CP violating processes*, that distinguish particle from antiparticle interactions. Such kind of processes are known to exist in the EPSM, due to the mixings between the three families of quarks and leptons, and have been observed in the neutral kaons system. ii) There should exist *baryon number violating processes* that generate a net baryonic number. Such processes exist in Grand Unified Theories, and also as a non-perturbative effect in the minimal standard model. iii) There should exist *deviations from thermal equilibrium*. For if thermal equilibrium was always maintained, since particles and antiparticles have the same mass, their abundances would be always exactly the same. Much work has been done over the last two decades in building models that meet the Sakharov criteria, and predict the right amount of baryons in the Universe (the value of η). However, there is still not any conclusive explanation of baryogenesis at present [15].

7 Inflation

Despite the success of classical standard Cosmology in explaining the expansion of the Universe, the abundances of light elements, and the existence of a highly isotropic cosmic microwave background; this theoretical framework exhibits important limitations which we will discuss in this section.

On one hand, the assumed initial conditions for the evolution of the homogeneous FRW background are problematic. Thus, observations favor a universe with flat or almost flat spatial sections. However, such an universe is an unstable solution of the cosmological evolution equations. This implies that only a very small set of initial conditions could evolve into the presently observable Universe. This is usually referred to as the flatness problem. More important is the so-called horizon problem. Since the Universe had an origin in time, the maximum distance that light can travel from the Big Bang until a given time (particle horizon) is finite. This is also the maximum size that a causally connected region can have at that time. However, when comparing the size of the particle horizon at matter-radiation decoupling with the physical size of the presently observable Universe at that time, we find that the latter was much larger than the horizon size. This implies that not all of our observable Universe was inside a single causally connected region, and therefore there is no reason to expect that background radiation photons coming from different regions in the sky were at the same temperature. However, observations confirm that this is certainly the case, since the temperature anisotropies are extremely small. It is important to emphasize that these two problems arise because we are assuming that the evolution of the Universe is the standard one all the way down to the initial singularity. However, General Relativity is a classical field theory which is expected to break down at very short distances where quantum effects would dominate.

On the other hand, as explained above, the formation of large scale structures such as galaxies or galaxy clusters, is understood within the SCM as the amplification of initially small density perturbations, due to Jeans instability. However, although it is possible to determine the evolution of such perturbations within classical Cosmology, the model does not provide a mechanism for the generation of the primordial seeds, which are considered as an additional input. Indeed, there is a general argument which suggests that fluctuations generated within the horizon size at an early epoch by some mechanism (thermal fluctuations, ...) cannot be responsible for the observed structure at all scales. The argument reads as follows: consider a process which took place before nucleosynthesis when $T \gtrsim 100$ MeV. The mass within the horizon at that time was around $m_H \lesssim M_\odot$. Consider also large density fluctuations of order $\delta_H = (\delta\rho/\rho)|_H \sim 1$ on those scales. As larger and larger scales M enter the horizon, the dispersion of the mass fluctuations will decrease as $N^{-1/2}$ where $N = M/m_H$ is the number of small regions contained in the large one. Thus, the typical size of fluctuations on horizon scales at the time when a galactic scale with $M_{gal} \sim 10^{12} M_\odot$ entered the horizon, would be $\delta_{gal} \sim \delta_H N^{-1/2} \lesssim 10^{-6}$, which is too small to explain the present galactic density contrast. Thus, the existence of structures and temperature anisotropies at large scales is difficult to explain by causal phenomena in classical Cosmology and suggests the presence of perturbations on *super-horizon* scales, generated by some *exotic* mechanism (such as inflation).

In fact, if we insist on solving these problems ignoring possible quantum gravitational effects near the Big Bang singularity, then inflation can do the job [16]. Although its original motivation was to get rid of the overproduction of supermas-

sive relics (monopoles) which were predicted by certain models of the very early Universe, it was soon realized that inflation also provided a natural solution for the flatness and horizon problems.

Inflation is a short phase of accelerated expansion in the very early Universe. If inflation lasts for a sufficiently long period, i.e. for a large enough number of e-folds: $N_e = \ln(a_f/a_i)$, where $a_{f(i)}$ denotes the scale factor at the end (beginning) of inflation, ($N_e \gtrsim 50 - 70$ depending on the model); then it can be shown that the previously mentioned problems are automatically solved. Thus, the typical exponential growth of the scale factor during inflation makes the spatial curvature of the Universe to decline dramatically. In addition, during inflation, the physical size of the particle horizon grows at a similar rate as the physical distances, and as a consequence the presently visible patch of the Universe was at all times well inside the causally connected region.

Apart from the debatable importance of inflation as a solution for the flatness and horizon problems, its major success was the prediction of a (nearly) scale invariant spectrum of density fluctuations on super-horizon scales, in agreement with observations. In fact, during inflation, quantum fluctuations with sub-horizon physical wavelengths ($\lambda \ll H_I^{-1} \sim (10^{13} \text{ GeV})^{-1}$ in typical models) can be stretched by the Universe expansion up to scales comparable to the size of galaxies, galaxy clusters (kpc-Mpc) or even larger, at the present epoch. The amplitude of the quantum fluctuations of any *light* scalar field at horizon crossing is determined by the Hubble parameter H_I . Then, since this parameter is typically almost constant during inflation, a generic prediction of any inflationary model, is the mentioned flat spectrum of perturbations. Such form for the spectrum had been postulated many years before by Harrison and Zeldovich in order to explain galaxy formation. In addition to acting as seeds for structure formation as explained above, these perturbations are also responsible for the generation of anisotropies in the background radiation through the already mentioned Sachs-Wolfe effect.

It can be seen that those large scales which became larger than the Hubble radius (exit the horizon) at the beginning of inflation, re-entered the Hubble radius later, whereas the shorter wavelengths re-entered sooner, following a LOFI (last out, first in) scheme. Linear perturbation theory shows that once a given scale has entered the horizon, the corresponding density fluctuation can grow linearly with the scale factor a in the matter dominated era, whereas the growth is only logarithmic in the radiation dominated one. Thus, the scale $\lambda_{eq} \simeq 13(\Omega_M h^2)^{-1}$ Mpc corresponding to the size of the Hubble horizon at matter-radiation equality separates out the two behaviours. Therefore, the perturbations with shorter wavelengths have had more time to grow since reentering than the larger ones. So, there are definite predictions of inflation for structure formation and CMB anisotropies. Since the primordial spectrum is flat, i.e. perturbations have the same amplitude at all scales, the suppression factor (ignoring possible non-linear effects) in the density contrast will be given by $(a_{eq}/a) = (\lambda_{eq}/\lambda)^2$ for $\lambda > \lambda_{eq}$, whereas it will be only logarithmic for shorter wavelengths. The predictions agree reasonably well with observations (see Fig. 1). In addition, the presence of the so called Sachs-Wolfe plateau in the large angular scales region of the CMB power spectrum (see Fig. 3), corresponding to wavelengths of thousands of Mpc, can be traced back also to the inflationary prediction.

Although inflation is at present the only viable scenario of the early Universe, unfortunately it is not a complete theory. The mechanism responsible for the accelerated expansion is unknown. Most of the inflationary models are based on the

existence of a hypothetical scalar field called *inflaton* (either fundamental or effective) whose potential energy density dominates at early times, acting as an effective cosmological constant. However, only extensions of the Standard Model of elementary particles, such as supersymmetry, supergravity or string theory could naturally accommodate a scalar field with the required properties. These models are not free from difficulties either, since the required smallness of the slow-roll parameters for the inflaton potential, which is needed to fit observations, can only be maintained in particular cases.

The increasing observational precision in CMB anisotropies and large scale structure has allowed to improve the constraints on inflationary models. In particular, the primordial curvature power spectrum predicted by inflation can be parametrized as:

$$P_R(k) = A_S^2 \left(\frac{k}{k_0} \right)^{n_s - 1} \quad (14)$$

where A_S is the amplitude of scalar metric perturbations, n_s is the spectral index and k_0 is the normalization scale. Observations seem to be compatible with a simple power-law, gaussian, adiabatic spectrum, although a small contribution from isocurvature modes cannot be excluded. The combined analysis of CMB data from WMAP satellite, ground-based detectors such as CBI and ACBAR, the 2dF galaxy redshift survey and Ly- α forest information, provides the following 68% C.L. results at the $k_0 = 0.05 \text{ Mpc}^{-1}$ scale:

$$\begin{aligned} A_S &= (4.1 - 5.0) \cdot 10^{-5} \\ n_s &= 0.90 - 0.96 \\ dn_s/d\ln k &= -(0.049 - 0.015) \end{aligned} \quad (15)$$

We see the agreement with the Harrison-Zeldovich prediction $n_s \simeq 1$, and the small running of n_s with the scale. Although at present, these results do not exclude any kind of inflationary model, some particular form of the inflaton potential could be disfavored in the near future when new data are available.

Apart from scalar perturbations, inflation also predicts the generation of a gravity wave background, characterized too by its power spectrum:

$$P_T(k) = A_T^2 \left(\frac{k}{k_0} \right)^{n_t} \quad (16)$$

The amplitude of tensors is usually compared with the scalar amplitude in the tensor-scalar ratio $r = A_T^2/A_S^2$. In single-field models of inflation, the tensor spectral index is related to r through the consistency condition $n_t = -r/8$, so that the number of independent parameters can be reduced to $(A_s, n_s, r, dn_s/d\ln k)$. Since no gravity wave mode has been detected yet, the previous combined analysis gives only the constraint: $r < 0.9$ at the 95 % C.L. (a very recent fit from SDSS+WMAP+Ly- α +SNIa data has obtained a better bound, $r < 0.45$ at the 95 % C.L. [17]) Things can improve if we take into account the effect of CMB polarization. Gravity waves produce *magnetic* components of polarization (B-modes) which are not produced by scalar perturbations. Planck satellite, with polarized detectors, is expected to measure r with error bars around 0.13.

The possibility of testing different inflationary models by future experiments will open a fascinating window to the physics of the early Universe. Indeed, quantum

fluctuations generated during inflation well inside the Hubble radius have typical wavelengths much smaller than those probed by current particle accelerators. In other words, inflation tells us that the CMB temperature anisotropies and the large scale structures that we observe today, are the signals of very high-energy physics in the sky.

The Cosmic Inventory

Hubble parameter	$h = 0.71^{+0.04}_{-0.03}$
Baryon density	$\Omega_B = 0.044 \pm 0.004$
Matter density	$\Omega_M = 0.27 \pm 0.04$
Dark energy density	$\Omega_A = 0.73 \pm 0.04$
Total energy density	$\Omega_{tot} = 1.02 \pm 0.02$
Neutrino density	$\Omega_\nu h^2 \leq 0.0076$ (95 % C.L.)
Dark energy equation of state	$w < -0.78$ (95 % C.L.)

Table 1. Cosmological parameters with 68% C.L. intervals (except as otherwise stated). Results from the combined fit of WMAP, CBI, ACBAR, 2dF and Ly- α

Acknowledgements: This work has been partially supported by the DGICYT (Spain) under the project numbers FPA2000-0956 and BFM2002-01003.

References

1. S. Dodelson, *Modern Cosmology*, Academic Press, San Diego, California (2003)
2. J. D. Anderson and J. G. Williams. *Class. Quant. Grav.* **18**, 2447, (2001).
3. A. G. Riess et al. *Astron. J.* **116**, 1009, (1998). S. Perlmutter et al. *Astrophys. J.* **517**, 565, (1999).
4. 2dF Galaxy Redshift Survey home page:
<http://www.mso.anu.edu.au/2dFGRS/>
5. M. Tegmark et al., *Astrophys.J.* **606**, 702-740 (2004)
Sloan Digital Sky Survey home page. <http://www.sdss.org/sdds.html>
6. L. Pietronero, F. Sylos-Labini. ArXiv:astro-ph/0406202
7. W. L. Freedman, and M. Turner *Rev. Mod. Phys.* **75**, 1433, (2003).
8. A. G. Riess et al. arXiv:astro-ph/0402512
9. SNAP home page: <http://snap.lbl.gov/>
10. G. Bertone, D. Hooper an J. Silk, hep-ph/0404175, to appear in *Phys. Rep.*
11. D. N. Spergel et al. *Astrophys. J. Suppl.* **148**, 175, (2003).
12. WMAP home page: <http://map.gsfc.nasa.gov/index.html>

13. Planck home page: <http://www.rssd.esa.int/index.php?project=PLANCK>
14. For a review see K.A. Olive et al. *Phys. Rep.* **333**, 389 (2000)
15. M. Dine and A. Kusenko, *Rev. Mod. Phys.* **76**, 1 (2004)
16. A.R. Liddle and D.H. Lyth, *Cosmological Inflation and Large Scale Structure*, Cambridge University Press, (2000)
17. U. Seljak et al., arXiv:astro-ph/0407372

COSMOLOGY AND ASTROPHYSICS

J. García-Bellido

Departamento de Física Teórica, Universidad Autónoma de Madrid, Cantoblanco 28049 Madrid, Spain

Abstract

In these lectures I review the present status of the so-called Standard Cosmological Model, based on the hot Big Bang Theory and the Inflationary Paradigm. I will make special emphasis on the recent developments in observational cosmology, mainly the acceleration of the universe, the precise measurements of the microwave background anisotropies, and the formation of structure like galaxies and clusters of galaxies from tiny primordial fluctuations generated during inflation.

1. INTRODUCTION

The last five years have seen the coming of age of Modern Cosmology, a mature branch of science based on the hot Big Bang theory and the Inflationary Paradigm. In particular, we can now define rather precisely a Standard Model of Cosmology, where the basic parameters are determined within small uncertainties, of just a few percent, thanks to a host of experiments and observations. This precision era of cosmology has become possible thanks to important experimental developments in all fronts, from measurements of supernovae at high redshifts to the microwave background anisotropies, as well as to the distribution of matter in galaxies and clusters of galaxies.

In these lecture notes I will first introduce the basic concepts and equations associated with hot Big Bang cosmology, defining the main cosmological parameters and their corresponding relationships. Then I will address in detail the three fundamental observations that have shaped our present knowledge: the recent acceleration of the universe, the distribution of matter on large scales and the anisotropies in the microwave background. Together these observations allow the precise determination of a handful of cosmological parameters, in the context of the inflationary plus cold dark matter paradigm.

2. BIG BANG COSMOLOGY

Our present understanding of the universe is based upon the successful hot Big Bang theory, which explains its evolution from the first fraction of a second to our present age, around 13.6 billion years later. This theory rests upon four robust pillars, a theoretical framework based on general relativity, as put forward by Albert Einstein [1] and Alexander A. Friedmann [2] in the 1920s, and three basic observational facts: First, the expansion of the universe, discovered by Edwin P. Hubble [3] in the 1930s, as a recession of galaxies at a speed proportional to their distance from us. Second, the relative abundance of light elements, explained by George Gamow [4] in the 1940s, mainly that of helium, deuterium and lithium, which were cooked from the nuclear reactions that took place at around a second to a few minutes after the Big Bang, when the universe was a few times hotter than the core of the sun. Third, the cosmic microwave background (CMB), the afterglow of the Big Bang, discovered in 1965 by Arno A. Penzias and Robert W. Wilson [5] as a very isotropic blackbody radiation at a temperature of about 3 degrees Kelvin, emitted when the universe was cold enough to form neutral atoms, and photons decoupled from matter, approximately 380,000 years after the Big Bang. Today, these observations are confirmed to within a few percent accuracy, and have helped establish the hot Big Bang as the preferred model of the universe.

Modern Cosmology begun as a quantitative science with the advent of Einstein's general relativity and the realization that the geometry of space-time, and thus the general attraction of matter, is

determined by the energy content of the universe [6]

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}. \quad (1)$$

These non-linear equations are simply too difficult to solve without invoking some symmetries of the problem at hand: the universe itself.

We live on Earth, just 8 light-minutes away from our star, the Sun, which is orbiting at 8.5 kpc from the center of our galaxy,¹ the Milky Way, an ordinary galaxy within the Virgo cluster, of size a few Mpc, itself part of a supercluster of size a few 100 Mpc, within the visible universe, approximately 10,000 Mpc in size. Although at small scales the universe looks very inhomogeneous and anisotropic, the deepest galaxy catalogs like 2dF GRS and SDSS suggest that the universe on large scales (beyond the supercluster scales) is very homogeneous and isotropic. Moreover, the cosmic microwave background, which contains information about the early universe, indicates that the deviations from homogeneity and isotropy were just a few parts per million at the time of photon decoupling. Therefore, we can safely impose those symmetries to the universe at large and determine the corresponding evolution equations. The most general metric satisfying homogeneity and isotropy is the Friedmann-Robertson-Walker (FRW) metric, written here in terms of the invariant geodesic distance $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$ in four dimensions [6] $\mu = 0, 1, 2, 3$,²

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

characterized by just two quantities, a *scale factor* $a(t)$, which determines the physical size of the universe, and a constant K , which characterizes the *spatial curvature* of the universe,

$$(3)R = \frac{6K}{a^2(t)} \quad \begin{cases} K = -1 & \text{OPEN} \\ K = 0 & \text{FLAT} \\ K = +1 & \text{CLOSED} \end{cases} \quad (3)$$

Spatially open, flat and closed universes have different three-geometries. Light geodesics on these universes behave differently, and thus could in principle be distinguished observationally, as we shall discuss later. Apart from the three-dimensional spatial curvature, we can also compute a four-dimensional *space-time curvature*,

$$(4)R = 6\frac{\ddot{a}}{a} + 6\left(\frac{\dot{a}}{a}\right)^2 + 6\frac{K}{a^2}. \quad (4)$$

Depending on the dynamics (and thus on the matter/energy content) of the universe, we will have different possible outcomes of its evolution. The universe may expand for ever, recollapse in the future or approach an asymptotic state in between.

2.1 The matter and energy content of the universe

The most general matter fluid consistent with the assumption of homogeneity and isotropy is a perfect fluid, one in which an observer *comoving with the fluid* would see the universe around it as isotropic. The energy momentum tensor associated with such a fluid can be written as [6]

$$T^{\mu\nu} = p g^{\mu\nu} + (p + \rho) U^\mu U^\nu, \quad (5)$$

where $p(t)$ and $\rho(t)$ are the pressure and energy density of the fluid at a given time in the expansion, as measured by this comoving observer, and U^μ is the comoving four-velocity, satisfying $U^\mu U_\mu = -1$. For such a comoving observer, the matter content looks isotropic (in its rest frame),

$$T^\mu_\nu = \text{diag}(-\rho(t), p(t), p(t), p(t)). \quad (6)$$

¹One parallax second (1 pc), *parsec* for short, corresponds to a distance of about 3.26 light-years or 3.09×10^{18} cm.

²I am using $c = 1$ everywhere, unless specified, and a metric signature $(-, +, +, +)$.

The conservation of energy ($T^{\mu\nu}_{;\nu} = 0$), a direct consequence of the general covariance of the theory ($G^{\mu\nu}_{;\nu} = 0$), can be written in terms of the FRW metric and the perfect fluid tensor (5) as

$$\dot{\rho} + 3\frac{\dot{a}}{a}(p + \rho) = 0. \quad (7)$$

In order to find explicit solutions, one has to supplement the conservation equation with an *equation of state* relating the pressure and the density of the fluid, $p = p(\rho)$. The most relevant fluids in cosmology are barotropic, i.e. fluids whose pressure is linearly proportional to the density, $p = w\rho$, and therefore the speed of sound is constant in those fluids.

We will restrict ourselves in these lectures to three main types of barotropic fluids:

- *Radiation*, with equation of state $p_R = \rho_R/3$, associated with relativistic degrees of freedom (i.e. particles with temperatures much greater than their mass). In this case, the energy density of radiation decays as $\rho_R \sim a^{-4}$ with the expansion of the universe.
- *Matter*, with equation of state $p_M \simeq 0$, associated with nonrelativistic degrees of freedom (i.e. particles with temperatures much smaller than their mass). In this case, the energy density of matter decays as $\rho_M \sim a^{-3}$ with the expansion of the universe.
- *Vacuum energy*, with equation of state $p_V = -\rho_V$, associated with quantum vacuum fluctuations. In this case, the vacuum energy density remains constant with the expansion of the universe.

This is all we need in order to solve the Einstein equations. Let us now write the equations of motion of observers comoving with such a fluid in an expanding universe. According to general relativity, these equations can be deduced from the Einstein equations (1), by substituting the FRW metric (2) and the perfect fluid tensor (5). The $\mu = i$, $\nu = j$ component of the Einstein equations, together with the $\mu = 0$, $\nu = 0$ component constitute the so-called Friedmann equations,

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{K}{a^2}, \quad (8)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}. \quad (9)$$

These equations contain all the relevant dynamics, since the energy conservation equation (7) can be obtained from these.

2.2 The Cosmological Parameters

I will now define the most important cosmological parameters. Perhaps the best known is the *Hubble parameter* or rate of expansion today, $H_0 = \dot{a}/a(t_0)$. We can write the Hubble parameter in units of $100 \text{ km s}^{-1}\text{Mpc}^{-1}$, which can be used to estimate the order of magnitude for the present size and age of the universe,

$$H_0 \equiv 100 h \text{ km s}^{-1}\text{Mpc}^{-1}, \quad (10)$$

$$cH_0^{-1} = 3000 h^{-1} \text{ Mpc}, \quad (11)$$

$$H_0^{-1} = 9.773 h^{-1} \text{ Gyr}. \quad (12)$$

The parameter h was measured to be in the range $0.4 < h < 1$ for decades, and only in the last few years has it been found to lie within 4% of $h = 0.70$. I will discuss those recent measurements in the next Section.

Using the present rate of expansion, one can define a *critical* density ρ_c , that which corresponds to a flat universe,

$$\rho_c \equiv \frac{3H_0^2}{8\pi G} = 1.88 h^2 10^{-29} \text{ g/cm}^3 \quad (13)$$

$$= 2.77 h^{-1} 10^{11} M_\odot/(h^{-1} \text{ Mpc})^3 \quad (14)$$

$$= 11.26 h^2 \text{ protons/m}^3, \quad (15)$$

where $M_\odot = 1.989 \times 10^{33}$ g is a solar mass unit. The critical density ρ_c corresponds to approximately 6 protons per cubic meter, certainly a very dilute fluid!

In terms of the critical density it is possible to define the density parameter

$$\Omega_0 \equiv \frac{8\pi G}{3H_0^2} \rho(t_0) = \frac{\rho}{\rho_c}(t_0), \quad (16)$$

whose sign can be used to determine the spatial (three-)curvature. Closed universes ($K = +1$) have $\Omega_0 > 1$, flat universes ($K = 0$) have $\Omega_0 = 1$, and open universes ($K = -1$) have $\Omega_0 < 1$, no matter what are the individual components that sum up to the density parameter.

In particular, we can define the individual ratios $\Omega_i \equiv \rho_i/\rho_c$, for matter, radiation, cosmological constant and even curvature, today,

$$\Omega_M = \frac{8\pi G \rho_M}{3H_0^2} \quad \Omega_R = \frac{8\pi G \rho_R}{3H_0^2} \quad (17)$$

$$\Omega_\Lambda = \frac{\Lambda}{3H_0^2} \quad \Omega_K = -\frac{K}{a_0^2 H_0^2}. \quad (18)$$

For instance, we can evaluate today the radiation component Ω_R , corresponding to relativistic particles, from the density of microwave background photons, $\rho_{\text{CMB}} = \pi^2 k^4 T_{\text{CMB}}^4 / (15\hbar^3 c^3) = 4.5 \times 10^{-34}$ g/cm³, which gives $\Omega_{\text{CMB}} = 2.4 \times 10^{-5} h^{-2}$. Three approximately massless neutrinos would contribute a similar amount. Therefore, we can safely neglect the contribution of relativistic particles to the total density of the universe today, which is dominated either by non-relativistic particles (baryons, dark matter or massive neutrinos) or by a cosmological constant, and write the rate of expansion in terms of its value today, as

$$H^2(a) = H_0^2 \left(\Omega_R \frac{a_0^4}{a^4} + \Omega_M \frac{a_0^3}{a^3} + \Omega_\Lambda + \Omega_K \frac{a_0^2}{a^2} \right). \quad (19)$$

An interesting consequence of these definitions is that one can now write the Friedmann equation today, $a = a_0$, as a *cosmic sum rule*,

$$1 = \Omega_M + \Omega_\Lambda + \Omega_K, \quad (20)$$

where we have neglected Ω_R today. That is, in the context of a FRW universe, the total fraction of matter density, cosmological constant and spatial curvature today must add up to one. For instance, if we measure one of the three components, say the spatial curvature, we can deduce the sum of the other two.

Looking now at the second Friedmann equation (9), we can define another basic parameter, the *deceleration parameter*,

$$q_0 = -\frac{a \ddot{a}}{\dot{a}^2}(t_0) = \frac{4\pi G}{3H_0^2} [\rho(t_0) + 3p(t_0)], \quad (21)$$

defined so that it is positive for ordinary matter and radiation, expressing the fact that the universe expansion should slow down due to the gravitational attraction of matter. We can write this parameter using the definitions of the density parameter for known and unknown fluids (with density Ω_x and arbitrary equation of state w_x) as

$$q_0 = \Omega_R + \frac{1}{2}\Omega_M - \Omega_\Lambda + \frac{1}{2} \sum_x (1 + 3w_x) \Omega_x. \quad (22)$$

Uniform expansion corresponds to $q_0 = 0$ and requires a cancellation between the matter and vacuum energies. For matter domination, $q_0 > 0$, while for vacuum domination, $q_0 < 0$. As we will see in a moment, we are at present probing the time dependence of the deceleration parameter and can determine with some accuracy the moment at which the universe went from a decelerating phase, dominated by dark matter, into an acceleration phase at present, which seems to indicate the dominance of some kind of vacuum energy.

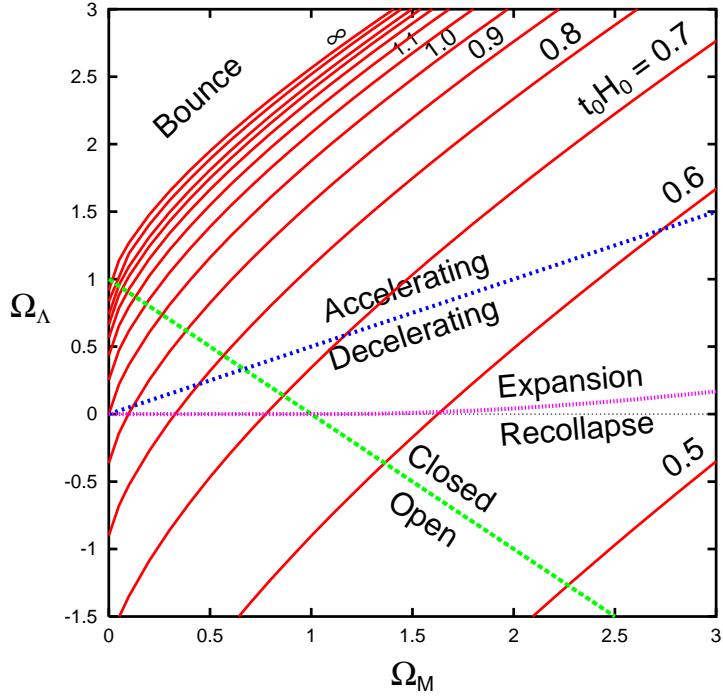


Fig. 1: Parameter space $(\Omega_M, \Omega_\Lambda)$. The green (dashed) line $\Omega_\Lambda = 1 - \Omega_M$ corresponds to a flat universe, $\Omega_K = 0$, separating open from closed universes. The blue (dotted) line $\Omega_\Lambda = \Omega_M/2$ corresponds to uniform expansion, $q_0 = 0$, separating accelerating from decelerating universes. The violet (dot-dashed) line corresponds to critical universes, separating eternal expansion from recollapse in the future. Finally, the red (continuous) lines correspond to $t_0 H_0 = 0.5, 0.6, \dots, \infty$, beyond which the universe has a bounce.

2.3 The $(\Omega_M, \Omega_\Lambda)$ plane

Now that we know that the universe is accelerating, one can parametrize the matter/energy content of the universe with just two components: the matter, characterized by Ω_M , and the vacuum energy Ω_Λ . Different values of these two parameters completely specify the universe evolution. It is thus natural to plot the results of observations in the plane $(\Omega_M, \Omega_\Lambda)$, in order to check whether we arrive at a consistent picture of the present universe from several different angles (different sets of cosmological observations).

Moreover, different regions of this plane specify different behaviors of the universe. The boundaries between regions are well defined curves that can be computed for a given model. I will now describe the various regions and boundaries.

- *Uniform expansion* ($q_0 = 0$). Corresponds to the line $\Omega_\Lambda = \Omega_M/2$. Points above this line correspond to universes that are accelerating today, while those below correspond to decelerating universes, in particular the old cosmological model of Einstein-de Sitter (EdS), with $\Omega_\Lambda = 0$, $\Omega_M = 1$. Since 1998, all the data from Supernovae of type Ia appear above this line, many standard deviations away from EdS universes.
- *Flat universe* ($\Omega_K = 0$). Corresponds to the line $\Omega_\Lambda = 1 - \Omega_M$. Points to the right of this line correspond to closed universes, while those to the left correspond to open ones. In the last few years we have mounting evidence that the universe is spatially flat (in fact Euclidean).
- *Bounce* ($t_0 H_0 = \infty$). Corresponds to a complicated function of $\Omega_\Lambda(\Omega_M)$, normally expressed as an integral equation, where

$$t_0 H_0 = \int_0^1 da [1 + \Omega_M(1/a - 1) + \Omega_\Lambda(a^2 - 1)]^{-1/2}$$

is the product of the age of the universe and the present rate of expansion. Points above this line correspond to universes that have contracted in the past and have later rebounded. At present, these universes are ruled out by observations of galaxies and quasars at high redshift (up to $z = 10$).

- *Critical Universe* ($H = \dot{H} = 0$). Corresponds to the boundary between eternal expansion in the future and recollapse. For $\Omega_M \leq 1$, it is simply the line $\Omega_\Lambda = 0$, but for $\Omega_M > 1$, it is a more complicated curve,

$$\Omega_\Lambda = 4\Omega_M \sin^3 \left[\frac{1}{3} \arcsin \left(\frac{\Omega_M - 1}{\Omega_M} \right) \right] \simeq \frac{4}{27} \frac{(\Omega_M - 1)^3}{\Omega_M^2}.$$

These critical solutions are asymptotic to the EdS model.

These boundaries, and the regions they delimit, can be seen in Fig. 1, together with the lines of equal $t_0 H_0$ values.

In summary, the basic cosmological parameters that are now been hunted by a host of cosmological observations are the following: the present rate of expansion H_0 ; the age of the universe t_0 ; the deceleration parameter q_0 ; the spatial curvature Ω_K ; the matter content Ω_M ; the vacuum energy Ω_Λ ; the baryon density Ω_B ; the neutrino density Ω_ν , and many other that characterize the perturbations responsible for the large scale structure (LSS) and the CMB anisotropies.

2.4 The accelerating universe

Let us first describe the effect that the expansion of the universe has on the objects that live in it. In the absence of other forces but those of gravity, the trajectory of a particle is given by general relativity in terms of the geodesic equation

$$\frac{du^\mu}{ds} + \Gamma_{\nu\lambda}^\mu u^\nu u^\lambda = 0, \quad (23)$$

where $u^\mu = (\gamma, \gamma v^i)$, with $\gamma^2 = 1 - v^2$ and v^i is the peculiar velocity. Here $\Gamma_{\nu\lambda}^\mu$ is the Christoffel connection [6], whose only non-zero component is $\Gamma_{ij}^0 = (\dot{a}/a) g_{ij}$; substituting into the geodesic equation, we obtain $|\vec{u}| \propto 1/a$, and thus the particle's momentum decays with the expansion like $p \propto 1/a$. In the case of a photon, satisfying the de Broglie relation $p = h/\lambda$, one obtains the well known *photon redshift*

$$\frac{\lambda_1}{\lambda_0} = \frac{a(t_1)}{a(t_0)} \Rightarrow z \equiv \frac{\lambda_0 - \lambda_1}{\lambda_1} = \frac{a_0}{a_1} - 1, \quad (24)$$

where λ_0 is the wavelength measured by an observer at time t_0 , while λ_1 is the wavelength emitted when the universe was younger ($t_1 < t_0$). Normally we measure light from stars in distant galaxies and compare their observed spectra with our laboratory (restframe) spectra. The fraction (24) then gives the redshift z of the object. We are assuming, of course, that both the emitted and the restframe spectra are identical, so that we can actually measure the effect of the intervening expansion, i.e. the growth of the scale factor from t_1 to t_0 , when we compare the two spectra. Note that if the emitting galaxy and our own participated in the expansion, i.e. if our measuring rods (our rulers) also expanded with the universe, we would see no effect! The reason we can measure the redshift of light from a distant galaxy is because our galaxy is a gravitationally bounded object that has decoupled from the expansion of the universe. It is the distance between galaxies that changes with time, not the sizes of galaxies, nor the local measuring rods.

We can now evaluate the relationship between physical distance and redshift as a function of the rate of expansion of the universe. Because of homogeneity we can always choose our position to be at the origin $r = 0$ of our spatial section. Imagine an object (a star) emitting light at time t_1 , at coordinate distance r_1 from the origin. Because of isotropy we can ignore the angular coordinates (θ, ϕ) . Then the physical distance, to first order, will be $d = a_0 r_1$. Since light travels along null geodesics [6], we can

write $0 = -dt^2 + a^2(t) dr^2/(1 - Kr^2)$, and therefore,

$$\int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_0^{r_1} \frac{dr}{\sqrt{1 - Kr^2}} \equiv f(r_1) = \begin{cases} \arcsin r_1 & K = 1 \\ r_1 & K = 0 \\ \operatorname{arcsinh} r_1 & K = -1 \end{cases} \quad (25)$$

If we now Taylor expand the scale factor to first order,

$$\frac{1}{1+z} = \frac{a(t)}{a_0} = 1 + H_0(t - t_0) + \mathcal{O}(t - t_0)^2, \quad (26)$$

we find, to first approximation,

$$r_1 \approx f(r_1) = \frac{1}{a_0}(t_0 - t_1) + \dots = \frac{z}{a_0 H_0} + \dots$$

Putting all together we find the famous Hubble law

$$H_0 d = a_0 H_0 r_1 = z \simeq vc, \quad (27)$$

which is just a kinematical effect (we have not included yet any dynamics, i.e. the matter content of the universe). Note that at low redshift ($z \ll 1$), one is tempted to associate the observed change in wavelength with a Doppler effect due to a hypothetical recession velocity of the distant galaxy. This is only an approximation. In fact, the redshift cannot be ascribed to the relative velocity of the distant galaxy because in general relativity (i.e. in curved spacetimes) one cannot compare velocities through parallel transport, since the value depends on the path! If the distance to the galaxy is small, i.e. $z \ll 1$, the physical spacetime is not very different from Minkowsky and such a comparison is approximately valid. As z becomes of order one, such a relation is manifestly false: galaxies cannot travel at speeds greater than the speed of light; it is the stretching of spacetime which is responsible for the observed redshift.

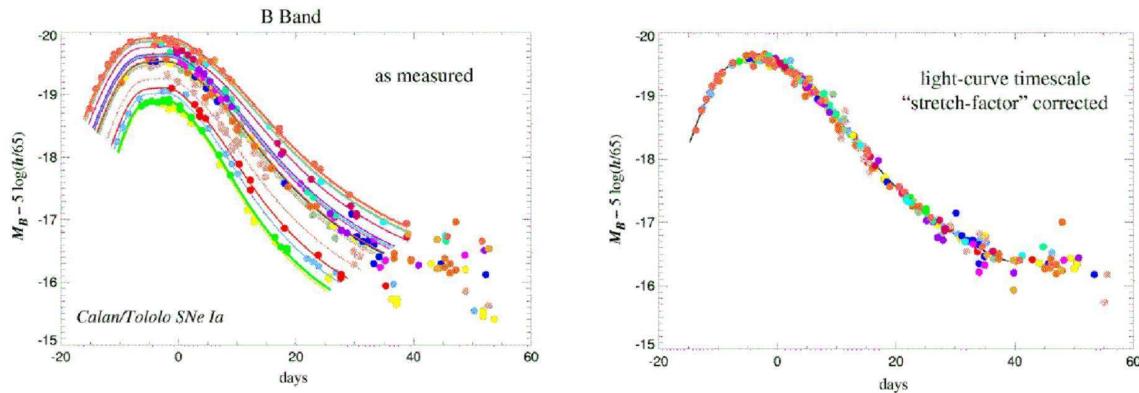


Fig. 2: The Type Ia supernovae observed nearby show a relationship between their absolute luminosity and the timescale of their light curve: the brighter supernovae are slower and the fainter ones are faster. A simple linear relation between the absolute magnitude and a ‘stretch factor’ multiplying the light curve timescale fits the data quite well. From Ref. [7].

Hubble’s law has been confirmed by observations ever since the 1920s, with increasing precision, which have allowed cosmologists to determine the Hubble parameter H_0 with less and less systematic errors. Nowadays, the best determination of the Hubble parameter was made by the Hubble Space Telescope Key Project [8], $H_0 = 72 \pm 8$ km/s/Mpc. This determination is based on objects at distances up to 500 Mpc, corresponding to redshifts $z \leq 0.1$.

Nowadays, we are beginning to probe much greater distances, corresponding to $z \simeq 1$, thanks to type Ia supernovae. These are white dwarf stars at the end of their life cycle that accrete matter from a companion until they become unstable and violently explode in a natural thermonuclear explosion that out-shines their progenitor galaxy. The intensity of the distant flash varies in time, it takes about three weeks to reach its maximum brightness and then it declines over a period of months. Although the maximum luminosity varies from one supernova to another, depending on their original mass, their environment, etc., there is a pattern: brighter explosions last longer than fainter ones. By studying the characteristic light curves, see Fig. 2, of a reasonably large statistical sample, cosmologists from the Supernova Cosmology Project [7] and the High-redshift Supernova Project [9], are now quite confident that they can use this type of supernova as a standard candle. Since the light coming from some of these rare explosions has travelled a large fraction of the size of the universe, one expects to be able to infer from their distribution the spatial curvature and the rate of expansion of the universe.

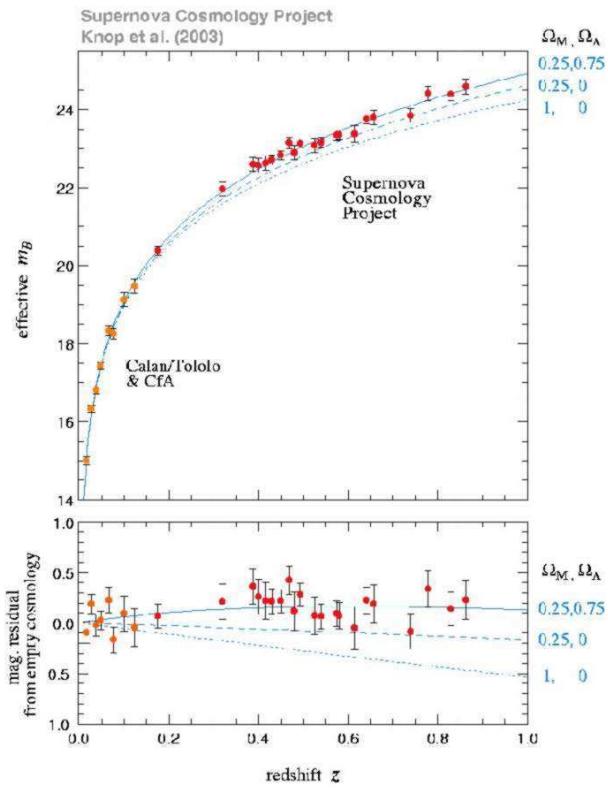


Fig. 3: Upper panel: The Hubble diagram in linear redshift scale. Supernovae with $\Delta z < 0.01$ of each other have been weighted-averaged binned. The solid curve represents the best-fit flat universe model, $(\Omega_M = 0.25, \Omega_\Lambda = 0.75)$. Two other cosmological models are shown for comparison, $(\Omega_M = 0.25, \Omega_\Lambda = 0)$ and $(\Omega_M = 1, \Omega_\Lambda = 0)$. Lower panel: Residuals of the averaged data relative to an empty universe. From Ref. [7].

The connection between observations of high redshift supernovae and cosmological parameters is done via the luminosity distance, defined as the distance d_L at which a source of absolute luminosity (energy emitted per unit time) \mathcal{L} gives a flux (measured energy per unit time and unit area of the detector) $\mathcal{F} = \mathcal{L}/4\pi d_L^2$. One can then evaluate, within a given cosmological model, the expression for d_L as a

function of redshift [10],

$$H_0 d_L(z) = \frac{(1+z)}{|\Omega_K|^{1/2}} \text{sinn} \left[\int_0^z \frac{|\Omega_K|^{1/2} dz'}{\sqrt{(1+z')^2(1+z'\Omega_M) - z'(2+z')\Omega_\Lambda}} \right], \quad (28)$$

where $\text{sinn}(x) = x$ if $K = 0$; $\sin(x)$ if $K = +1$ and $\sinh(x)$ if $K = -1$, and we have used the cosmic sum rule (20).

Astronomers measure the relative luminosity of a distant object in terms of what they call the effective magnitude, which has a peculiar relation with distance,

$$m(z) \equiv M + 5 \log_{10} \left[\frac{d_L(z)}{\text{Mpc}} \right] + 25 = \bar{M} + 5 \log_{10} [H_0 d_L(z)]. \quad (29)$$

Since 1998, several groups have obtained serious evidence that high redshift supernovae appear fainter than expected for either an open ($\Omega_M < 1$) or a flat ($\Omega_M = 1$) universe, see Fig. 3. In fact, the universe appears to be accelerating instead of decelerating, as was expected from the general attraction of matter, see Eq. (22); something seems to be acting as a repulsive force on very large scales. The most natural explanation for this is the presence of a cosmological constant, a diffuse vacuum energy that permeates all space and, as explained above, gives the universe an acceleration that tends to separate gravitationally bound systems from each other. The best-fit results from the Supernova Cosmology Project [11] give a linear combination

$$0.8 \Omega_M - 0.6 \Omega_\Lambda = -0.16 \pm 0.05 \quad (1\sigma),$$

which is now many sigma away from an EdS model with $\Lambda = 0$. In particular, for a flat universe this gives

$$\Omega_\Lambda = 0.71 \pm 0.05 \quad \text{and} \quad \Omega_M = 0.29 \pm 0.05 \quad (1\sigma).$$

Surprising as it may seem, arguments for a significant dark energy component of the universe were proposed long before these observations, in order to accommodate the ages of globular clusters, as well as a flat universe with a matter content below critical, which was needed in order to explain the observed distribution of galaxies, clusters and voids.

Taylor expanding the scale factor to third order,

$$\frac{a(t)}{a_0} = 1 + H_0(t-t_0) - \frac{q_0}{2!} H_0^2(t-t_0)^2 + \frac{j_0}{3!} H_0^3(t-t_0)^3 + \mathcal{O}(t-t_0)^4, \quad (30)$$

where

$$q_0 = -\frac{\ddot{a}}{aH^2}(t_0) = \frac{1}{2} \sum_i (1+3w_i)\Omega_i = \frac{1}{2} \Omega_M - \Omega_\Lambda, \quad (31)$$

$$j_0 = +\frac{\dddot{a}}{aH^3}(t_0) = \frac{1}{2} \sum_i (1+3w_i)(2+3w_i)\Omega_i = \Omega_M + \Omega_\Lambda, \quad (32)$$

are the deceleration and “jerk” parameters. Substituting into Eq. (28) we find

$$H_0 d_L(z) = z + \frac{1}{2}(1-q_0)z^2 - \frac{1}{6}(1-q_0-3q_0^2+j_0)z^3 + \mathcal{O}(z^4). \quad (33)$$

This expression goes beyond the leading linear term, corresponding to the Hubble law, into the second and third order terms, which are sensitive to the cosmological parameters Ω_M and Ω_Λ . It is only recently that cosmological observations have gone far enough back into the early universe that we can begin to probe these terms, see Fig. 4.

This extra component of the critical density would have to resist gravitational collapse, otherwise it would have been detected already as part of the energy in the halos of galaxies. However, if most of the

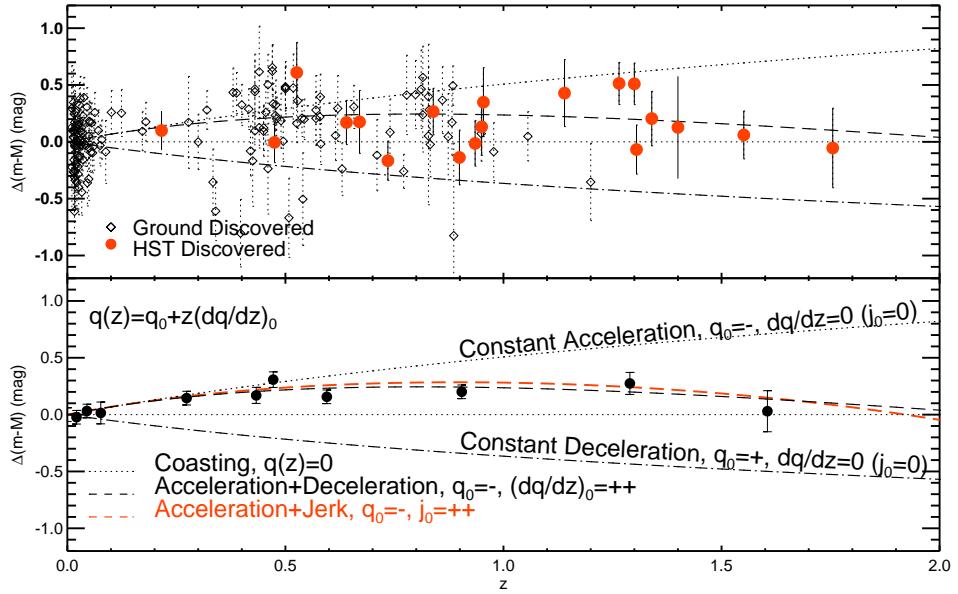


Fig. 4: The Supernovae Ia residual Hubble diagram. Upper panel: Ground-based discoveries are represented by diamonds, HST-discovered SNe Ia are shown as filled circles. Lower panel: The same but with weighted averaged in fixed redshift bins. Kinematic models of the expansion history are shown relative to an eternally coasting model $q(z) = 0$. From Ref. [12].

energy of the universe resists gravitational collapse, it is impossible for structure in the universe to grow. This dilemma can be resolved if the hypothetical dark energy was negligible in the past and only recently became the dominant component. According to general relativity, this requires that the dark energy have negative pressure, since the ratio of dark energy to matter density goes like $a(t)^{-3p/\rho}$. This argument would rule out almost all of the usual suspects, such as cold dark matter, neutrinos, radiation, and kinetic energy, since they all have zero or positive pressure. Thus, we expect something like a cosmological constant, with a negative pressure, $p \approx -\rho$, to account for the missing energy.

However, if the universe was dominated by dark matter in the past, in order to form structure, and only recently became dominated by dark energy, we must be able to see the effects of the transition from the deceleration into the acceleration phase in the luminosity of distant type Ia supernovae. This has been searched for since 1998, when the first convincing results on the present acceleration appeared. However, only recently [12] do we have clear evidence of this transition point in the evolution of the universe. This *coasting point* is defined as the time, or redshift, at which the deceleration parameter vanishes,

$$q(z) = -1 + (1 + z) \frac{d}{dz} \ln H(z) = 0, \quad (34)$$

where

$$H(z) = H_0 \left[\Omega_M (1 + z)^3 + \Omega_x e^{3 \int_0^z (1+w_x(z')) \frac{dz'}{1+z'}} + \Omega_K (1 + z)^2 \right]^{1/2}, \quad (35)$$

and we have assumed that the dark energy is parametrized by a density Ω_x today, with a redshift-dependent equation of state, $w_x(z)$, not necessarily equal to -1 . Of course, in the case of a true cosmological constant, this reduces to the usual expression.

Let us suppose for a moment that the barotropic parameter w is constant, then the coasting redshift can be determined from

$$q(z) = \frac{1}{2} \left[\frac{\Omega_M + (1 + 3w) \Omega_x (1 + z)^{3w}}{\Omega_M + \Omega_x (1 + z)^{3w} + \Omega_K (1 + z)^{-1}} \right] = 0, \quad (36)$$

$$\Rightarrow z_c = \left(\frac{(3|w|-1)\Omega_x}{\Omega_M} \right)^{\frac{1}{3|w|}} - 1, \quad (37)$$

which, in the case of a true cosmological constant, reduces to

$$z_c = \left(\frac{2\Omega_\Lambda}{\Omega_M} \right)^{1/3} - 1. \quad (38)$$

When substituting $\Omega_\Lambda \simeq 0.7$ and $\Omega_M \simeq 0.3$, one obtains $z_c \simeq 0.6$, in excellent agreement with recent observations [12]. The plane $(\Omega_M, \Omega_\Lambda)$ can be seen in Fig. 5, which shows a significant improvement with respect to previous data.

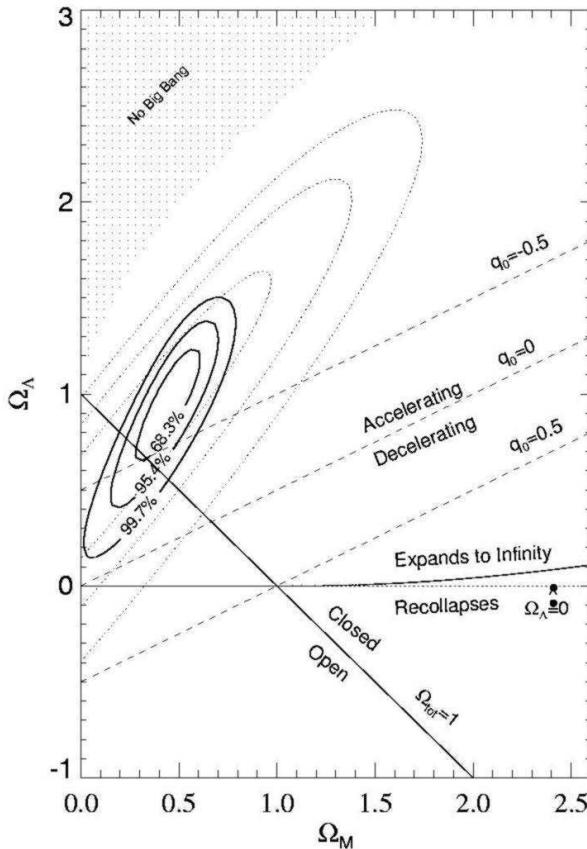


Fig. 5: The recent supernovae data on the $(\Omega_M, \Omega_\Lambda)$ plane. Shown are the 1-, 2- and 3- σ contours, as well as the data from 1998, for comparison. It is clear that the old EdS cosmological model at $(\Omega_M = 1, \Omega_\Lambda = 0)$ is many standard deviations away from the data. From Ref. [12].

Now, if we have to live with this vacuum energy, we might as well try to comprehend its origin. For the moment it is a complete mystery, perhaps the biggest mystery we have in physics today [13]. We measure its value but we don't understand why it has the value it has. In fact, if we naively predict it using the rules of quantum mechanics, we find a number that is many (many!) orders of magnitude off the mark. Let us describe this calculation in some detail. In non-gravitational physics, the zero-point energy of the system is irrelevant because forces arise from gradients of potential energies. However, we know from general relativity that even a constant energy density gravitates. Let us write down the most general energy momentum tensor compatible with the symmetries of the metric and that is covariantly conserved. This is precisely of the form $T_{\mu\nu}^{(vac)} = p_V g_{\mu\nu} = -\rho_V g_{\mu\nu}$, see Fig. 6. Substituting into

the Einstein equations (1), we see that the cosmological constant and the vacuum energy are completely equivalent, $\Lambda = 8\pi G \rho_V$, so we can measure the vacuum energy with the observations of the acceleration of the universe, which tells us that $\Omega_\Lambda \simeq 0.7$.

On the other hand, we can estimate the contribution to the vacuum energy coming from the quantum mechanical zero-point energy of the quantum oscillators associated with the fluctuations of all quantum fields,

$$\rho_V^{th} = \sum_i \int_0^{\Lambda_{UV}} \frac{d^2 k}{(2\pi)^3} \frac{1}{2} \hbar \omega_i(k) = \frac{\hbar \Lambda_{UV}^4}{16\pi^2} \sum_i (-1)^F N_i + \mathcal{O}(m_i^2 \Lambda_{UV}^2), \quad (39)$$

where Λ_{UV} is the ultraviolet cutoff signaling the scale of new physics. Taking the scale of quantum gravity, $\Lambda_{UV} = M_{Pl}$, as the cutoff, and barring any fortuitous cancellations, then the theoretical expectation (39) appears to be 120 orders of magnitude larger than the observed vacuum energy associated with the acceleration of the universe,

$$\rho_V^{th} \simeq 1.4 \times 10^{74} \text{ GeV}^4 = 3.2 \times 10^{91} \text{ g/cm}^3, \quad (40)$$

$$\rho_V^{obs} \simeq 0.7 \rho_c = 0.66 \times 10^{-29} \text{ g/cm}^3 = 2.9 \times 10^{-11} \text{ eV}^4. \quad (41)$$

Even if we assumed that the ultraviolet cutoff associated with quantum gravity was as low as the electroweak scale (and thus around the corner, liable to be explored in the LHC), the theoretical expectation would still be 60 orders of magnitude too big. This is by far the worst mismatch between theory and observations in all of science. There must be something seriously wrong in our present understanding of gravity at the most fundamental level. Perhaps we don't understand the vacuum and its energy does not gravitate after all, or perhaps we need to impose a new principle (or a symmetry) at the quantum gravity level to accommodate such a flagrant mismatch.

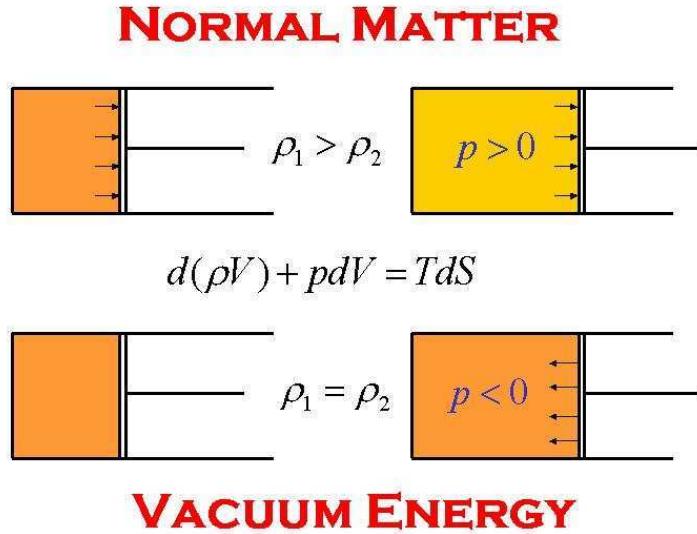


Fig. 6: Ordinary matter dilutes as it expands. According to the second law of Thermodynamics, its pressure on the walls should be positive, which exerts a force, and energy is lost in the expansion. On the other hand, vacuum energy is always the same, independent of the volume of the region, and thus, according to the second law, its pressure must be negative and of the same magnitude as the energy density. This negative pressure means that the volume tends to increase more and more rapidly, which explains the exponential expansion of the universe dominated by a cosmological constant.

In the meantime, one can at least parametrize our ignorance by making variations on the idea of a *constant* vacuum energy. Let us assume that it actually evolves slowly with time. In that case, we do not

expect the equation of state $p = -\rho$ to remain true, but instead we expect the barotropic parameter $w(z)$ to depend on redshift. Such phenomenological models have been proposed, and until recently produced results that were compatible with $w = -1$ today, but with enough uncertainty to speculate on alternatives to a truly constant vacuum energy. However, with the recent supernovae results [12], there seems to be little space for variations, and models of a time-dependent vacuum energy are less and less favoured. In the near future, the SNAP satellite [14] will measure several thousand supernovae at high redshift and therefore map the redshift dependence of both the dark energy density and its equation of state with great precision. This will allow a much better determination of the cosmological parameters Ω_M and Ω_Λ .

2.5 Thermodynamics of an expanding plasma

In this section I will describe the main concepts associated with ensembles of particles in thermal equilibrium and the brief periods in which the universe fell out of equilibrium. To begin with, let me make contact between the covariant energy conservation law (7) and the second law of thermodynamics,

$$T dS = dU + p dV, \quad (42)$$

where $U = \rho V$ is the total energy of the fluid, and $p = w \rho$ is its barotropic pressure. Taking a comoving volume for the universe, $V = a^3$, we find

$$T \frac{dS}{dt} = \frac{d}{dt}(\rho a^3) + p \frac{d}{dt}(a^3) = 0, \quad (43)$$

where we have used (7). Therefore, entropy is conserved during the expansion of the universe, $dS = 0$; i.e., the expansion is adiabatic even in those epochs in which the equation of state changes, like in the matter-radiation transition (not a proper phase transition). Using (7), we can write

$$\frac{d}{dt} \ln(\rho a^3) = -3H w. \quad (44)$$

Thus, our universe expands like a gaseous fluid in thermal equilibrium at a temperature T . This temperature decreases like that of any expanding fluid, in a way that is inversely proportional to the cubic root of the volume. This implies that in the past the universe was necessarily denser and hotter. As we go back in time we reach higher and higher temperatures, which implies that the mean energy of plasma particles is larger and thus certain fundamental reactions are now possible and even common, giving rise to processes that today we can only attain in particle physics accelerators. That is the reason why it is so important, for the study of early universe, to know the nature of the fundamental interactions at high energies, and the basic connection between cosmology and high energy particle physics. However, I should clarify a misleading statement that is often used: “high energy particle physics colliders reproduce the early universe” by inducing collisions among relativistic particles. Although the energies of some of the interactions at those collisions reach similar values as those attained in the early universe, the physical conditions are rather different. The interactions within the detectors of the great particle physics accelerators occur typically in the perturbative regime, locally, and very far from thermal equilibrium, lasting a minute fraction of a second; on the other hand, the same interactions occurred within a hot plasma in equilibrium in the early universe while it was expanding adiabatically and its duration could be significantly larger, with a distribution in energy that has nothing to do with those associated with particle accelerators. What is true, of course, is that the fundamental parameters corresponding to those interactions – masses and couplings – are assumed to be the same, and therefore present terrestrial experiments can help us imagine what it could have been like in the early universe, and make predictions about the evolution of the universe, in the context of an expanding plasma a high temperatures and high densities, and in thermal equilibrium.

2.51 Fluids in thermal equilibrium

In order to understand the thermodynamical behaviour of a plasma of different species of particles at high temperatures we will consider a gas of particles with g internal degrees of freedom weakly interacting. The degrees of freedom corresponding to the different particles can be seen in Table 1. For example, leptons and quarks have 4 degrees of freedom since they correspond to the two helicities for both particle and antiparticle. However, the nature of neutrinos is still unknown. If they happen to be Majorana fermions, then they would be their own antiparticle and the number of degrees of freedom would reduce to 2. For photons and gravitons (without mass) their 2 d.o.f. correspond to their states of polarization. The 8 gluons (also without mass) are the gauge bosons responsible for the strong interaction between quarks, and also have 2 d.o.f. each. The vector bosons W^\pm and Z^0 are massive and thus, apart from the transverse components of the polarization, they also have longitudinal components.

Particle	Spin	Degrees of freedom (g)	Nature
Higgs	0	1	Massive scalar
photon	1	2	Massless vector
graviton	2	2	Massless tensor
gluon	1	2	Massless vector
W y Z	1	3	Massive vector
leptons & quarks	1/2	4	Dirac Fermion
neutrinos	1/2	4 (2)	Dirac (Majorana) Fermion

Table 1: The internal degrees of freedom of various fundamental particles.

For each of these particles we can compute the number density n , the energy density ρ and the pressure p , in thermal equilibrium at a given temperature T ,

$$n = g \int \frac{d^3\mathbf{p}}{(2\pi)^3} f(\mathbf{p}), \quad (45)$$

$$\rho = g \int \frac{d^3\mathbf{p}}{(2\pi)^3} E(\mathbf{p}) f(\mathbf{p}), \quad (46)$$

$$p = g \int \frac{d^3\mathbf{p}}{(2\pi)^3} \frac{|\mathbf{p}|^2}{3E} f(\mathbf{p}), \quad (47)$$

where the energy is given by $E^2 = |\mathbf{p}|^2 + m^2$ and the momentum distribution in thermal (kinetic) equilibrium is

$$f(\mathbf{p}) = \frac{1}{e^{(E-\mu)/T} \pm 1} \quad \begin{cases} -1 & \text{Bose - Einstein} \\ +1 & \text{Fermi - Dirac} \end{cases} \quad (48)$$

The chemical potential μ is conserved in these reactions if they are in thermal equilibrium. For example, for reactions of the type $i + j \longleftrightarrow k + l$, we have $\mu_i + \mu_j = \mu_k + \mu_l$. For example, the chemical potential of the photon vanishes $\mu_\gamma = 0$, and thus particles and antiparticles have opposite chemical potentials.

From the equilibrium distributions one can obtain the number density n , the energy ρ and the pressure p , of a particle of mass m with chemical potential μ at the temperature T ,

$$n = \frac{g}{2\pi^2} \int_m^\infty dE \frac{E(E^2 - m^2)^{1/2}}{e^{(E-\mu)/T} \pm 1}, \quad (49)$$

$$\rho = \frac{g}{2\pi^2} \int_m^\infty dE \frac{E^2(E^2 - m^2)^{1/2}}{e^{(E-\mu)/T} \pm 1}, \quad (50)$$

$$p = \frac{g}{6\pi^2} \int_m^\infty dE \frac{(E^2 - m^2)^{3/2}}{e^{(E-\mu)/T} \pm 1}. \quad (51)$$

For a non-degenerate ($\mu \ll T$) relativistic gas ($m \ll T$), we find

$$n = \frac{g}{2\pi^2} \int_0^\infty \frac{E^2 dE}{e^{E/T} \pm 1} = \begin{cases} \frac{\zeta(3)}{\pi^2} g T^3 & \text{Bosons} \\ \frac{3}{4} \frac{\zeta(3)}{\pi^2} g T^3 & \text{Fermions} \end{cases}, \quad (52)$$

$$\rho = \frac{g}{2\pi^2} \int_0^\infty \frac{E^3 dE}{e^{E/T} \pm 1} = \begin{cases} \frac{\pi^2}{30} g T^4 & \text{Bosons} \\ \frac{7}{8} \frac{\pi^2}{30} g T^4 & \text{Fermions} \end{cases}, \quad (53)$$

$$p = \frac{1}{3} \rho, \quad (54)$$

where $\zeta(3) = 1.20206\dots$ is the Riemann Zeta function. For relativistic fluids, the energy density per particle is

$$\langle E \rangle \equiv \frac{\rho}{n} = \begin{cases} \frac{\pi^4}{30\zeta(3)} T \simeq 2.701 T & \text{Bosons} \\ \frac{7\pi^2}{180\zeta(3)} T \simeq 3.151 T & \text{Fermions} \end{cases} \quad (55)$$

For relativistic bosons or fermions with $\mu < 0$ and $|\mu| < T$, we have

$$n = \frac{g}{\pi^2} T^3 e^{\mu/T}, \quad (56)$$

$$\rho = \frac{3g}{\pi^2} T^4 e^{\mu/T}, \quad (57)$$

$$p = \frac{1}{3} \rho. \quad (58)$$

For a bosonic particle, a positive chemical potential, $\mu > 0$, indicates the presence of a Bose-Einstein condensate, and should be treated separately from the rest of the modes.

On the other hand, for a non-relativistic gas ($m \gg T$), with arbitrary chemical potential μ , we find

$$n = g \left(\frac{mT}{2\pi} \right)^{3/2} e^{-(m-\mu)/T}, \quad (59)$$

$$\rho = m n, \quad (60)$$

$$p = n T \ll \rho. \quad (61)$$

The average energy density per particle is

$$\langle E \rangle \equiv \frac{\rho}{n} = m + \frac{3}{2} T. \quad (62)$$

Note that, at any given temperature T , the contribution to the energy density of the universe coming from non-relativistic particles in thermal equilibrium is exponentially suppressed with respect to that of

relativistic particles, therefore we can write

$$\rho_R = \frac{\pi^2}{30} g_* T^4, \quad p_R = \frac{1}{3} \rho_R, \quad (63)$$

$$g_*(T) = \sum_{\text{bosons}} g_i \left(\frac{T_i}{T} \right)^4 + \frac{7}{8} \sum_{\text{fermions}} g_i \left(\frac{T_i}{T} \right)^4, \quad (64)$$

where the factor $7/8$ takes into account the difference between the Fermi and Bose statistics; g_* is the total number of light d.o.f. ($m \ll T$), and we have also considered the possibility that particle species i (bosons or fermions) have an equilibrium distribution at a temperature T_i different from that of photons, as happens for example when a given relativistic species decouples from the thermal bath, as we will discuss later. This number, g_* , strongly depends on the temperature of the universe, since as it expands and cools, different particles go out of equilibrium or become non-relativistic ($m \gg T$) and thus become exponentially suppressed from that moment on. A plot of the time evolution of $g_*(T)$ can be seen in Fig. 7.

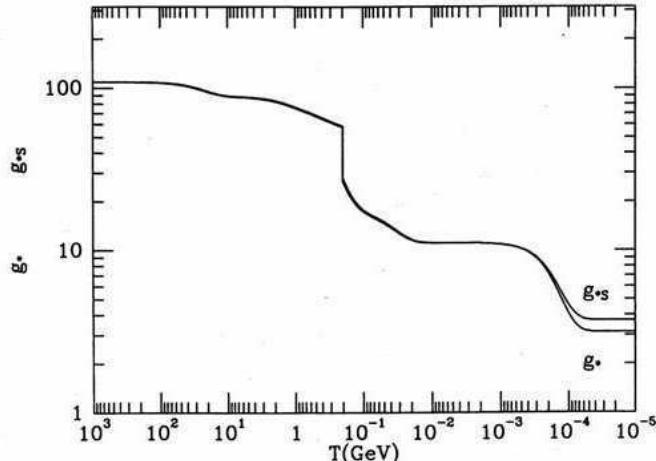


Fig. 7: the light degrees of freedom g_* and g_{*S} as a function of the temperature of the universe. From Ref. [15].

For example, for $T \ll 1$ MeV, i.e. after the time of primordial Big Bang Nucleosynthesis (BBN) and neutrino decoupling, the only relativistic species are the 3 light neutrinos and the photons; since the temperature of the neutrinos is $T_\nu = (4/11)^{1/3} T_\gamma = 1.90$ K, see below, we have $g_* = 2 + 3 \times \frac{7}{4} \times \left(\frac{4}{11}\right)^{4/3} = 3.36$, while $g_{*S} = 2 + 3 \times \frac{7}{4} \times \left(\frac{4}{11}\right) = 3.91$.

For $1 \text{ MeV} \ll T \ll 100 \text{ MeV}$, i.e. between BBN and the phase transition from a quark-gluon plasma to hadrons and mesons, we have, as relativistic species, apart from neutrinos and photons, also the electrons and positrons, so $g_* = 2 + 3 \times \frac{7}{4} + 2 \times \frac{7}{4} = 10.75$.

For $T \gg 250 \text{ GeV}$, i.e. above the electroweak (EW) symmetry breaking scale, we have one photon (2 polarizations), 8 gluons (massless), the W^\pm and Z^0 (massive), 3 families of quarks & leptones, a Higgs (still undiscovered), with which one finds $g_* = \frac{427}{4} = 106.75$.

At temperatures well above the electroweak transition we ignore the number of d.o.f. of particles, since we have never explored those energies in particle physics accelerators. Perhaps in the near future, with the results of the Large Hadron Collider (LHC) at CERN, we may predict the behaviour of the universe at those energy scales. For the moment we even ignore whether the universe was in thermal equilibrium at those temperatures. The highest energy scale at which we can safely say the universe

was in thermal equilibrium is that of BBN, i.e. 1 MeV, due to the fact that we observe the present relative abundances of the light element produced at that time. For instance, we can't even claim that the universe went through the quark-gluon phase transition, at ~ 200 MeV, since we have not observed yet any signature of such an event, not to mention the electroweak phase transition, at ~ 1 TeV.

Let us now use the relation between the rate of expansion and the temperature of relativistic particles to obtain the time scale of the universe as a function of its temperature,

$$H = 1.66 g_*^{1/2} \frac{T^2}{M_P} = \frac{1}{2t} \quad \Rightarrow \quad t = 0.301 g_*^{-1/2} \frac{M_P}{T^2} \sim \left(\frac{T}{\text{MeV}} \right)^{-2} \text{s}, \quad (65)$$

thus, e.g. at the EW scale (100 GeV) the universe was just 10^{-10} s old, while during the primordial BBN (1 – 0.1 MeV), it was 1 s to 3 min old.

2.52 The entropy of the universe

During most of the history of the universe, the rates of reaction, Γ_{int} , of particles in the thermal bath are much bigger than the rate of expansion of the universe, H , so that local thermal equilibrium was maintained. In this case, the entropy per comoving volume remained constant. In an expanding universe, the second law of thermodynamics, applied to the element of comoving volume, of unit coordinate volume and physical volume $V = a^3$, can be written as, see (42),

$$T dS = d(\rho V) + p dV = d[(\rho + p)V] - V dp. \quad (66)$$

Using the Maxwell condition of integrability, $\frac{\partial^2 S}{\partial T \partial V} = \frac{\partial^2 S}{\partial V \partial T}$, we find that $dp = (\rho + p)dT/T$, so that

$$dS = d \left[(\rho + p) \frac{V}{T} + \text{const.} \right], \quad (67)$$

i.e. the entropy in a comoving volume is $S = (\rho + p)\frac{V}{T}$, except for a constant. Using now the first law, the covariant conservation of energy, $T^{\mu\nu}_{;\nu} = 0$, we have

$$d[(\rho + p)V] = V dp \quad \Rightarrow \quad d \left((\rho + p) \frac{V}{T} \right) = 0, \quad (68)$$

and thus, in thermal equilibrium, the total entropy in a comoving volume, $S = a^3(\rho + p)/T$, is conserved. During most of the evolution of the universe, this entropy was dominated by the contribution from relativistic particles,

$$S = \frac{2\pi^2}{45} g_{*S} (aT)^3 = \text{const.}, \quad (69)$$

$$g_*(T) = \sum_{\text{bosons}} g_i \left(\frac{T_i}{T} \right)^3 + \frac{7}{8} \sum_{\text{fermions}} g_i \left(\frac{T_i}{T} \right)^3, \quad (70)$$

where g_{*S} is the number of “entropic” degrees of freedom, as we can see in Fig. 7. Above the electron-positron annihilation, all relativistic particles had the same temperature and thus $g_{*S} = g_*$. It may be also useful to realize that the entropy density, $s = S/a^3$, is proportional to the number density of relativistic particles, and in particular to the number density of photons, $s = 1.80 g_{*S} n_\gamma$; today, $s = 7.04 n_\gamma$. However, since g_{*S} in general is a function of temperature, we can't always interchange s and n_γ .

The conservation of S implies that the entropy density satisfies $s \propto a^{-3}$, and thus the physical size of the comoving volume is $a^3 \propto s^{-1}$; therefore, the number of particles of a given species in a comoving

volume, $N = a^3 n$, is proportional to the number density of that species over the entropy density s ,

$$N \sim \frac{n}{s} = \begin{cases} \frac{45\zeta(3)g}{2\pi^4 g_{*S}} & T \gg m, \mu \\ \frac{45g}{4\pi^5 \sqrt{2} g_{*S}} \left(\frac{m}{T}\right)^{3/2} e^{-\frac{m-\mu}{T}} & T \ll m \end{cases} \quad (71)$$

If this number does not change, i.e. if those particles are neither created nor destroyed, then n/s remains constant. As a useful example, we will consider the barionic number in a comoving volume,

$$\frac{n_B}{s} \equiv \frac{n_b - n_{\bar{b}}}{s}. \quad (72)$$

As long as the interactions that violate barion number occur sufficiently slowly, the barionic number per comoving volume, n_B/s , will remain constant. Although

$$\eta \equiv \frac{n_B}{n_\gamma} = 1.80 g_{*S} \frac{n_B}{s}, \quad (73)$$

the ratio between barion and photon numbers it does not remain constant during the whole evolution of the universe since g_{*S} varies; e.g. during the annihilation of electrons and positrons, the number of photons per comoving volume, $N_\gamma = a^3 n_\gamma$, grows a factor 11/4, and η decreases by the same factor. After this epoch, however, g_* is constant so that $\eta \simeq 7n_B/s$ and n_B/s can be used indistinctly.

Another consequence of Eq. (69) is that $S = \text{const.}$ implies that the temperature of the universe evolves as

$$T \propto g_{*S}^{-1/3} a^{-1}. \quad (74)$$

As long as g_{*S} remains constant, we recover the well known result that the universe cools as it expands according to $T \propto 1/a$. The factor $g_{*S}^{-1/3}$ appears because when a species becomes non-relativistic (when $T \leq m$), and effectively disappears from the energy density of the universe, its entropy is transferred to the rest of the relativistic particles in the plasma, making T decrease not as quickly as $1/a$, until g_{*S} again becomes constant.

From the observational fact that the universe expands today one can deduce that in the past it must have been hotter and denser, and that in the future it will be colder and more dilute. Since the ratio of scale factors is determined by the redshift parameter z , we can obtain (to very good approximation) the temperature of the universe in the past with

$$T = T_0 (1 + z). \quad (75)$$

This expression has been spectacularly confirmed thanks to the absorption spectra of distant quasars [16]. These spectra suggest that the radiation background was acting as a thermal bath for the molecules in the interstellar medium with a temperature of 9 K at a redshift $z \sim 2$, and thus that in the past the photon background was hotter than today. Furthermore, observations of the anisotropies in the microwave background confirm that the universe at a redshift $z = 1089$ had a temperature of 0.3 eV, in agreement with Eq. (75).

2.6 The thermal evolution of the universe

In a strict mathematical sense, it is impossible for the universe to have been always in thermal equilibrium since the FRW model does not have a timelike Killing vector. In practice, however, we can say that the universe has been most of its history very close to thermal equilibrium. Of course, those periods in which there were deviations from thermal equilibrium have been crucial for its evolution thereafter (e.g. baryogenesis, QCD transition, primordial nucleosynthesis, recombination, etc.); without these the universe today would be very different and probably we would not be here to tell the story.

The key to understand the thermal history of the universe is the comparison between the rates of interaction between particles (microphysics) and the rate of expansion of the universe (macrophysics). Ignoring for the moment the dependence of g_* on temperature, the rate of change of T is given directly by the rate of expansion, $\dot{T}/T = -H$. As long as the local interactions – necessary in order that the particle distribution function adjusts *adiabatically* to the change of temperature – are sufficiently fast compared with the rate of expansion of the universe, the latter will evolve as a succession of states very close to thermal equilibrium, with a temperature proportional to a^{-1} . If we evaluate the interaction rates as

$$\Gamma_{\text{int}} \equiv \langle n \sigma |v| \rangle, \quad (76)$$

where $n(t)$ is the number density of target particles, σ is the cross section on the interaction and v is the relative velocity of the reaction, all averaged on a thermal distribution; then a rule of thumb for ensuring that thermal equilibrium is maintained is

$$\Gamma_{\text{int}} \gtrsim H. \quad (77)$$

This criterium is understandable. Suppose, as often occurs, that the interaction rate in thermal equilibrium is $\Gamma_{\text{int}} \propto T^n$, with $n > 2$; then, the number of interactions of a particle after time t is

$$N_{\text{int}} = \int_t^\infty \Gamma_{\text{int}}(t') dt' = \frac{1}{n-2} \frac{\Gamma_{\text{int}}}{H}(t), \quad (78)$$

therefore the particle interacts less than once from the moment in which $\Gamma_{\text{int}} \approx H$. If $\Gamma_{\text{int}} \gtrsim H$, the species remains coupled to the thermal plasma. This doesn't mean that, necessarily, the particle is out of local thermal equilibrium, since we have seen already that relativistic particles that have decoupled retain their equilibrium distribution, only at a different temperature from that of the rest of the plasma.

In order to obtain an approximate description of the decoupling of a particle species in an expanding universe, let us consider two types of interaction:

i) interactions mediated by massless gauge bosons, like for example the photon. In this case, the cross section for particles with significant momentum transfer can be written as $\sigma \sim \alpha^2/T^2$, with $\alpha = g^2/4\pi$ the coupling constant of the interaction. Assuming local thermal equilibrium, $n(t) \sim T^3$ and thus the interaction rate becomes $\Gamma \sim n \sigma |v| \sim \alpha^2 T$. Therefore,

$$\frac{\Gamma}{H} \sim \alpha^2 \frac{M_P}{T}, \quad (79)$$

so that for temperatures of the universe $T \lesssim \alpha^2 M_P \sim 10^{16}$ GeV, the reactions are fast enough and the plasma is in equilibrium, while for $T \gtrsim 10^{16}$ GeV, reactions are too slow to maintain equilibrium and it is said that they are “frozen-out”. An important consequence of this result is that the universe could never have been in thermal equilibrium above the grand unification (GUT) scale.

ii) interactions mediated by massive gauge bosons, e.g. like the W^\pm and Z^0 , or those responsible for the GUT interactions, X and Y . We will generically call them X bosons. The cross section depends rather strongly on the temperature of the plasma,

$$\sigma \sim \begin{cases} G_X^2 T^2 & T \ll M_X \\ \frac{\alpha^2}{T^2} & T \gg M_X \end{cases} \quad (80)$$

where $G_X \sim \alpha/M_X^2$ is the effective coupling constant of the interaction at energies well below the mass of the vector boson, analogous to the Fermi constant of the electroweak interaction, $G_F = g^2/(4\sqrt{2}M_W^2)$ at tree level. Note that for $T \gg M_X$ we recover the result for massless bosons, so we will concentrate here on the other case. For $T \leq M_X$, the rate of thermal interactions is $\Gamma \sim n \sigma |v| \sim G_X^2 T^5$. Therefore,

$$\frac{\Gamma}{H} \sim G_X^2 M_P T^3, \quad (81)$$

such that at temperatures in the range

$$M_X \gtrsim T \gtrsim G_X^{-2/3} M_P^{-1/3} \sim \left(\frac{M_X}{100 \text{ GeV}} \right)^{4/3} \text{ MeV}, \quad (82)$$

reactions occur so fast that the plasma is in thermal equilibrium, while for $T \lesssim (M_X/100 \text{ GeV})^{4/3} \text{ MeV}$, those reactions are too slow for maintaining equilibrium and they effective freeze-out, see Eq. (78).

2.61 The decoupling of relativistic particles

Those relativistic particles that have decoupled from the thermal bath do not participate in the transfer of entropy when the temperature of the universe falls below the mass threshold of a given species $T \simeq m$; in fact, the temperature of the decoupled relativistic species falls as $T \propto 1/a$, as we will now show. Suppose that a relativistic particle is initially in local thermal equilibrium, and that it decouples at a temperature T_D and time t_D . The phase space distribution at the time of decoupling is given by the equilibrium distribution,

$$f(\mathbf{p}, t_D) = \frac{1}{e^{E/T_D} \pm 1}. \quad (83)$$

After decoupling, the energy of each massless particle suffers redshift, $E(t) = E_D (a_D/a(t))$. The number density of particles also decreases, $n(t) = n_D (a_D/a(t))^3$. Thus, the phase space distribution at a time $t > t_D$ is

$$f(\mathbf{p}, t) = \frac{d^3 n}{d^3 \mathbf{p}} = f\left(\mathbf{p} \frac{a}{a_D}, t_D\right) = \frac{1}{e^{E a / a_D T_D} \pm 1} = \frac{1}{e^{E/T} \pm 1}, \quad (84)$$

so that we conclude that the distribution function of a particle that has decoupled while being relativistic remains self-similar as the universe expands, with a temperature that decreases as

$$T = T_D \frac{a_D}{a} \propto a^{-1}, \quad (85)$$

and *not* as $g_{*S}^{-1/3} a^{-1}$, like the rest of the plasma in equilibrium (74).

2.62 The decoupling of non-relativistic particles

Those particles that decoupled from the thermal bath when they were non-relativistic ($m \gg T$) behave differently. Let us study the evolution of the distribution function of a non-relativistic particle that was in local thermal equilibrium at a time t_D , when the universe had a temperature T_D . The moment of each particle suffers redshift as the universe expands, $|\mathbf{p}| = |\mathbf{p}_D| (a_D/a)$, see Eq. (24). Therefore, their kinetic energy satisfies $E = E_D (a_D/a)^2$. On the other hand, the particle number density also varies, $n(t) = n_D (a_D/a(t))^3$, so that a decoupled non-relativistic particle will have an equilibrium distribution function characterized by a temperature

$$T = T_D \frac{a_D^2}{a^2} \propto a^{-2}, \quad (86)$$

and a chemical potential

$$\mu(t) = m + (\mu_D - m) \frac{T}{T_D}, \quad (87)$$

whose variation is precisely that which is needed for the number density of particle to decrease as a^{-3} .

In summary, a particle species that decouples from the thermal bath follows an equilibrium distribution function with a temperature that decreases like $T_R \propto a^{-1}$ for relativistic particles ($T_D \gg m$) or like $T_{NR} \propto a^{-2}$ for non-relativistic particles ($T_D \ll m$). On the other hand, for semi-relativistic particles ($T_D \sim m$), its phase space distribution *does not maintain* an equilibrium distribution function, and should be computed case by case.

2.63 Brief thermal history of the universe

I will briefly summarize here the thermal history of the universe, from the Planck era to the present. As we go back in time, the universe becomes hotter and hotter and thus the amount of energy available for particle interactions increases. As a consequence, the nature of interactions goes from those described at low energy by long range gravitational and electromagnetic physics, to atomic physics, nuclear physics, all the way to high energy physics at the electroweak scale, gran unification (perhaps), and finally quantum gravity. The last two are still uncertain since we do not have any experimental evidence for those ultra high energy phenomena, and perhaps Nature has followed a different path.

The way we know about the high energy interactions of matter is via particle accelerators, which are unravelling the details of those fundamental interactions as we increase in energy. However, one should bear in mind that the physical conditions that take place in our high energy colliders are very different from those that occurred in the early universe. These machines could never reproduce the conditions of density and pressure in the rapidly expanding thermal plasma of the early universe. Nevertheless, those experiments are crucial in understanding the nature and *rate* of the local fundamental interactions available at those energies. What interests cosmologists is the statistical and thermal properties that such a plasma should have, and the role that causal horizons play in the final outcome of the early universe expansion. For instance, of crucial importance is the time at which certain particles *decoupled* from the plasma, i.e. when their interactions were not quick enough compared with the expansion of the universe, and they were left out of equilibrium with the plasma.

One can trace the evolution of the universe from its origin till today. There is still some speculation about the physics that took place in the universe above the energy scales probed by present colliders. Nevertheless, the overall layout presented here is a plausible and hopefully testable proposal. According to the best accepted view, the universe must have originated at the Planck era (10^{19} GeV, 10^{-43} s) from a quantum gravity fluctuation. Needless to say, we don't have any experimental evidence for such a statement: Quantum gravity phenomena are still in the realm of physical speculation. However, it is plausible that a primordial era of cosmological *inflation* originated then. Its consequences will be discussed below. Soon after, the universe may have reached the Grand Unified Theories (GUT) era (10^{16} GeV, 10^{-35} s). Quantum fluctuations of the inflaton field most probably left their imprint then as tiny perturbations in an otherwise very homogenous patch of the universe. At the end of inflation, the huge energy density of the inflaton field was converted into particles, which soon thermalized and became the origin of the hot Big Bang as we know it. Such a process is called *reheating* of the universe. Since then, the universe became radiation dominated. It is probable (although by no means certain) that the asymmetry between matter and antimatter originated at the same time as the rest of the energy of the universe, from the decay of the inflaton. This process is known under the name of *baryogenesis* since baryons (mostly quarks at that time) must have originated then, from the leftovers of their annihilation with antibaryons. It is a matter of speculation whether baryogenesis could have occurred at energies as low as the electroweak scale (100 GeV, 10^{-10} s). Note that although particle physics experiments have reached energies as high as 100 GeV, we still do not have observational evidence that the universe actually went through the EW phase transition. If confirmed, baryogenesis would constitute another "window" into the early universe. As the universe cooled down, it may have gone through the quark-gluon phase transition (10^2 MeV, 10^{-5} s), when baryons (mainly protons and neutrons) formed from their constituent quarks.

The furthest window we have on the early universe at the moment is that of *primordial nucleosynthesis* ($1 - 0.1$ MeV, 1 s – 3 min), when protons and neutrons were cold enough that bound systems could form, giving rise to the lightest elements, soon after *neutrino decoupling*: It is the realm of nuclear physics. The observed relative abundances of light elements are in agreement with the predictions of the hot Big Bang theory. Immediately afterwards, electron-positron annihilation occurs (0.5 MeV, 1 min) and all their energy goes into photons. Much later, at about (1 eV, $\sim 10^5$ yr), matter and radiation have equal energy densities. Soon after, electrons become bound to nuclei to form atoms (0.3 eV, 3×10^5

yr), in a process known as *recombination*: It is the realm of atomic physics. Immediately after, photons decouple from the plasma, travelling freely since then. Those are the photons we observe as the cosmic microwave background. Much later ($\sim 1 - 10$ Gyr), the small inhomogeneities generated during inflation have grown, via gravitational collapse, to become galaxies, clusters of galaxies, and superclusters, characterizing the epoch of *structure formation*. It is the realm of long range gravitational physics, perhaps dominated by a vacuum energy in the form of a cosmological constant. Finally (3K, 13 Gyr), the Sun, the Earth, and biological life originated from previous generations of stars, and from a primordial soup of organic compounds, respectively.

I will now review some of the more robust features of the Hot Big Bang theory of which we have precise observational evidence.

2.64 Primordial nucleosynthesis and light element abundance

In this subsection I will briefly review Big Bang nucleosynthesis and give the present observational constraints on the amount of baryons in the universe. In 1920 Eddington suggested that the sun might derive its energy from the fusion of hydrogen into helium. The detailed reactions by which stars burn hydrogen were first laid out by Hans Bethe in 1939. Soon afterwards, in 1946, George Gamow realized that similar processes might have occurred also in the hot and dense early universe and gave rise to the first light elements [4]. These processes could take place when the universe had a temperature of around $T_{\text{NS}} \sim 1 - 0.1$ MeV, which is about 100 times the temperature in the core of the Sun, while the density is $\rho_{\text{NS}} = \frac{\pi^2}{30} g_* T_{\text{NS}}^4 \sim 82 \text{ g cm}^{-3}$, about the same density as the core of the Sun. Note, however, that although both processes are driven by identical thermonuclear reactions, the physical conditions in star and Big Bang nucleosynthesis are very different. In the former, gravitational collapse heats up the core of the star and reactions last for billions of years (except in supernova explosions, which last a few minutes and creates all the heavier elements beyond iron), while in the latter the universe expansion cools the hot and dense plasma in just a few minutes. Nevertheless, Gamow reasoned that, although the early period of cosmic expansion was much shorter than the lifetime of a star, there was a large number of free neutrons at that time, so that the lighter elements could be built up quickly by successive neutron captures, starting with the reaction $n + p \rightarrow D + \gamma$. The abundances of the light elements would then be correlated with their neutron capture cross sections, in rough agreement with observations [6, 17].

Nowadays, Big Bang nucleosynthesis (BBN) codes compute a chain of around 30 coupled nuclear reactions [18], to produce all the light elements up to beryllium-7.³ Only the first four or five elements can be computed with accuracy better than 1% and compared with cosmological observations. These light elements are H , 4He , D , 3He , 7Li , and perhaps also 6Li . Their observed relative abundance to hydrogen is $[1 : 0.25 : 3 \cdot 10^{-5} : 2 \cdot 10^{-5} : 2 \cdot 10^{-10}]$ with various errors, mainly systematic. The BBN codes calculate these abundances using the laboratory measured nuclear reaction rates, the decay rate of the neutron, the number of light neutrinos and the homogeneous FRW expansion of the universe, as a function of *only* one variable, the number density fraction of baryons to photons, $\eta \equiv n_B/n_\gamma$. In fact, the present observations are only consistent, see Fig. 8 and Ref. [17, 18, 19], with a very narrow range of values of

$$\eta_{10} \equiv 10^{10} \eta = 6.2 \pm 0.6. \quad (88)$$

Such a small value of η indicates that there is about one baryon per 10^9 photons in the universe today. Any acceptable theory of baryogenesis should account for such a small number. Furthermore, the present baryon fraction of the critical density can be calculated from η_{10} as

$$\Omega_B h^2 = 3.6271 \times 10^{-3} \eta_{10} = 0.0224 \pm 0.0022 \quad (95\% \text{ c.l.}) \quad (89)$$

Clearly, this number is well below closure density, so baryons cannot account for all the matter in the universe, as I shall discuss below.

³The rest of nuclei, up to iron (Fe), are produced in heavy stars, and beyond Fe in novae and supernovae explosions.

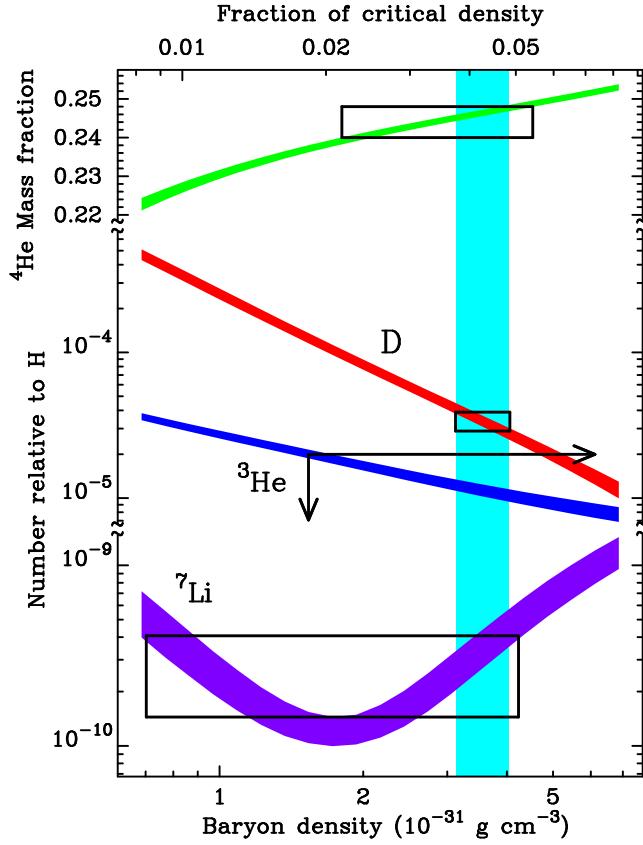


Fig. 8: The relative abundance of light elements to Hydrogen. Note the large range of scales involved. From Ref. [17].

2.65 Neutrino decoupling

Just before the nucleosynthesis of the lightest elements in the early universe, weak interactions were too slow to keep neutrinos in thermal equilibrium with the plasma, so they decoupled. We can estimate the temperature at which decoupling occurred from the weak interaction cross section, $\sigma_w \simeq G_F^2 T^2$ at finite temperature T , where $G_F = 1.2 \times 10^{-5}$ GeV $^{-2}$ is the Fermi constant. The neutrino interaction rate, via W boson exchange in $n + \nu \leftrightarrow p + e^-$ and $p + \bar{\nu} \leftrightarrow n + e^+$, can be written as [15]

$$\Gamma_\nu = n_\nu \langle \sigma_w |v| \rangle \simeq G_F^2 T^5, \quad (90)$$

while the rate of expansion of the universe at that time ($g_* = 10.75$) was $H \simeq 5.4 T^2 / M_P$, where $M_P = 1.22 \times 10^{19}$ GeV is the Planck mass. Neutrinos decouple when their interaction rate is slower than the universe expansion, $\Gamma_\nu \leq H$ or, equivalently, at $T_{\nu-\text{dec}} \simeq 0.8$ MeV. Below this temperature, neutrinos are no longer in thermal equilibrium with the rest of the plasma, and their temperature continues to decay inversely proportional to the scale factor of the universe. Since neutrinos decoupled before e^+e^- annihilation, the cosmic background of neutrinos has a temperature today lower than that of the microwave background of photons. Let us compute the difference. At temperatures above the mass of the electron, $T > m_e = 0.511$ MeV, and below 0.8 MeV, the only particle species contributing to the entropy of the universe are the photons ($g_* = 2$) and the electron-positron pairs ($g_* = 4 \times \frac{7}{8}$); total number of degrees of freedom $g_* = \frac{11}{2}$. At temperatures $T \simeq m_e$, electrons and positrons annihilate into photons, heating up the plasma (but not the neutrinos, which had decoupled already). At temperatures $T < m_e$, only photons contribute to the entropy of the universe, with $g_* = 2$ degrees of freedom.

Therefore, from the conservation of entropy, we find that the ratio of T_γ and T_ν today must be

$$\frac{T_\gamma}{T_\nu} = \left(\frac{11}{4}\right)^{1/3} = 1.401 \quad \Rightarrow \quad T_\nu = 1.945 \text{ K}, \quad (91)$$

where I have used $T_{\text{CMB}} = 2.725 \pm 0.002$ K. We still have not measured such a relic background of neutrinos, and probably will remain undetected for a long time, since they have an average energy of order 10^{-4} eV, much below that required for detection by present experiments (of order GeV), precisely because of the relative weakness of the weak interactions. Nevertheless, it would be fascinating if, in the future, ingenious experiments were devised to detect such a background, since it would confirm one of the most robust features of Big Bang cosmology.

2.66 Matter-radiation equality

Relativistic species have energy densities proportional to the quartic power of temperature and therefore scale as $\rho_R \propto a^{-4}$, while non-relativistic particles have essentially zero pressure and scale as $\rho_M \propto a^{-3}$. Therefore, there will be a time in the evolution of the universe in which both energy densities are equal $\rho_R(t_{\text{eq}}) = \rho_M(t_{\text{eq}})$. Since then both decay differently, and thus

$$1 + z_{\text{eq}} = \frac{a_0}{a_{\text{eq}}} = \frac{\Omega_M}{\Omega_R} = 3.1 \times 10^4 \Omega_M h^2, \quad (92)$$

where I have used $\Omega_R h^2 = \Omega_{\text{CMB}} h^2 + \Omega_\nu h^2 = 3.24 \times 10^{-5}$ for three massless neutrinos at $T = T_\nu$. As I will show later, the matter content of the universe today is below critical, $\Omega_M \simeq 0.3$, while $h \simeq 0.71$, and therefore $(1 + z_{\text{eq}}) \simeq 3400$, or about $t_{\text{eq}} = 1308 (\Omega_M h^2)^{-2} \text{yr} \simeq 61,000$ years after the origin of the universe. Around the time of matter-radiation equality, the rate of expansion (19) can be written as ($a_0 \equiv 1$)

$$H(a) = H_0 \left(\Omega_R a^{-4} + \Omega_M a^{-3} \right)^{1/2} = H_0 \Omega_M^{1/2} a^{-3/2} \left(1 + \frac{a_{\text{eq}}}{a} \right)^{1/2}. \quad (93)$$

The *horizon size* is the coordinate distance travelled by a photon since the beginning of the universe, $d_H \sim H^{-1}$, i.e. the size of causally connected regions in the universe. The *comoving* horizon size is then given by

$$d_H = \frac{c}{aH(a)} = c H_0^{-1} \Omega_M^{-1/2} a^{1/2} \left(1 + \frac{a_{\text{eq}}}{a} \right)^{-1/2}. \quad (94)$$

Thus the horizon size at matter-radiation equality ($a = a_{\text{eq}}$) is

$$d_H(a_{\text{eq}}) = \frac{c H_0^{-1}}{\sqrt{2}} \Omega_M^{-1/2} a_{\text{eq}}^{1/2} \simeq 12 (\Omega_M h)^{-1} h^{-1} \text{Mpc}. \quad (95)$$

This scale plays a very important role in theories of structure formation.

2.67 Recombination and photon decoupling

As the temperature of the universe decreased, electrons could eventually become bound to protons to form neutral hydrogen. Nevertheless, there is always a non-zero probability that a rare energetic photon ionizes hydrogen and produces a free electron. The *ionization fraction* of electrons in equilibrium with the plasma at a given temperature is given by the Saha equation [15]

$$\frac{1 - X_e^{\text{eq}}}{X_e^{\text{eq}}} = \frac{4\sqrt{2}\zeta(3)}{\sqrt{\pi}} \eta \left(\frac{T}{m_e} \right)^{3/2} e^{E_{\text{ion}}/T}, \quad (96)$$

where $E_{\text{ion}} = 13.6$ eV is the ionization energy of hydrogen, and η is the baryon-to-photon ratio (88). If we now use Eq. (75), we can compute the ionization fraction X_e^{eq} as a function of redshift z . Note that

the huge number of photons with respect to electrons (in the ratio ${}^4\text{He} : \text{H} : \gamma \simeq 1 : 4 : 10^{10}$) implies that even at a very low temperature, the photon distribution will contain a sufficiently large number of high-energy photons to ionize a significant fraction of hydrogen. In fact, *defining* recombination as the time at which $X_e^{\text{eq}} \equiv 0.1$, one finds that the recombination temperature is $T_{\text{rec}} = 0.31 \text{ eV} \ll E_{\text{ion}}$, for $\eta_{10} \simeq 6.2$. Comparing with the present temperature of the microwave background, we deduce the corresponding redshift at recombination, $(1 + z_{\text{rec}}) \simeq 1331$.

Photons remain in thermal equilibrium with the plasma of baryons and electrons through elastic Thomson scattering, with cross section

$$\sigma_T = \frac{8\pi\alpha^2}{3m_e^2} = 6.65 \times 10^{-25} \text{ cm}^2 = 0.665 \text{ barn}, \quad (97)$$

where $\alpha = 1/137.036$ is the dimensionless electromagnetic coupling constant. The mean free path of photons λ_γ in such a plasma can be estimated from the photon interaction rate, $\lambda_\gamma^{-1} \sim \Gamma_\gamma = n_e \sigma_T$. For temperatures above a few eV, the mean free path is much smaller than the causal horizon at that time and photons suffer multiple scattering: the plasma is like a dense fog. Photons will decouple from the plasma when their interaction rate cannot keep up with the expansion of the universe and the mean free path becomes larger than the horizon size: the universe becomes transparent. We can estimate this moment by evaluating $\Gamma_\gamma = H$ at photon decoupling. Using $n_e = X_e \eta n_\gamma$, one can compute the decoupling temperature as $T_{\text{dec}} = 0.26 \text{ eV}$, and the corresponding redshift as $1 + z_{\text{dec}} \simeq 1100$. Recently, WMAP measured this redshift to be $1 + z_{\text{dec}} \simeq 1089 \pm 1$ [20]. This redshift defines the so called *last scattering surface*, when photons last scattered off protons and electrons and travelled freely ever since. This decoupling occurred when the universe was approximately $t_{\text{dec}} = 1.5 \times 10^5 (\Omega_M h^2)^{-1/2} \simeq 380,000$ years old.

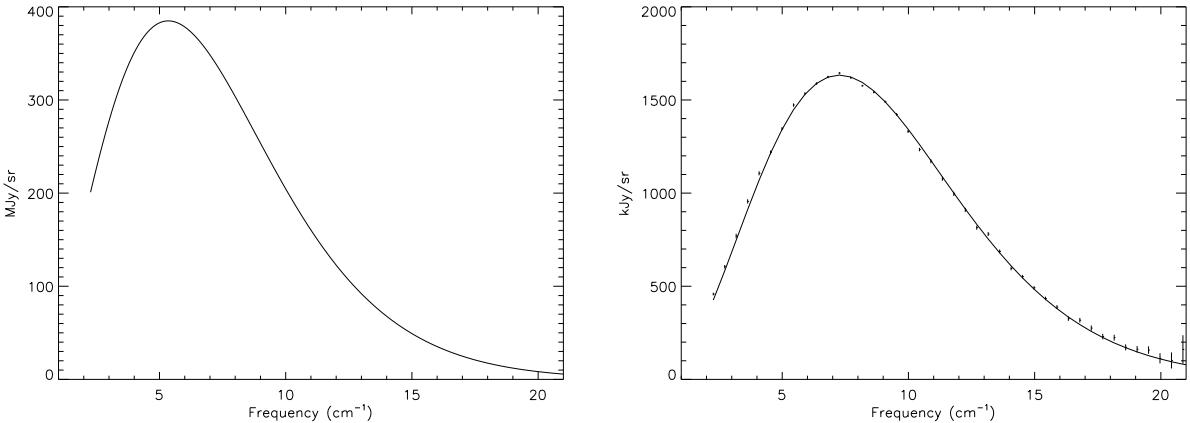


Fig. 9: The Cosmic Microwave Background Spectrum seen by the FIRAS instrument on COBE. The left panel corresponds to the monopole spectrum, $T_0 = 2.725 \pm 0.002 \text{ K}$, where the error bars are smaller than the line width. The right panel shows the dipole spectrum, $\delta T_1 = 3.372 \pm 0.014 \text{ mK}$. From Ref. [21].

2.68 The microwave background

One of the most remarkable observations ever made by mankind is the detection of the relic background of photons from the Big Bang. This background was predicted by George Gamow and collaborators in the 1940s, based on the consistency of primordial nucleosynthesis with the observed helium abundance. They estimated a value of about 10 K, although a somewhat more detailed analysis by Alpher and Herman in 1950 predicted $T_\gamma \approx 5 \text{ K}$. Unfortunately, they had doubts whether the radiation would have survived until the present, and this remarkable prediction slipped into obscurity, until Dicke, Peebles,

Roll and Wilkinson [22] studied the problem again in 1965. Before they could measure the photon background, they learned that Penzias and Wilson had observed a weak isotropic background signal at a radio wavelength of 7.35 cm, corresponding to a blackbody temperature of $T_\gamma = 3.5 \pm 1$ K. They published their two papers back to back, with that of Dicke et al. explaining the fundamental significance of their measurement [6].

Since then many different experiments have confirmed the existence of the microwave background. The most outstanding one has been the Cosmic Background Explorer (COBE) satellite, whose FIRAS instrument measured the photon background with great accuracy over a wide range of frequencies ($\nu = 1 - 97$ cm $^{-1}$), see Ref. [21], with a spectral resolution $\frac{\Delta\nu}{\nu} = 0.0035$. Nowadays, the photon spectrum is confirmed to be a blackbody spectrum with a temperature given by [21]

$$T_{\text{CMB}} = 2.725 \pm 0.002 \text{ K} \text{ (systematic, 95\% c.l.)} \pm 7 \mu\text{K} \text{ (1}\sigma \text{ statistical)} \quad (98)$$

In fact, this is the best blackbody spectrum ever measured, see Fig. 9, with spectral distortions below the level of 10 parts per million (ppm).

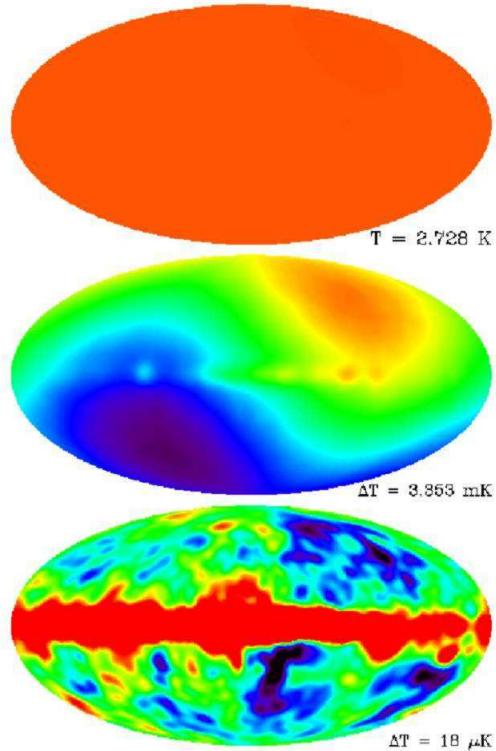


Fig. 10: The Cosmic Microwave Background Spectrum seen by the DMR instrument on COBE. The top figure corresponds to the monopole, $T_0 = 2.725 \pm 0.002$ K. The middle figure shows the dipole, $\delta T_1 = 3.372 \pm 0.014$ mK, and the lower figure shows the quadrupole and higher multipoles, $\delta T_2 = 18 \pm 2 \mu\text{K}$. The central region corresponds to foreground by the galaxy. From Ref. [23].

Moreover, the differential microwave radiometer (DMR) instrument on COBE, with a resolution of about 7° in the sky, has also confirmed that it is an extraordinarily isotropic background. The deviations from isotropy, i.e. differences in the temperature of the blackbody spectrum measured in different directions in the sky, are of the order of 20 μK on large scales, or one part in 10⁵, see Ref. [23]. There is, in fact, a dipole anisotropy of one part in 10³, $\delta T_1 = 3.372 \pm 0.007$ mK (95% c.l.), in the direction of the Virgo cluster, $(l, b) = (264.14^\circ \pm 0.30, 48.26^\circ \pm 0.30)$ (95% c.l.). Under the assumption that a Doppler effect is responsible for the entire CMB dipole, the velocity of the Sun with respect to the CMB

rest frame is $v_\odot = 371 \pm 0.5$ km/s, see Ref. [21].⁴ When subtracted, we are left with a whole spectrum of anisotropies in the higher multipoles (quadrupole, octupole, etc.), $\delta T_2 = 18 \pm 2$ μK (95% c.l.), see Ref. [23] and Fig. 10.

Soon after COBE, other groups quickly confirmed the detection of temperature anisotropies at around 30 μK and above, at higher multipole numbers or smaller angular scales. As I shall discuss below, these anisotropies play a crucial role in the understanding of the origin of structure in the universe.

2.69 Large-scale structure formation

Although the isotropic microwave background indicates that the universe in the *past* was extraordinarily homogeneous, we know that the universe *today* is not exactly homogeneous: we observe galaxies, clusters and superclusters on large scales. These structures are expected to arise from very small primordial inhomogeneities that grow in time via gravitational instability, and that may have originated from tiny ripples in the metric, as matter fell into their troughs. Those ripples must have left some trace as temperature anisotropies in the microwave background, and indeed such anisotropies were finally discovered by the COBE satellite in 1992. The reason why they took so long to be discovered was that they appear as perturbations in temperature of only one part in 10^5 .

While the predicted anisotropies have finally been seen in the CMB, not all kinds of matter and/or evolution of the universe can give rise to the structure we observe today. If we define the density contrast as [24]

$$\delta(\vec{x}, a) \equiv \frac{\rho(\vec{x}, a) - \bar{\rho}(a)}{\bar{\rho}(a)} = \int d^3 k \delta_k(a) e^{i\vec{k}\cdot\vec{x}}, \quad (99)$$

where $\bar{\rho}(a) = \rho_0 a^{-3}$ is the average cosmic density, we need a theory that will grow a density contrast with amplitude $\delta \sim 10^{-5}$ at the last scattering surface ($z = 1100$) up to density contrasts of the order of $\delta \sim 10^2$ for galaxies at redshifts $z \ll 1$, i.e. today. This is a *necessary* requirement for any consistent theory of structure formation [25].

Furthermore, the anisotropies observed by the COBE satellite correspond to a small-amplitude scale-invariant primordial power spectrum of inhomogeneities

$$P(k) = \langle |\delta_k|^2 \rangle \propto k^n, \quad \text{with} \quad n = 1, \quad (100)$$

where the brackets $\langle \cdot \rangle$ represent integration over an ensemble of different universe realizations. These inhomogeneities are like waves in the space-time metric. When matter fell in the troughs of those waves, it created density perturbations that collapsed gravitationally to form galaxies and clusters of galaxies, with a spectrum that is also scale invariant. Such a type of spectrum was proposed in the early 1970s by Edward R. Harrison, and independently by the Russian cosmologist Yakov B. Zel'dovich, see Ref. [26], to explain the distribution of galaxies and clusters of galaxies on very large scales in our observable universe.

Today various telescopes – like the Hubble Space Telescope, the twin Keck telescopes in Hawaii and the European Southern Observatory telescopes in Chile – are exploring the most distant regions of the universe and discovering the first galaxies at large distances. The furthest galaxies observed so far are at redshifts of $z \simeq 10$ (at a distance of 13.7 billion light years from Earth), whose light was emitted when the universe had only about 3% of its present age. Only a few galaxies are known at those redshifts, but there are at present various catalogs like the CfA and APM galaxy catalogs, and more recently the IRAS Point Source redshift Catalog, see Fig. 11, and Las Campanas redshift surveys, that study the spatial distribution of hundreds of thousands of galaxies up to distances of a billion light years, or $z < 0.1$, or the 2 degree Field Galaxy Redshift Survey (2dFGRS) and the Sloan Digital Sky Survey (SDSS), which reach $z < 0.5$ and study millions of galaxies. These catalogs are telling us about the evolution

⁴COBE even determined the annual variation due to the Earth's motion around the Sun – the ultimate proof of Copernicus' hypothesis.

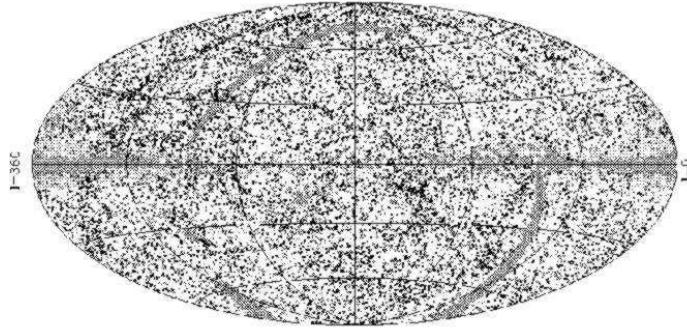


Fig. 11: The IRAS Point Source Catalog redshift survey contains some 15,000 galaxies, covering over 83% of the sky up to redshifts of $z \leq 0.05$. We show here the projection of the galaxy distribution in galactic coordinates. From Ref. [27].

of clusters and superclusters of galaxies in the universe, and already put constraints on the theory of structure formation. From these observations one can infer that most galaxies formed at redshifts of the order of 2 – 6; clusters of galaxies formed at redshifts of order 1, and superclusters are forming now. That is, cosmic structure formed from the bottom up: from galaxies to clusters to superclusters, and not the other way around. This fundamental difference is an indication of the type of matter that gave rise to structure.

We know from Big Bang nucleosynthesis that all the baryons in the universe cannot account for the observed amount of matter, so there must be some extra matter (dark since we don't see it) to account for its gravitational pull. Whether it is relativistic (hot) or non-relativistic (cold) could be inferred from observations: relativistic particles tend to diffuse from one concentration of matter to another, thus transferring energy among them and preventing the growth of structure on small scales. This is excluded by observations, so we conclude that most of the matter responsible for structure formation must be cold. How much there is is a matter of debate at the moment. Some recent analyses suggest that there is not enough cold dark matter to reach the critical density required to make the universe flat. If we want to make sense of the present observations, we must conclude that some other form of energy permeates the universe. In order to resolve this issue, 2dFGRS and SDSS started taking data a few years ago. The first has already been completed, but the second one is still taking data up to redshifts $z \simeq 5$ for quasars, over a large region of the sky. These important observations will help astronomers determine the nature of the dark matter and test the validity of the models of structure formation.

Before COBE discovered the anisotropies of the microwave background there were serious doubts whether gravity alone could be responsible for the formation of the structure we observe in the universe today. It seemed that a new force was required to do the job. Fortunately, the anisotropies were found with the right amplitude for structure to be accounted for by gravitational collapse of primordial inhomogeneities under the attraction of a large component of non-relativistic dark matter. Nowadays, the standard theory of structure formation is a cold dark matter model with a non vanishing cosmological constant in a spatially flat universe. Gravitational collapse amplifies the density contrast initially through linear growth and later on via non-linear collapse. In the process, overdense regions decouple from the Hubble expansion to become bound systems, which start attracting each other to form larger bound structures. In fact, the largest structures, superclusters, have not yet gone non-linear.

The primordial spectrum (100) is reprocessed by gravitational instability after the universe becomes matter dominated and inhomogeneities can grow. Linear perturbation theory shows that the grow-

ing mode⁵ of small density contrasts go like [24, 25]

$$\delta(a) \propto a^{1+3\omega} = \begin{cases} a^2, & a < a_{\text{eq}} \\ a, & a > a_{\text{eq}} \end{cases} \quad (101)$$

in the Einstein-de Sitter limit ($\omega = p/\rho = 1/3$ and 0, for radiation and matter, respectively). There are slight deviations for $a \gg a_{\text{eq}}$, if $\Omega_M \neq 1$ or $\Omega_\Lambda \neq 0$, but we will not be concerned with them here. The important observation is that, since the density contrast at last scattering is of order $\delta \sim 10^{-5}$, and the scale factor has grown since then only a factor $z_{\text{dec}} \sim 10^3$, one would expect a density contrast today of order $\delta_0 \sim 10^{-2}$. Instead, we observe structures like galaxies, where $\delta \sim 10^2$. So how can this be possible? The microwave background shows anisotropies due to fluctuations in the baryonic matter component only (to which photons couple, electromagnetically). If there is an additional matter component that only couples through very weak interactions, fluctuations in that component could grow as soon as it decoupled from the plasma, well before photons decoupled from baryons. The reason why baryonic inhomogeneities cannot grow is because of photon pressure: as baryons collapse towards denser regions, radiation pressure eventually halts the contraction and sets up acoustic oscillations in the plasma that prevent the growth of perturbations, until photon decoupling. On the other hand, a weakly interacting cold dark matter component could start gravitational collapse much earlier, even before matter-radiation equality, and thus reach the density contrast amplitudes observed today. The resolution of this mismatch is one of the strongest arguments for the existence of a weakly interacting cold dark matter component of the universe.

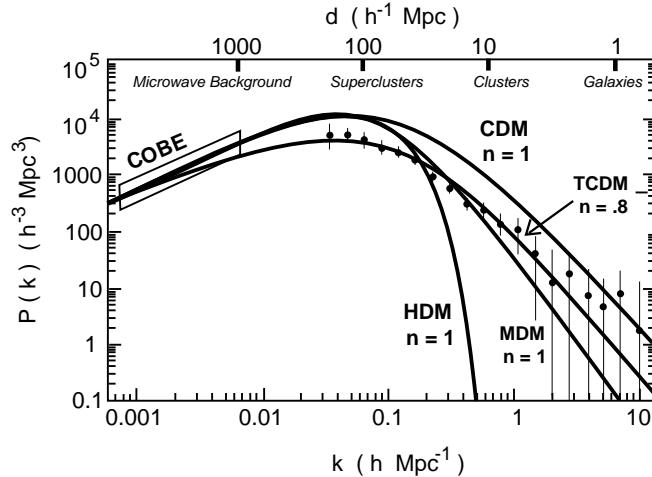


Fig. 12: The power spectrum for cold dark matter (CDM), tilted cold dark matter (TCDM), hot dark matter (HDM), and mixed hot plus cold dark matter (MDM), normalized to COBE, for large-scale structure formation. From Ref. [28].

How much dark matter there is in the universe can be deduced from the actual power spectrum (the Fourier transform of the two-point correlation function of density perturbations) of the observed large scale structure. One can decompose the density contrast in Fourier components, see Eq. (99). This is very convenient since in linear perturbation theory individual Fourier components evolve independently. A comoving wavenumber k is said to “enter the horizon” when $k = d_H^{-1}(a) = aH(a)$. If a certain perturbation, of wavelength $\lambda = k^{-1} < d_H(a_{\text{eq}})$, enters the horizon before matter-radiation equality, the fast radiation-driven expansion prevents dark-matter perturbations from collapsing. Since light can only cross regions that are smaller than the horizon, the suppression of growth due to radiation is restricted to scales smaller than the horizon, while large-scale perturbations remain unaffected. This is the reason

⁵The decaying modes go like $\delta(t) \sim t^{-1}$, for all ω .

why the horizon size at equality, Eq. (95), sets an important scale for structure growth,

$$k_{\text{eq}} = d_H^{-1}(a_{\text{eq}}) \simeq 0.083 (\Omega_M h) h \text{ Mpc}^{-1}. \quad (102)$$

The suppression factor can be easily computed from (101) as $f_{\text{sup}} = (a_{\text{enter}}/a_{\text{eq}})^2 = (k_{\text{eq}}/k)^2$. In other words, the processed power spectrum $P(k)$ will have the form:

$$P(k) \propto \begin{cases} k, & k \ll k_{\text{eq}} \\ k^{-3}, & k \gg k_{\text{eq}} \end{cases} \quad (103)$$

This is precisely the shape that large-scale galaxy catalogs are bound to test in the near future, see Fig. 12. Furthermore, since relativistic Hot Dark Matter (HDM) transfer energy between clumps of matter, they will wipe out small scale perturbations, and this should be seen as a distinctive signature in the matter power spectra of future galaxy catalogs. On the other hand, non-relativistic Cold Dark Matter (CDM) allow structure to form on *all* scales via gravitational collapse. The dark matter will then pull in the baryons, which will later shine and thus allow us to see the galaxies.

Naturally, when baryons start to collapse onto dark matter potential wells, they will convert a large fraction of their potential energy into kinetic energy of protons and electrons, ionizing the medium. As a consequence, we expect to see a large fraction of those baryons constituting a hot ionized gas surrounding large clusters of galaxies. This is indeed what is observed, and confirms the general picture of structure formation.

3. DETERMINATION OF COSMOLOGICAL PARAMETERS

In this Section, I will restrict myself to those recent measurements of the cosmological parameters by means of standard cosmological techniques, together with a few instances of new results from recently applied techniques. We will see that a large host of observations are determining the cosmological parameters with some reliability of the order of 10%. However, the majority of these measurements are dominated by large systematic errors. Most of the recent work in observational cosmology has been the search for virtually systematic-free observables, like those obtained from the microwave background anisotropies, and discussed in Section 4.4. I will devote, however, this Section to the more ‘classical’ measurements of the following cosmological parameters: The rate of expansion H_0 ; the matter content Ω_M ; the cosmological constant Ω_Λ ; the spatial curvature Ω_K , and the age of the universe t_0 .

3.1 The rate of expansion H_0

Over most of last century the value of H_0 has been a constant source of disagreement [29]. Around 1929, Hubble measured the rate of expansion to be $H_0 = 500 \text{ km s}^{-1} \text{Mpc}^{-1}$, which implied an age of the universe of order $t_0 \sim 2 \text{ Gyr}$, in clear conflict with geology. Hubble’s data was based on Cepheid standard candles that were incorrectly calibrated with those in the Large Magellanic Cloud. Later on, in 1954 Baade recalibrated the Cepheid distance and obtained a lower value, $H_0 = 250 \text{ km s}^{-1} \text{Mpc}^{-1}$, still in conflict with ratios of certain unstable isotopes. Finally, in 1958 Sandage realized that the brightest stars in galaxies were ionized HII regions, and the Hubble rate dropped down to $H_0 = 60 \text{ km s}^{-1} \text{Mpc}^{-1}$, still with large (factor of two) systematic errors. Fortunately, in the past 15 years there has been significant progress towards the determination of H_0 , with systematic errors approaching the 10% level. These improvements come from two directions. First, technological, through the replacement of photographic plates (almost exclusively the source of data from the 1920s to 1980s) with charged couple devices (CCDs), i.e. solid state detectors with excellent flux sensitivity per pixel, which were previously used successfully in particle physics detectors. Second, by the refinement of existing methods for measuring extragalactic distances (e.g. parallax, Cepheids, supernovae, etc.). Finally, with the development of completely new methods to determine H_0 , which fall into totally independent and very broad categories: a) Gravitational lensing; b) Sunyaev-Zel’dovich effect; c) Extragalactic distance scale, mainly

Cepheid variability and type Ia Supernovae; d) Microwave background anisotropies. I will review here the first three, and leave the last method for Section 4.4, since it involves knowledge about the primordial spectrum of inhomogeneities.

3.11 Gravitational lensing

Imagine a quasi-stellar object (QSO) at large redshift ($z \gg 1$) whose light is lensed by an intervening galaxy at redshift $z \sim 1$ and arrives to an observer at $z = 0$. There will be at least two different images of the same background *variable* point source. The arrival times of photons from two different gravitationally lensed images of the quasar depend on the different path lengths and the gravitational potential traversed. Therefore, a measurement of the time delay and the angular separation of the different images of a variable quasar can be used to determine H_0 with great accuracy. This method, proposed in 1964 by Refsdal [30], offers tremendous potential because it can be applied at great distances and it is based on very solid physical principles [31].

Unfortunately, there are very few systems with both a favourable geometry (i.e. a known mass distribution of the intervening galaxy) and a variable background source with a measurable time delay. That is the reason why it has taken so much time since the original proposal for the first results to come out. Fortunately, there are now very powerful telescopes that can be used for these purposes. The best candidate to-date is the QSO 0957 + 561, observed with the 10m Keck telescope, for which there is a model of the lensing mass distribution that is consistent with the measured velocity dispersion. Assuming a flat space with $\Omega_M = 0.25$, one can determine [32]

$$H_0 = 72 \pm 7 \text{ (1}\sigma \text{ statistical)} \pm 15\% \text{ (systematic)} \text{ km s}^{-1} \text{Mpc}^{-1}. \quad (104)$$

The main source of systematic error is the degeneracy between the mass distribution of the lens and the value of H_0 . Knowledge of the velocity dispersion within the lens as a function of position helps constrain the mass distribution, but those measurements are very difficult and, in the case of lensing by a cluster of galaxies, the dark matter distribution in those systems is usually unknown, associated with a complicated cluster potential. Nevertheless, the method is just starting to give promising results and, in the near future, with the recent discovery of several systems with optimum properties, the prospects for measuring H_0 and lowering its uncertainty with this technique are excellent.

3.12 Sunyaev-Zel'dovich effect

As discussed in the previous Section, the gravitational collapse of baryons onto the potential wells generated by dark matter gave rise to the reionization of the plasma, generating an X-ray halo around rich clusters of galaxies, see Fig. 13. The inverse-Compton scattering of microwave background photons off the hot electrons in the X-ray gas results in a measurable distortion of the blackbody spectrum of the microwave background, known as the Sunyaev-Zel'dovich (SZ) effect. Since photons acquire extra energy from the X-ray electrons, we expect a shift towards higher frequencies of the spectrum, $(\Delta\nu/\nu) \simeq (k_B T_{\text{gas}}/m_e c^2) \sim 10^{-2}$. This corresponds to a *decrement* of the microwave background temperature at low frequencies (Rayleigh-Jeans region) and an increment at high frequencies, see Ref. [33].

Measuring the *spatial* distribution of the SZ effect (3 K spectrum), together with a high resolution X-ray map (10^8 K spectrum) of the cluster, one can determine the density and temperature distribution of the hot gas. Since the X-ray flux is distance-dependent ($\mathcal{F} = \mathcal{L}/4\pi d_L^2$), while the SZ decrement is not (because the energy of the CMB photons increases as we go back in redshift, $\nu = \nu_0(1+z)$, and exactly compensates the redshift in energy of the photons that reach us), one can determine from there the distance to the cluster, and thus the Hubble rate H_0 .

The advantages of this method are that it can be applied to large distances and it is based on clear physical principles. The main systematics come from possible clumpiness of the gas (which would

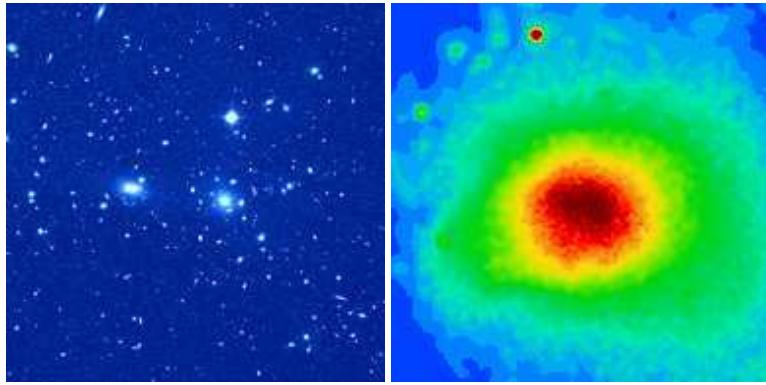


Fig. 13: The Coma cluster of galaxies, seen here in an optical image (left) and an X-ray image (right), taken by the recently launched Chandra X-ray Observatory. From Ref. [34].

reduce H_0), projection effects (if the clusters are prolate, H_0 could be larger), the assumption of hydrostatic equilibrium of the X-ray gas, details of models for the gas and electron densities, and possible contaminations from point sources. Present measurements give the value [33]

$$H_0 = 60 \pm 10 \text{ (1}\sigma \text{ statistical)} \pm 20\% \text{ (systematic)} \text{ km s}^{-1} \text{Mpc}^{-1}, \quad (105)$$

compatible with other determinations. A great advantage of this completely new and independent method is that nowadays more and more clusters are observed in the X-ray, and soon we will have high-resolution 2D maps of the SZ decrement from several balloon flights, as well as from future microwave background satellites, together with precise X-ray maps and spectra from the Chandra X-ray observatory recently launched by NASA, as well as from the European X-ray satellite XMM launched a few months ago by ESA, which will deliver orders of magnitude better resolution than the existing Einstein X-ray satellite.

3.13 Cepheid variability

Cepheids are low-mass variable stars with a period-luminosity relation based on the helium ionization cycles inside the star, as it contracts and expands. This time variability can be measured, and the star's absolute luminosity determined from the calibrated relationship. From the observed flux one can then deduce the luminosity distance, see Eq. (28), and thus the Hubble rate H_0 . The Hubble Space Telescope (HST) was launched by NASA in 1990 (and repaired in 1993) with the specific project of calibrating the extragalactic distance scale and thus determining the Hubble rate with 10% accuracy. The most recent results from HST are the following [35]

$$H_0 = 71 \pm 4 \text{ (random)} \pm 7 \text{ (systematic)} \text{ km s}^{-1} \text{Mpc}^{-1}. \quad (106)$$

The main source of systematic error is the distance to the Large Magellanic Cloud, which provides the fiducial comparison for Cepheids in more distant galaxies. Other systematic uncertainties that affect the value of H_0 are the internal extinction correction method used, a possible metallicity dependence of the Cepheid period-luminosity relation and cluster population incompleteness bias, for a set of 21 galaxies within 25 Mpc, and 23 clusters within $z \lesssim 0.03$.

With better telescopes already taking data, like the Very Large Telescope (VLT) interferometer of the European Southern Observatory (ESO) in the Chilean Atacama desert, with 8 synchronized telescopes, and others coming up soon, like the Next Generation Space Telescope (NGST) proposed by NASA for 2008, and the Gran TeCan of the European Northern Observatory in the Canary Islands, for 2010, it is expected that much better resolution and therefore accuracy can be obtained for the determination of H_0 .

3.2 Dark Matter

In the 1920s Hubble realized that the so called nebulae were actually distant galaxies very similar to our own. Soon afterwards, in 1933, Zwicky found dynamical evidence that there is possibly ten to a hundred times more mass in the Coma cluster than contributed by the luminous matter in galaxies [36]. However, it was not until the 1970s that the existence of dark matter began to be taken more seriously. At that time there was evidence that rotation curves of galaxies did not fall off with radius and that the dynamical mass was increasing with scale from that of individual galaxies up to clusters of galaxies. Since then, new possible extra sources to the matter content of the universe have been accumulating:

$$\Omega_M = \Omega_{B, \text{lum}} \quad (\text{stars in galaxies}) \quad (107)$$

$$+ \Omega_{B, \text{dark}} \quad (\text{MACHOs?}) \quad (108)$$

$$+ \Omega_{CDM} \quad (\text{weakly interacting : axion, neutralino?}) \quad (109)$$

$$+ \Omega_{HDM} \quad (\text{massive neutrinos?}) \quad (110)$$

The empirical route to the determination of Ω_M is nowadays one of the most diversified of all cosmological parameters. The matter content of the universe can be deduced from the mass-to-light ratio of various objects in the universe; from the rotation curves of galaxies; from microlensing and the direct search of Massive Compact Halo Objects (MACHOs); from the cluster velocity dispersion with the use of the Virial theorem; from the baryon fraction in the X-ray gas of clusters; from weak gravitational lensing; from the observed matter distribution of the universe via its power spectrum; from the cluster abundance and its evolution; from direct detection of massive neutrinos at SuperKamiokande; from direct detection of Weakly Interacting Massive Particles (WIMPs) at CDMS, DAMA or UKDMC, and finally from microwave background anisotropies. I will review here just a few of them.

3.2.1 Rotation curves of spiral galaxies

The flat rotation curves of spiral galaxies provide the most direct evidence for the existence of large amounts of dark matter. Spiral galaxies consist of a central bulge and a very thin disk, stabilized against gravitational collapse by angular momentum conservation, and surrounded by an approximately spherical halo of dark matter. One can measure the orbital velocities of objects orbiting around the disk as a function of radius from the Doppler shifts of their spectral lines.

The rotation curve of the Andromeda galaxy was first measured by Babcock in 1938, from the stars in the disk. Later it became possible to measure galactic rotation curves far out into the disk, and a trend was found [37]. The orbital velocity rose linearly from the center outward until it reached a typical value of 200 km/s, and then remained flat out to the largest measured radii. This was completely unexpected since the observed surface luminosity of the disk falls off exponentially with radius [37], $I(r) = I_0 \exp(-r/r_D)$. Therefore, one would expect that most of the galactic mass is concentrated within a few disk lengths r_D , such that the rotation velocity is determined as in a Keplerian orbit, $v_{\text{rot}} = (GM/r)^{1/2} \propto r^{-1/2}$. No such behaviour is observed. In fact, the most convincing observations come from radio emission (from the 21 cm line) of neutral hydrogen in the disk, which has been measured to much larger galactic radii than optical tracers. A typical case is that of the spiral galaxy NGC 6503, where $r_D = 1.73$ kpc, while the furthest measured hydrogen line is at $r = 22.22$ kpc, about 13 disk lengths away. Nowadays, thousands of galactic rotation curves are known, see Fig. 14, and all suggest the existence of about ten times more mass in the halos of spiral galaxies than in the stars of the disk. Recent numerical simulations of galaxy formation in a CDM cosmology [38] suggest that galaxies probably formed by the infall of material in an overdense region of the universe that had decoupled from the overall expansion.

The dark matter is supposed to undergo violent relaxation and create a virialized system, i.e. in hydrostatic equilibrium. This picture has led to a simple model of dark-matter halos as isothermal spheres, with density profile $\rho(r) = \rho_c/(r_c^2 + r^2)$, where r_c is a core radius and $\rho_c = v_\infty^2/4\pi G$, with

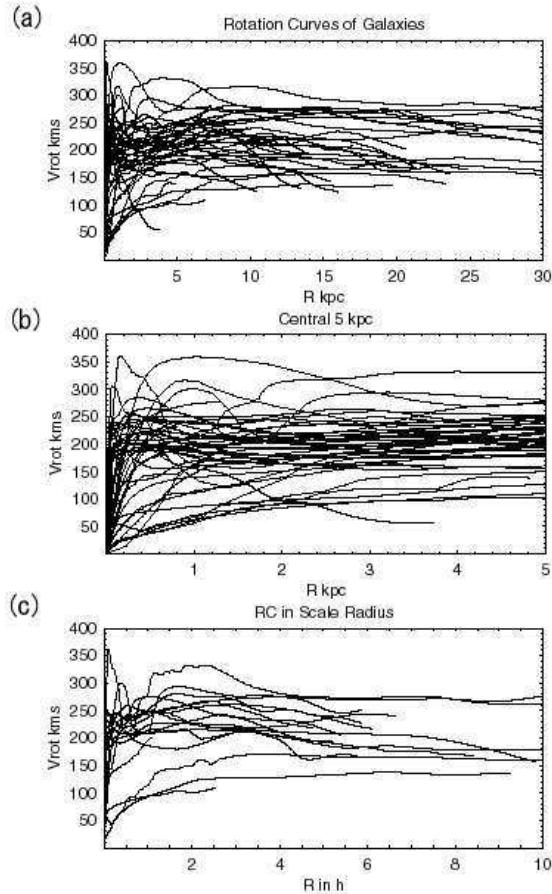


Fig. 14: The rotation curves of several hundred galaxies. Upper panel: As a function of their radii in kpc. Middle panel: The central 5 kpc. Lower panel: As a function of scale radius.

v_∞ equal to the plateau value of the flat rotation curve. This model is consistent with the universal rotation curves seen in Fig. 6. At large radii the dark matter distribution leads to a flat rotation curve. The question is for how long. In dense galaxy clusters one expects the galactic halos to overlap and form a continuum, and therefore the rotation curves should remain flat from one galaxy to another. However, in field galaxies, far from clusters, one can study the rotation velocities of substructures (like satellite dwarf galaxies) around a given galaxy, and determine whether they fall off at sufficiently large distances according to Kepler's law, as one would expect, once the edges of the dark matter halo have been reached. These observations are rather difficult because of uncertainties in distinguishing between true satellites and interlopers. Recently, a group from the Sloan Digital Sky Survey Collaboration claim that they have seen the edges of the dark matter halos around field galaxies by confirming the fall-off at large distances of their rotation curves [39]. These results, if corroborated by further analysis, would constitute a tremendous support to the idea of dark matter as a fluid surrounding galaxies and clusters, while at the same time eliminates the need for modifications of Newtonian or even Einsteinian gravity at the scales of galaxies, to account for the flat rotation curves.

That's fine, but how much dark matter is there at the galactic scale? Adding up all the matter in galactic halos up to a maximum radii, one finds

$$\Omega_{\text{halo}} \simeq 10 \Omega_{\text{lum}} \geq 0.03 - 0.05. \quad (111)$$

Of course, it would be extraordinary if we could confirm, through direct detection, the existence of dark matter in our own galaxy. For that purpose, one should measure its rotation curve, which is much more

difficult because of obscuration by dust in the disk, as well as problems with the determination of reliable galactocentric distances for the tracers. Nevertheless, the rotation curve of the Milky Way has been measured and conforms to the usual picture, with a plateau value of the rotation velocity of 220 km/s. For dark matter searches, the crucial quantity is the dark matter density in the solar neighbourhood, which turns out to be (within a factor of two uncertainty depending on the halo model) $\rho_{\text{DM}} = 0.3 \text{ GeV/cm}^3$. We will come back to direct searched of dark matter in a later subsection.

3.22 Baryon fraction in clusters

Since large clusters of galaxies form through gravitational collapse, they scoop up mass over a large volume of space, and therefore the ratio of baryons over the total matter in the cluster should be representative of the entire universe, at least within a 20% systematic error. Since the 1960s, when X-ray telescopes became available, it is known that galaxy clusters are the most powerful X-ray sources in the sky [40]. The emission extends over the whole cluster and reveals the existence of a hot plasma with temperature $T \sim 10^7 - 10^8 \text{ K}$, where X-rays are produced by electron bremsstrahlung. Assuming the gas to be in hydrostatic equilibrium and applying the virial theorem one can estimate the total mass in the cluster, giving general agreement (within a factor of 2) with the virial mass estimates. From these estimates one can calculate the baryon fraction of clusters

$$f_B h^{3/2} = 0.08 \quad \Rightarrow \quad \frac{\Omega_B}{\Omega_M} \approx 0.14, \quad \text{for } h = 0.70. \quad (112)$$

Since $\Omega_{\text{lum}} \simeq 0.002 - 0.006$, the previous expression suggests that clusters contain far more baryonic matter in the form of hot gas than in the form of stars in galaxies. Assuming this fraction to be representative of the entire universe, and using the Big Bang nucleosynthesis value of $\Omega_B = 0.04 \pm 0.01$, for $h = 0.7$, we find

$$\Omega_M = 0.3 \pm 0.1 \text{ (statistical)} \pm 20\% \text{ (systematic)}. \quad (113)$$

This value is consistent with previous determinations of Ω_M . If some baryons are ejected from the cluster during gravitational collapse, or some are actually bound in nonluminous objects like planets, then the actual value of Ω_M is smaller than this estimate.

3.23 Weak gravitational lensing

Since the mid 1980s, deep surveys with powerful telescopes have observed huge arc-like features in galaxy clusters. The spectroscopic analysis showed that the cluster and the giant arcs were at very different redshifts. The usual interpretation is that the arc is the image of a distant background galaxy which is in the same line of sight as the cluster so that it appears distorted and magnified by the gravitational lens effect: the giant arcs are essentially partial Einstein rings. From a systematic study of the cluster mass distribution one can reconstruct the shear field responsible for the gravitational distortion [41]. This analysis shows that there are large amounts of dark matter in the clusters, in rough agreement with the virial mass estimates, although the lensing masses tend to be systematically larger. At present, the estimates indicate $\Omega_M = 0.2 - 0.3$ on scales $\lesssim 6 h^{-1} \text{ Mpc}$.

3.24 Large scale structure formation and the matter power spectrum

Although the isotropic microwave background indicates that the universe in the *past* was extraordinarily homogeneous, we know that the universe *today* is far from homogeneous: we observe galaxies, clusters and superclusters on large scales. These structures are expected to arise from very small primordial inhomogeneities that grow in time via gravitational instability, and that may have originated from tiny ripples in the metric, as matter fell into their troughs. Those ripples must have left some trace as temperature anisotropies in the microwave background, and indeed such anisotropies were finally discovered by the

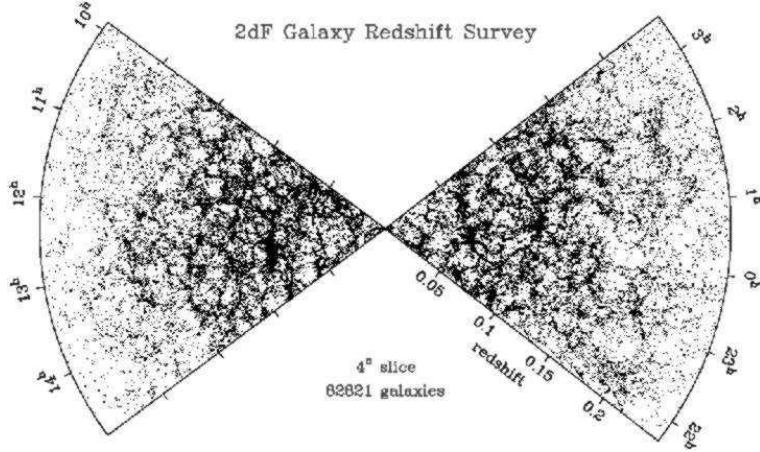


Fig. 15: The 2 degree Field Galaxy Redshift Survey contains some 250,000 galaxies, covering a large fraction of the sky up to redshifts of $z \leq 0.25$. From Ref. [42].

COBE satellite in 1992. However, not all kinds of matter and/or evolution of the universe can give rise to the structure we observe today. If we define the density contrast as

$$\delta(\vec{x}, a) \equiv \frac{\rho(\vec{x}, a) - \bar{\rho}(a)}{\bar{\rho}(a)} = \int d^3 k \delta_k(a) e^{i\vec{k}\cdot\vec{x}}, \quad (114)$$

where $\bar{\rho}(a) = \rho_0 a^{-3}$ is the average cosmic density, we need a theory that will grow a density contrast with amplitude $\delta \sim 10^{-5}$ at the last scattering surface ($z = 1100$) up to density contrasts of the order of $\delta \sim 10^2$ for galaxies at redshifts $z \ll 1$, i.e. today. This is a *necessary* requirement for any consistent theory of structure formation.

Furthermore, the anisotropies observed by the COBE satellite correspond to a small-amplitude scale-invariant primordial power spectrum of inhomogeneities

$$P(k) = \langle |\delta_k|^2 \rangle \propto k^n, \quad \text{with} \quad n = 1, \quad (115)$$

These inhomogeneities are like waves in the space-time metric. When matter fell in the troughs of those waves, it created density perturbations that collapsed gravitationally to form galaxies and clusters of galaxies, with a spectrum that is also scale invariant. Such a type of spectrum was proposed in the early 1970s by Edward R. Harrison, and independently by the Russian cosmologist Yakov B. Zel'dovich [26], to explain the distribution of galaxies and clusters of galaxies on very large scales in our observable universe, see Fig. 15.

Since the primordial spectrum is very approximately represented by a scale-invariant *Gaussian random field*, the best way to present the results of structure formation is by working with the 2-point correlation function in Fourier space, the so-called *power spectrum*. If the reprocessed spectrum of inhomogeneities remains Gaussian, the power spectrum is all we need to describe the galaxy distribution. Non-Gaussian effects are expected to arise from the non-linear gravitational collapse of structure, and may be important at small scales. The power spectrum measures the degree of inhomogeneity in the mass distribution on different scales, see Fig. 16. It depends upon a few basic ingredients: a) the primordial spectrum of inhomogeneities, whether they are Gaussian or non-Gaussian, whether *adiabatic* (perturbations in the energy density) or *isocurvature* (perturbations in the entropy density), whether the primordial spectrum has *tilt* (deviations from scale-invariance), etc.; b) the recent creation of inhomogeneities, whether *cosmic strings* or some other topological defect from an early phase transition are responsible for the formation of structure today; and c) the cosmic evolution of the inhomogeneity,

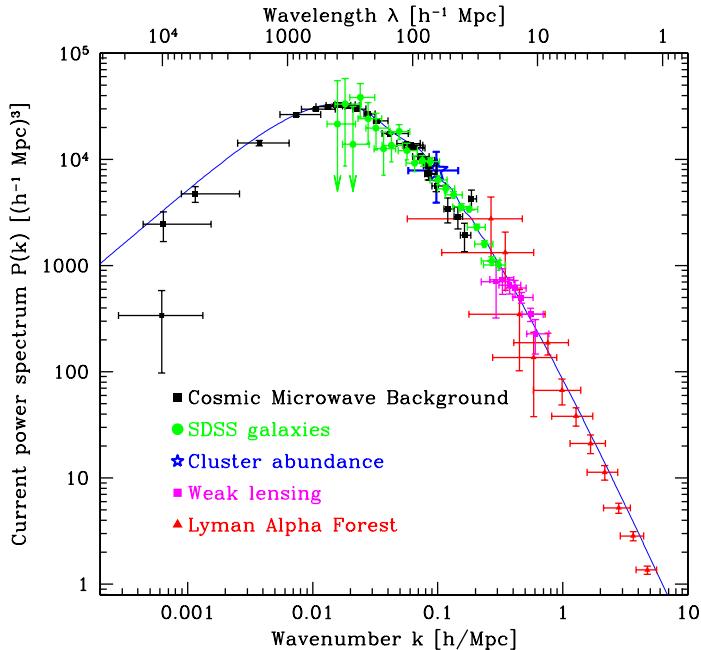


Fig. 16: The measured power spectrum $P(k)$ as a function of wavenumber k . From observations of the Sloan Digital Sky Survey, CMB anisotropies, cluster abundance, gravitational lensing and Lyman- α forest. From Ref. [43].

whether the universe has been dominated by cold or hot dark matter or by a cosmological constant since the beginning of structure formation, and also depending on the rate of expansion of the universe.

The working tools used for the comparison between the observed power spectrum and the predicted one are very precise N-body numerical simulations and theoretical models that predict the *shape* but not the *amplitude* of the present power spectrum. Even though a large amount of work has gone into those analyses, we still have large uncertainties about the nature and amount of matter necessary for structure formation. A model that has become a working paradigm is a flat cold dark matter model with a cosmological constant and $\Omega_M \sim 0.3$. This model is now been confronted with the recent very precise measurements from 2dFGRS [42] and SDSS [43].

3.25 The new redshift catalogs, 2dF and Sloan Digital Sky Survey

Our view of the large-scale distribution of luminous objects in the universe has changed dramatically during the last 25 years: from the simple pre-1975 picture of a distribution of field and cluster galaxies, to the discovery of the first single superstructures and voids, to the most recent results showing an almost regular web-like network of interconnected clusters, filaments and walls, separating huge nearly empty volumes. The increased efficiency of redshift surveys, made possible by the development of spectrographs and – specially in the last decade – by an enormous increase in multiplexing gain (i.e. the ability to collect spectra of several galaxies at once, thanks to fibre-optic spectrographs), has allowed us not only to do *cartography* of the nearby universe, but also to statistically characterize some of its properties. At the same time, advances in theoretical modeling of the development of structure, with large high-resolution gravitational simulations coupled to a deeper yet limited understanding of how to form galaxies within the dark matter halos, have provided a more realistic connection of the models to the observable quantities. Despite the large uncertainties that still exist, this has transformed the study of

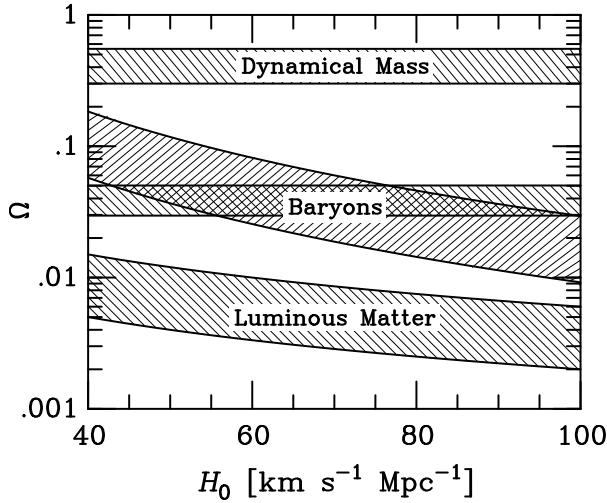


Fig. 17: The observed cosmic matter components as functions of the Hubble expansion parameter. The luminous matter component is given by $0.002 \leq \Omega_{\text{lum}} \leq 0.006$; the galactic halo component is the horizontal band, $0.03 \leq \Omega_{\text{halo}} \leq 0.05$, crossing the baryonic component from BBN, $\Omega_B h^2 = 0.0244 \pm 0.0024$; and the dynamical mass component from large scale structure analysis is given by $\Omega_M = 0.3 \pm 0.1$. Note that in the range $H_0 = 70 \pm 7 \text{ km/s/Mpc}$, there are *three* dark matter problems, see the text. From Ref. [44].

cosmology and large-scale structure into a truly quantitative science, where theory and observations can progress together.

3.26 Summary of the matter content

We can summarize the present situation with Fig. 17, for Ω_M as a function of H_0 . There are four bands, the luminous matter Ω_{lum} ; the baryon content Ω_B , from BBN; the galactic halo component Ω_{halo} , and the dynamical mass from clusters, Ω_M . From this figure it is clear that there are in fact *three* dark matter problems: The first one is where are 90% of the baryons? Between the fraction predicted by BBN and that seen in stars and diffuse gas there is a huge fraction which is in the form of dark baryons. They could be in small clumps of hydrogen that have not started thermonuclear reactions and perhaps constitute the dark matter of spiral galaxies' halos. Note that although Ω_B and Ω_{halo} coincide at $H_0 \simeq 70 \text{ km/s/Mpc}$, this could be just a coincidence. The second problem is what constitutes 90% of matter, from BBN baryons to the mass inferred from cluster dynamics? This is the standard dark matter problem and could be solved in the future by direct detection of a weakly interacting massive particle in the laboratory. And finally, since we know from observations of the CMB that the universe is flat, the rest, up to $\Omega_0 = 1$, must be a diffuse vacuum energy, which affects the very large scales and late times, and seems to be responsible for the present acceleration of the universe, see Section 3. Nowadays, multiple observations seem to converge towards a common determination of $\Omega_M = 0.25 \pm 0.08$ (95% c.l.), see Fig. 18.

3.27 Massive neutrinos

One of the ‘usual suspects’ when addressing the problem of dark matter are neutrinos. They are the only candidates known to exist. If neutrinos have a mass, could they constitute the missing matter? We know from the Big Bang theory, see Section 2.6.5, that there is a cosmic neutrino background at a temperature of approximately 2K. This allows one to compute the present number density in the form of neutrinos, which turns out to be, for massless neutrinos, $n_\nu(T_\nu) = \frac{3}{11} n_\gamma(T_\gamma) = 112 \text{ cm}^{-3}$, per species of neutrino.

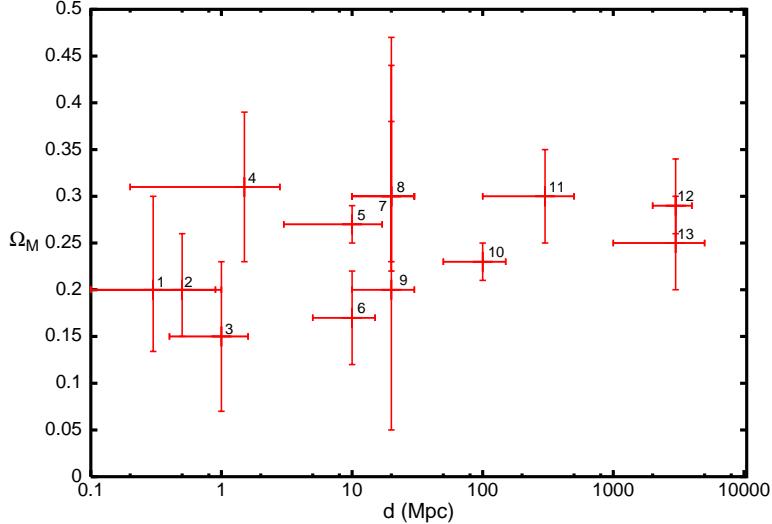


Fig. 18: Different determinations of Ω_M as a function of distance, from various sources: 1. peculiar velocities; 2. weak gravitational lensing; 3. shear autocorrelation function; 4. local group of galaxies; 5. baryon mass fraction; 6. cluster mass function; 7. virgocentric flow; 8. mean relative velocities; 9. redshift space distortions; 10. mass power spectrum; 11. integrated Sachs-Wolfe effect; 12. angular diameter distance: SNe; 13. cluster baryon fraction. While a few years ago the dispersion among observed values was huge and strongly dependent on scale, at present the observed value of the matter density parameter falls well within a narrow range, $\Omega_M = 0.25 \pm 0.07$ (95% c.l.) and is essentially independent on scale, from 100 kpc to 5000 Mpc. Adapted from Ref. [45].

If neutrinos have mass, as recent experiments seem to suggest,⁶ see Fig. 19, the cosmic energy density in massive neutrinos would be $\rho_\nu = \sum n_\nu m_\nu = \frac{3}{11} n_\gamma \sum m_\nu$, and therefore its contribution today,

$$\Omega_\nu h^2 = \frac{\sum m_\nu}{93.2 \text{ eV}}. \quad (116)$$

The discussion in the previous Sections suggest that $\Omega_M \leq 0.4$, and thus, for any of the three families of neutrinos, $m_\nu \leq 40$ eV. Note that this limit improves by six orders of magnitude the present bound on the tau-neutrino mass [19]. Supposing that the missing mass in non-baryonic cold dark matter arises from a single particle dark matter (PDM) component, its contribution to the critical density is bounded by $0.05 \leq \Omega_{\text{PDM}} h^2 \leq 0.4$, see Fig. 17.

I will now go through the various logical arguments that exclude neutrinos as the *dominant* component of the missing dark matter in the universe. Is it possible that neutrinos with a mass $4 \text{ eV} \leq m_\nu \leq 40$ eV be the non-baryonic PDM component? For instance, could massive neutrinos constitute the dark matter halos of galaxies? For neutrinos to be gravitationally bound to galaxies it is necessary that their velocity be less than the escape velocity v_{esc} , and thus their maximum momentum is $p_{\text{max}} = m_\nu v_{\text{esc}}$. How many neutrinos can be packed in the halo of a galaxy? Due to the Pauli exclusion principle, the maximum number density is given by that of a completely degenerate Fermi gas with momentum $p_F = p_{\text{max}}$, i.e. $n_{\text{max}} = p_{\text{max}}^3 / 3\pi^2$. Therefore, the maximum local density in dark matter neutrinos is $\rho_{\text{max}} = n_{\text{max}} m_\nu = m_\nu^4 v_{\text{esc}}^3 / 3\pi^2$, which must be greater than the typical halo density $\rho_{\text{halo}} = 0.3 \text{ GeV cm}^{-3}$. For a typical spiral galaxy, this constraint, known as the Tremaine-Gunn limit, gives $m_\nu \geq 40$ eV, see Ref. [47]. However, this mass, even for a single species, say the tau-neutrino, gives a value for $\Omega_\nu h^2 = 0.5$, which is far too high for structure formation. Neutrinos of such a low mass would constitute a relativistic hot dark matter component, which would wash-out structure below the supercluster scale, against evidence from present observations, see Fig. 19. Furthermore, apply-

⁶For a review on Neutrino properties, see González-García's lectures on these Proceedings.

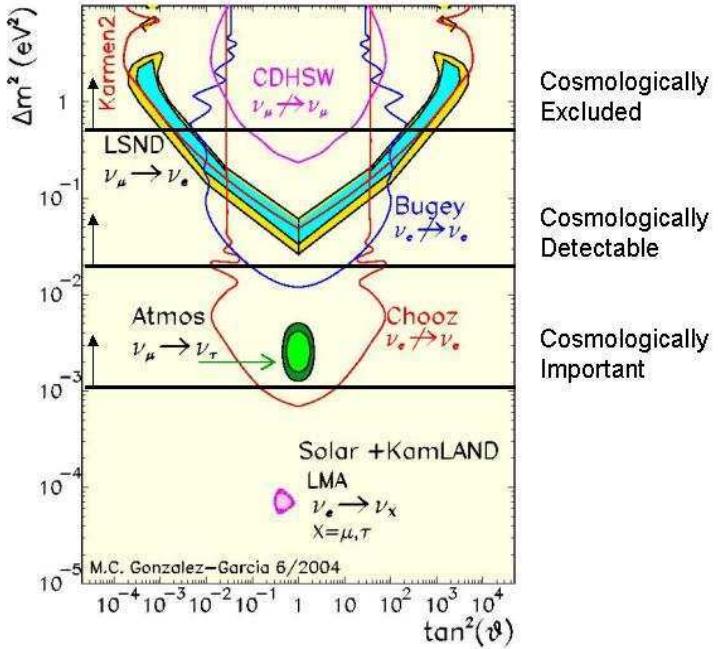


Fig. 19: The neutrino parameter space, mixing angle against Δm^2 , including the results from the different solar and atmospheric neutrino oscillation experiments. Note the threshold of cosmologically important masses, cosmologically detectable neutrinos (by CMB and LSS observations), and cosmologically excluded range of masses. Adapted from Refs. [46] and [91].

ing the same phase-space argument to the neutrinos as dark matter in the halo of dwarf galaxies gives $m_\nu \geq 100$ eV, beyond closure density (116). We must conclude that the simple idea that light neutrinos could constitute the particle dark matter on all scales is ruled out. They could, however, still play a role as a sub-dominant hot dark matter component in a flat CDM model. In that case, a neutrino mass of order 1 eV is not cosmological excluded, see Fig. 19.

Another possibility is that neutrinos have a large mass, of order a few GeV. In that case, their number density at decoupling, see Section 2.5.1, is suppressed by a Boltzmann factor, $\sim \exp(-m_\nu/T_{\text{dec}})$. For masses $m_\nu > T_{\text{dec}} \simeq 0.8$ MeV, the present energy density has to be computed as a solution of the corresponding Boltzmann equation. Apart from a logarithmic correction, one finds $\Omega_\nu h^2 \simeq 0.1(10 \text{ GeV}/m_\nu)^2$ for Majorana neutrinos and slightly smaller for Dirac neutrinos. In either case, neutrinos could be the dark matter only if their mass was a few GeV. Laboratory limits for ν_τ of around 18 MeV [19], and much more stringent ones for ν_μ and ν_e , exclude the known light neutrinos. However, there is always the possibility of a fourth unknown heavy and stable (perhaps sterile) neutrino. If it couples to the Z boson and has a mass below 45 GeV for Dirac neutrinos (39.5 GeV for Majorana neutrinos), then it is ruled out by measurements at LEP of the invisible width of the Z. There are two logical alternatives, either it is a sterile neutrino (it does not couple to the Z), or it does couple but has a larger mass. In the case of a Majorana neutrino (its own antiparticle), their abundance, for this mass range, is too small for being cosmologically relevant, $\Omega_\nu h^2 \leq 0.005$. If it were a Dirac neutrino there could be a lepton asymmetry, which may provide a higher abundance (similar to the case of baryogenesis). However, neutrinos scatter on nucleons via the weak axial-vector current (spin-dependent) interaction. For the small momentum transfers imparted by galactic WIMPs, such collisions are essentially coherent over an entire nucleus, leading to an enhancement of the effective cross section. The relatively large detection rate in this case allows one to exclude fourth-generation Dirac neutrinos for the galactic dark matter [48]. Anyway, it would be very implausible to have such a massive neutrino today, since it would have to be stable, with a life-time greater than the age of the universe, and there is no theoretical reason

to expect a massive sterile neutrino that does not oscillate into the other neutrinos.

Of course, the definitive test to the possible contribution of neutrinos to the overall density of the universe would be to measure *directly* their mass in laboratory experiments. There are at present two types of experiments: neutrino oscillation experiments, which measure only *differences* in squared masses, and direct mass-searches experiments, like the tritium β -spectrum and the neutrinoless double- β decay experiments, which measure directly the mass of the electron neutrino. The former experiments give a bound $m_{\nu_e} \lesssim 2.3$ eV (95% c.l.) [49], while the latter claim [50] they have a positive evidence for a Majorana neutrino of mass $m_\nu = 0.05 - 0.89$ eV (95% c.l.), although this result still awaits confirmation by other experiments. Neutrinos with such a mass could very well constitute the HDM component of the universe, $\Omega_{\text{HDM}} \lesssim 0.15$. The oscillation experiments give a range of possibilities for $\Delta m_\nu^2 = 0.3 - 3$ eV² from LSND (not yet confirmed by Miniboone), to the atmospheric neutrino oscillations from SuperKamiokande ($\Delta m_\nu^2 \simeq 2.2 \pm 0.5 \times 10^{-3}$ eV², $\tan^2 \theta = 1.0 \pm 0.3$) and the solar neutrino oscillations from KamLAND and the Sudbury Neutrino Observatory ($\Delta m_\nu^2 \simeq 8.2 \pm 0.3 \times 10^{-5}$ eV², $\tan^2 \theta = 0.39 \pm 0.05$), see Ref. [46]. Only the first two possibilities would be cosmologically relevant, see Fig. 19. Thanks to recent observations by WMAP, 2dFGRS and SDSS, we can put stringent limits on the absolute scale of neutrino masses, see below (Section 3.4).

3.28 Weakly Interacting Massive Particles

Unless we drastically change the theory of gravity on large scales, baryons cannot make up the bulk of the dark matter. Massive neutrinos are the only alternative among the known particles, but they are essentially ruled out as a universal dark matter candidate, even if they may play a subdominant role as a hot dark matter component. There remains the mystery of what is the physical nature of the dominant cold dark matter component. Something like a heavy stable neutrino, a generic Weakly Interacting Massive Particle (WIMP), could be a reasonable candidate because its present abundance could fall within the expected range,

$$\Omega_{\text{PDM}} h^2 \sim \frac{G^{3/2} T_0^3 h^2}{H_0^2 \langle \sigma_{\text{ann}} v_{\text{rel}} \rangle} = \frac{3 \times 10^{-27} \text{ cm}^3 \text{s}^{-1}}{\langle \sigma_{\text{ann}} v_{\text{rel}} \rangle}. \quad (117)$$

Here v_{rel} is the relative velocity of the two incoming dark matter particles and the brackets $\langle \cdot \rangle$ denote a thermal average at the freeze-out temperature, $T_f \simeq m_{\text{PDM}}/20$, when the dark matter particles go out of equilibrium with radiation. The value of $\langle \sigma_{\text{ann}} v_{\text{rel}} \rangle$ needed for $\Omega_{\text{PDM}} \approx 1$ is remarkably close to what one would expect for a WIMP with a mass $m_{\text{PDM}} = 100$ GeV, $\langle \sigma_{\text{ann}} v_{\text{rel}} \rangle \sim \alpha^2 / 8\pi m_{\text{PDM}} \sim 3 \times 10^{-27} \text{ cm}^3 \text{s}^{-1}$. We still do not know whether this is just a coincidence or an important hint on the nature of dark matter.

There are a few theoretical candidates for WIMPs, like the neutralino, coming from supersymmetric extensions of the standard model of particle physics,⁷ but at present there is no empirical evidence that such extensions are indeed realized in nature. In fact, the non-observation of supersymmetric particles at current accelerators places stringent limits on the neutralino mass and interaction cross section [52]. If WIMPs constitute the dominant component of the halo of our galaxy, it is expected that some may cross the Earth at a reasonable rate to be detected. The direct experimental search for them rely on elastic WIMP collisions with the nuclei of a suitable target. Dark matter WIMPs move at a typical galactic “virial” velocity of around 200 – 300 km/s, depending on the model. If their mass is in the range 10 – 100 GeV, the recoil energy of the nuclei in the elastic collision would be of order 10 keV. Therefore, one should be able to identify such energy depositions in a macroscopic sample of the target. There are at present three different methods: First, one could search for scintillation light in NaI crystals or in liquid xenon; second, search for an ionization signal in a semiconductor, typically a very pure germanium crystal; and third, use a cryogenic detector at 10 mK and search for a measurable temperature

⁷For a review of Supersymmetry (SUSY), see Kazakov’s contribution to these Proceedings.

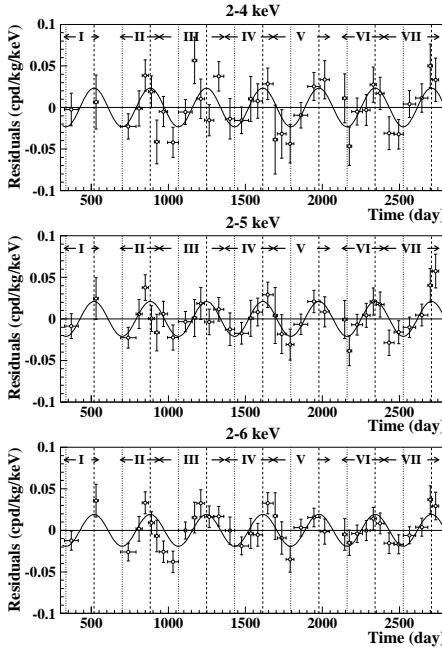


Fig. 20: The annual-modulation signal accumulated over 7 years is consistent with a neutralino of mass of $m_\chi = 59^{+17}_{-14}$ GeV and a proton cross section of $\xi\sigma_p = 7.0^{+0.4}_{-1.2} \times 10^{-6}$ pb, according to DAMA. From Ref. [51].

increase of the sample. The main problem with such a type of experiment is the low expected signal rate, with a typical number below 1 event/kg/day. To reduce natural radioactive contamination one must use extremely pure substances, and to reduce the background caused by cosmic rays requires that these experiments be located deeply underground.

The best limits on WIMP scattering cross sections come from some germanium experiments, like the Criogenic Dark Matter Search (CDMS) collaboration at Stanford and the Soudan mine [53], as well as from the NaI scintillation detectors of the UK dark matter collaboration (UKDMC) in the Boulby salt mine in England [54], and the DAMA experiment in the Gran Sasso laboratory in Italy [51]. Current experiments already touch the parameter space expected from supersymmetric particles, see Fig. 21, and therefore there is a chance that they actually discover the nature of the missing dark matter. The problem, of course, is to attribute a tentative signal unambiguously to galactic WIMPs rather than to some unidentified radioactive background.

One specific signature is the annual modulation which arises as the Earth moves around the Sun.⁸ Therefore, the net speed of the Earth relative to the galactic dark matter halo varies, causing a modulation of the expected counting rate. The DAMA/NaI experiment has actually reported such a modulation signal, from the combined analysis of their 7-year data, see Fig. 20 and Ref. [51], which provides a confidence level of 99.6% for a neutralino mass of $m_\chi = 52^{+10}_{-8}$ GeV and a proton cross section of $\xi\sigma_p = 7.2^{+0.4}_{-0.9} \times 10^{-6}$ pb, where $\xi = \rho_\chi/0.3$ GeV cm⁻³ is the local neutralino energy density in units of the galactic halo density. There has been no confirmation yet of this result from other dark matter search groups. In fact, the CDMS collaboration claims an exclusion of the DAMA region at the 3 sigma level, see Fig. 21. Hopefully in the near future we will have much better sensitivity at low masses from the Cryogenic Rare Event Search with Superconducting Thermometers (CRESST) experiment at Gran Sasso. The CRESST experiment [55] uses sapphire crystals as targets and a new method to simultaneously measure the phonons and the scintillating light from particle interactions inside

⁸The time scale of the Sun's orbit around the center of the galaxy is too large to be relevant in the analysis.

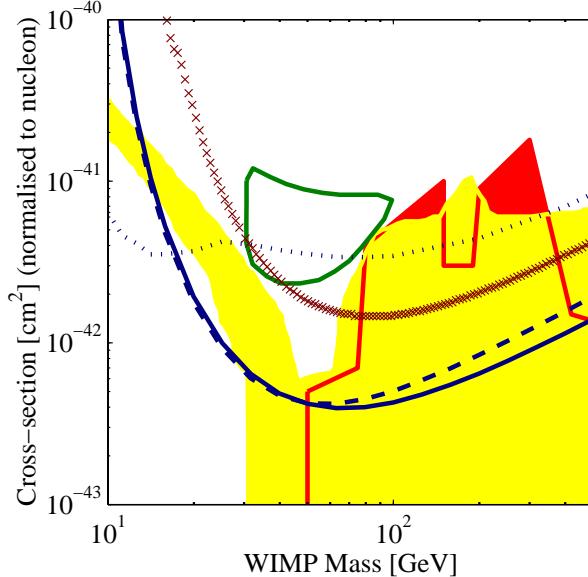


Fig. 21: Exclusion range for the spin-independent WIMP scattering cross section per nucleon from the NaI experiments and the Ge detectors. The blue lines come from the CDMS experiment, which exclude the DAMA region at more than 3 sigma. Also shown in yellow and red is the range of expected counting rates for neutralinos in the MSSM. From Ref. [53].

the crystal, which allows excellent background discrimination. Very recently there has been also the proposal of a completely new method based on a Superheated Droplet Detector (SDD), which claims to have already a similar sensitivity as the more standard methods described above, see Ref. [56].

There exist other *indirect* methods to search for galactic WIMPs [57]. Such particles could self-annihilate at a certain rate in the galactic halo, producing a potentially detectable background of high energy photons or antiprotons. The absence of such a background in both gamma ray satellites and the Alpha Matter Spectrometer [58] imposes bounds on their density in the halo. Alternatively, WIMPs traversing the solar system may interact with the matter that makes up the Earth or the Sun so that a small fraction of them will lose energy and be trapped in their cores, building up over the age of the universe. Their annihilation in the core would thus produce high energy neutrinos from the center of the Earth or from the Sun which are detectable by neutrino telescopes. In fact, SuperKamiokande already covers a large part of SUSY parameter space. In other words, neutrino telescopes are already competitive with direct search experiments. In particular, the AMANDA experiment at the South Pole [59], which has approximately 10^3 Cherenkov detectors several km deep in very clear ice, over a volume $\sim 1 \text{ km}^3$, is competitive with the best direct searches proposed. The advantages of AMANDA are also directional, since the arrays of Cherenkov detectors will allow one to reconstruct the neutrino trajectory and thus its source, whether it comes from the Earth or the Sun. AMANDA recently reported the detection of TeV neutrinos [59].

3.3 The age of the universe t_0

The universe must be older than the oldest objects it contains. Those are believed to be the stars in the oldest clusters in the Milky Way, globular clusters. The most reliable ages come from the application of theoretical models of stellar evolution to observations of old stars in globular clusters. For about 30 years, the ages of globular clusters have remained reasonable stable, at about 15 Gyr [60]. However, recently these ages have been revised downward [61].

During the 1980s and 1990s, the globular cluster age estimates have improved as both new obser-

vations have been made with CCDs, and since refinements to stellar evolution models, including opacities, consideration of mixing, and different chemical abundances have been incorporated [62]. From the theory side, uncertainties in globular cluster ages come from uncertainties in convection models, opacities, and nuclear reaction rates. From the observational side, uncertainties arise due to corrections for dust and chemical composition. However, the dominant source of systematic errors in the globular cluster age is the uncertainty in the cluster distances. Fortunately, the Hipparcos satellite recently provided geometric parallax measurements for many nearby old stars with low metallicity, typical of globular clusters, thus allowing for a new calibration of the ages of stars in globular clusters, leading to a downward revision to 10 – 13 Gyr [62]. Moreover, there were very few stars in the Hipparcos catalog with both small parallax errors and low metal abundance. Hence, an increase in the sample size could be critical in reducing the statistical uncertainties for the calibration of the globular cluster ages. There are already proposed two new parallax satellites, NASA’s Space Interferometry Mission (SIM) and ESA’s mission, called GAIA, that will give 2 or 3 orders of magnitude more accurate parallaxes than Hipparcos, down to fainter magnitude limits, for several orders of magnitude more stars. Until larger samples are available, however, distance errors are likely to be the largest source of systematic uncertainty to the globular cluster age [29].

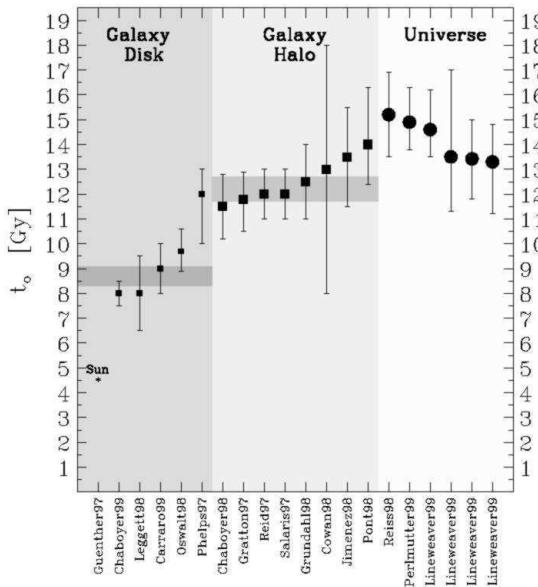


Fig. 22: The recent estimates of the age of the universe and that of the oldest objects in our galaxy. The last three points correspond to the combined analysis of 8 different measurements, for $h = 0.64, 0.68$ and 7.2 , which indicates a relatively weak dependence on h . The age of the Sun is accurately known and is included for reference. Error bars indicate 1σ limits. The averages of the ages of the Galactic Halo and Disk are shaded in gray. Note that there isn’t a single age estimate more than 2σ away from the average. The result $t_0 > t_{\text{gal}}$ is logically inevitable, but the standard EdS model does not satisfy this unless $h < 0.55$. From Ref. [63].

The supernovae groups can also determine the age of the universe from their high redshift observations. The high confidence regions in the $(\Omega_M, \Omega_\Lambda)$ plane are almost parallel to the contours of constant age. For any value of the Hubble constant less than $H_0 = 70$ km/s/Mpc, the implied age of the universe is greater than 13 Gyr, allowing enough time for the oldest stars in globular clusters to evolve [62]. Integrating over Ω_M and Ω_Λ , the best fit value of the age in Hubble-time units is $H_0 t_0 = 0.93 \pm 0.06$ or equivalently $t_0 = 14.1 \pm 1.0$ ($0.65 h^{-1}$) Gyr, see Ref. [7]. Furthermore, a combination of 8 independent recent measurements: CMB anisotropies, type Ia SNe, cluster mass-to-light ratios, cluster abundance evolution, cluster baryon fraction, deuterium-to-hydrogen ratios in quasar spectra, double-lobed radio sources and the Hubble constant, can be used to determine the present age of the universe [63]. The

result is shown in Fig. 22, compared to other recent determinations. The best fit value for the age of the universe is, according to this analysis, $t_0 = 13.4 \pm 1.6$ Gyr, about a billion years younger than other recent estimates [63].

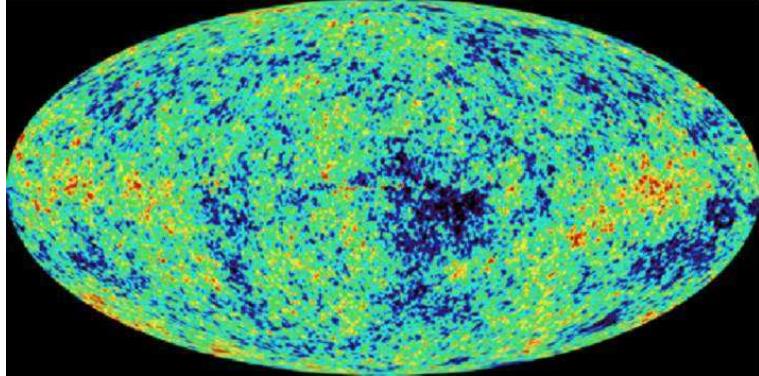


Fig. 23: The anisotropies of the microwave background measured by the WMAP satellite with 10 arcminute resolution. It shows the intrinsic CMB anisotropies at the level of a few parts in 10^5 . The galactic foreground has been properly subtracted. The amount of information contained in this map is enough to determine most of the cosmological parameters to few percent accuracy. From Ref. [20].

3.4 Cosmic Microwave Background Anisotropies

The cosmic microwave background has become in the last five years the Holy Grail of Cosmology, since precise observations of the temperature and polarization anisotropies allow in principle to determine the parameters of the Standard Model of Cosmology with very high accuracy. Recently, the WMAP satellite has provided with a very detailed map of the microwave anisotropies in the sky, see Fig. 23, and indeed has fulfilled our expectations, see Table 2.

The physics of the CMB anisotropies is relatively simple [64]. The universe just before recombination is a very tightly coupled fluid, due to the large electromagnetic Thomson cross section $\sigma_T = 8\pi\alpha^2/3m_e^2 \simeq 0.7$ barn. Photons scatter off charged particles (protons and electrons), and carry energy, so they feel the gravitational potential associated with the perturbations imprinted in the metric during inflation. An overdensity of baryons (protons and neutrons) does not collapse under the effect of gravity until it enters the causal Hubble radius. The perturbation continues to grow until radiation pressure opposes gravity and sets up acoustic oscillations in the plasma, very similar to sound waves. Since overdensities of the same size will enter the Hubble radius at the same time, they will oscillate in phase. Moreover, since photons scatter off these baryons, the acoustic oscillations occur also in the photon field and induces a pattern of peaks in the temperature anisotropies in the sky, at different angular scales, see Fig. 24. There are three different effects that determine the temperature anisotropies we observe in the CMB. First, *gravity*: photons fall in and escape off gravitational potential wells, characterized by Φ in the comoving gauge, and as a consequence their frequency is gravitationally blue- or red-shifted, $\delta\nu/\nu = \Phi$. If the gravitational potential is not constant, the photons will escape from a larger or smaller potential well than they fell in, so their frequency is also blue- or red-shifted, a phenomenon known as the Rees-Sciama effect. Second, *pressure*: photons scatter off baryons which fall into gravitational potential wells and the two competing forces create acoustic waves of compression and rarefaction. Finally, *velocity*: baryons accelerate as they fall into potential wells. They have minimum velocity at maximum compression and rarefaction. That is, their velocity wave is exactly 90° off-phase with the acoustic waves. These waves induce a Doppler effect on the frequency of the photons. The temperature anisotropy induced by

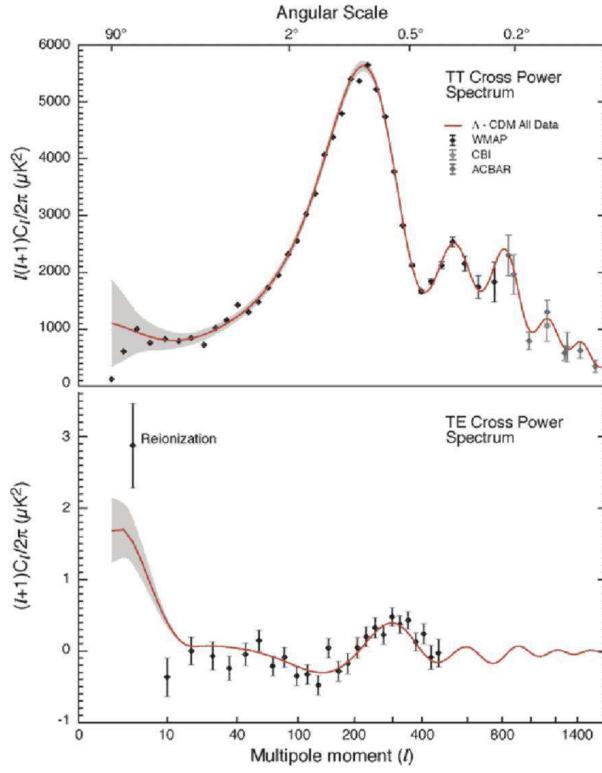


Fig. 24: The Angular Power Spectrum of CMB temperature anisotropies, compared with the cross-correlation of temperature-polarization anisotropies. From Ref. [20].

these three effects is therefore given by [64]

$$\frac{\delta T}{T}(\mathbf{r}) = \Phi(\mathbf{r}, t_{\text{dec}}) + 2 \int_{t_{\text{dec}}}^{t_0} \dot{\Phi}(\mathbf{r}, t) dt + \frac{1}{3} \frac{\delta \rho}{\rho} - \frac{\mathbf{r} \cdot \mathbf{v}}{c}. \quad (118)$$

Metric perturbations of different wavelengths enter the horizon at different times. The largest wavelengths, of size comparable to our present horizon, are entering now. There are perturbations with wavelengths comparable to the size of the horizon at the time of last scattering, of projected size about 1° in the sky today, which entered precisely at decoupling. And there are perturbations with wavelengths much smaller than the size of the horizon at last scattering, that entered much earlier than decoupling, all the way to the time of radiation-matter equality, which have gone through several acoustic oscillations before last scattering. All these perturbations of different wavelengths leave their imprint in the CMB anisotropies.

The baryons at the time of decoupling do not feel the gravitational attraction of perturbations with wavelength greater than the size of the horizon at last scattering, because of causality. Perturbations with exactly that wavelength are undergoing their first contraction, or acoustic compression, at decoupling. Those perturbations induce a large peak in the temperature anisotropies power spectrum, see Fig. 24. Perturbations with wavelengths smaller than these will have gone, after they entered the Hubble scale, through a series of acoustic compressions and rarefactions, which can be seen as secondary peaks in the power spectrum. Since the surface of last scattering is not a sharp discontinuity, but a region of $\Delta z \sim 100$, there will be scales for which photons, travelling from one energy concentration to another, will erase the perturbation on that scale, similarly to what neutrinos or HDM do for structure on small scales. That is the reason why we don't see all the acoustic oscillations with the same amplitude, but in fact they decay exponentially towards smaller angular scales, an effect known as Silk damping, due to photon diffusion [65, 64].

Table 2: **The parameters of the standard cosmological model.** The standard model of cosmology has about 20 different parameters, needed to describe the background space-time, the matter content and the spectrum of metric perturbations. We include here the present range of the most relevant parameters (with 1σ errors), as recently determined by WMAP, and the error with which the Planck satellite will be able to determine them in the near future. The rate of expansion is written in units of $H = 100 h \text{ km/s/Mpc}$.

physical quantity	symbol	WMAP	Planck
total density	Ω_0	1.02 ± 0.02	0.7%
baryonic matter	Ω_B	0.044 ± 0.004	0.6%
cosmological constant	Ω_Λ	0.73 ± 0.04	0.5%
cold dark matter	Ω_M	0.23 ± 0.04	0.6%
hot dark matter	$\Omega_\nu h^2$	< 0.0076 (95% c.l.)	1%
sum of neutrino masses	$\sum m_\nu$ (eV)	< 0.23 (95% c.l.)	1%
CMB temperature	T_0 (K)	2.725 ± 0.002	0.1%
baryon to photon ratio	$\eta \times 10^{10}$	6.1 ± 0.3	0.5%
baryon to matter ratio	Ω_B/Ω_M	0.17 ± 0.01	1%
spatial curvature	Ω_K	< 0.02 (95% c.l.)	0.5%
rate of expansion	h	0.71 ± 0.03	0.8%
age of the universe	t_0 (Gyr)	13.7 ± 0.2	0.1%
age at decoupling	t_{dec} (kyr)	379 ± 8	0.5%
age at reionization	t_r (Myr)	180 ± 100	5%
spectral amplitude	A	0.833 ± 0.085	0.1%
spectral tilt	n_s	0.98 ± 0.03	0.2%
spectral tilt variation	$dn_s/d \ln k$	-0.031 ± 0.017	0.5%
tensor-scalar ratio	r	< 0.71 (95% c.l.)	5%
reionization optical depth	τ	0.17 ± 0.04	5%
redshift of equality	z_{eq}	3233 ± 200	5%
redshift of decoupling	z_{dec}	1089 ± 1	0.1%
width of decoupling	Δz_{dec}	195 ± 2	1%
redshift of reionization	z_r	20 ± 10	2%

From the observations of the CMB anisotropies it is possible to determine most of the parameters of the Standard Cosmological Model with few percent accuracy, see Table 2. However, there are many degeneracies between parameters and it is difficult to disentangle one from another. For instance, as mentioned above, the first peak in the photon distribution corresponds to overdensities that have undergone half an oscillation, that is, a compression, and appear at a scale associated with the size of the horizon at last scattering, about 1° projected in the sky today. Since photons scatter off baryons, they will also feel the acoustic wave and create a peak in the correlation function. The height of the peak is proportional to the amount of baryons: the larger the baryon content of the universe, the higher the peak. The position of the peak in the power spectrum depends on the geometrical size of the particle horizon at last scattering. Since photons travel along geodesics, the projected size of the causal horizon at decoupling depends on whether the universe is flat, open or closed. In a flat universe the geodesics are straight lines and, by looking at the angular scale of the first acoustic peak, we would be measuring the actual size of the horizon at last scattering. In an open universe, the geodesics are inward-curved trajectories, and therefore the projected size on the sky appears smaller. In this case, the first acoustic peak should occur at higher multipoles or smaller angular scales. On the other hand, for a closed universe, the first peak occurs at smaller multipoles or larger angular scales. The dependence of the position of

the first acoustic peak on the spatial curvature can be approximately given by $l_{\text{peak}} \simeq 220 \Omega_0^{-1/2}$, where $\Omega_0 = \Omega_M + \Omega_\Lambda = 1 - \Omega_K$. Present observations by WMAP and other experiments give $\Omega_0 = 1.00 \pm 0.02$ at one standard deviation [20].

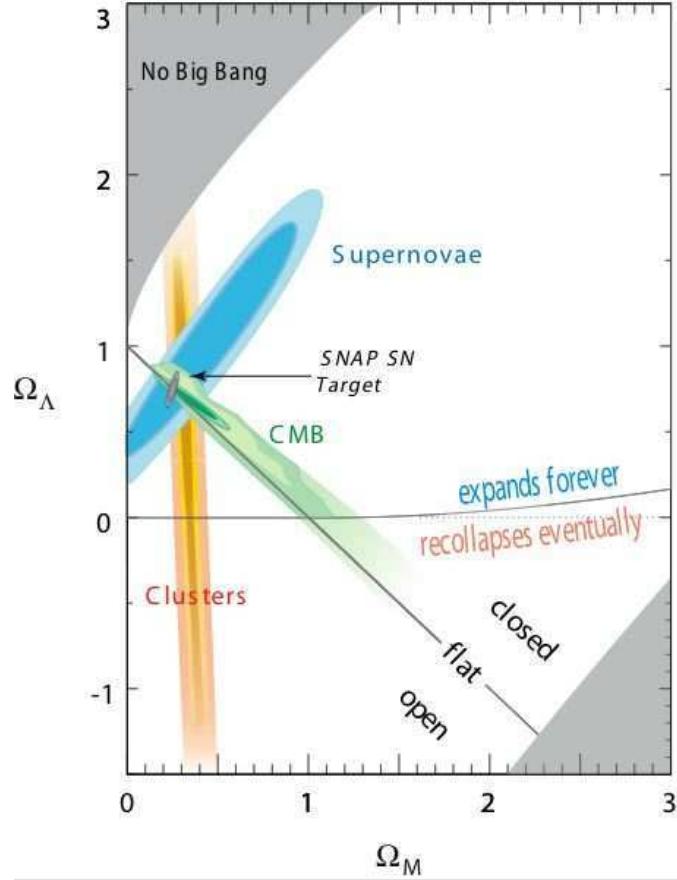


Fig. 25: The $(\Omega_M, \Omega_\Lambda)$ plane with the present data set of cosmological observations – the acceleration of the universe, the large scale structure and the CMB anisotropies – as well as the future determinations by SNAP and Planck of the fundamental parameters which define our Standard Model of Cosmology.

The other acoustic peaks occur at harmonics of this, corresponding to smaller angular scales. Since the amplitude and position of the primary and secondary peaks are directly determined by the sound speed (and, hence, the equation of state) and by the geometry and expansion of the universe, they can be used as a powerful test of the density of baryons and dark matter, and other cosmological parameters. With the joined data from WMAP, VSA, CBI and ACBAR, we have rather good evidence of the existence of the second and third acoustic peaks, which confirms one of the most important predictions of inflation – the non-causal origin of the primordial spectrum of perturbations –, and rules out cosmological defects as the dominant source of structure in the universe [66]. Moreover, since the observations of CMB anisotropies now cover almost three orders of magnitude in the size of perturbations, we can determine the much better accuracy the value of the spectral tilt, $n = 0.98 \pm 0.03$, which is compatible with the approximate scale invariant spectrum needed for structure formation, and is a prediction of the simplest models of inflation. Soon after the release of data from WMAP, there was some expectation at the claim of a scale-dependent tilt. Nowadays, with better resolution in the linear matter power spectrum from SDSS [67], we can not conclude that the spectral tilt has any observable dependence on scale.

The microwave background has become also a testing ground for theories of particle physics. In

particular, it already gives stringent constraints on the mass of the neutrino, when analysed together with large scale structure observations. Assuming a flat Λ CDM model, the 2-sigma upper bounds on the sum of the masses of light neutrinos is $\sum m_\nu < 1.0$ eV for degenerate neutrinos (i.e. without a large hierarchy between them) if we don't impose any priors, and it comes down to $\sum m_\nu < 0.6$ eV if one imposes the bounds coming from the HST measurements of the rate of expansion and the supernova data on the present acceleration of the universe [68]. The final bound on the neutrino density can be expressed as $\Omega_\nu h^2 = \sum m_\nu / 93.2$ eV ≤ 0.01 . In the future, both with Planck and with the Atacama Cosmology Telescope (ACT) we will be able to put constraints on the neutrino masses down to the 0.1 eV level.

Moreover, the present data is good enough that we can start to put constraints on the models of inflation that give rise to structure. In particular, multifield models of inflation predict a mixture of adiabatic and isocurvature perturbations,⁹ and their signatures in the cosmic microwave background anisotropies and the matter power spectrum of large scale structure are specific and perfectly distinguishable. Nowadays, thanks to precise CMB, LSS and SNIa data, one can put rather stringent limits on the relative fraction and correlation of the isocurvature modes to the dominant adiabatic perturbations [69].

We can summarize this Section by showing the region in parameter space where we stand nowadays, thanks to the recent cosmological observations. We have plotted that region in Fig. 25. One could also superimpose the contour lines corresponding to equal $t_0 H_0$ lines, as a cross check. It is extraordinary that only in the last few months we have been able to reduce the concordance region to where it stands today, where all the different observations seem to converge. There are still many uncertainties, mainly systematic; however, those are quickly decreasing and becoming predominantly statistical. In the near future, with precise observations of the anisotropies in the microwave background temperature and polarization anisotropies, thanks to Planck satellite, we will be able to reduce those uncertainties to the level of one percent. This is the reason why cosmologists are so excited and why it is claimed that we live in the Golden Age of Cosmology.

4. THE INFLATIONARY PARADIGM

The hot Big Bang theory is nowadays a very robust edifice, with many independent observational checks: the expansion of the universe; the abundance of light elements; the cosmic microwave background; a predicted age of the universe compatible with the age of the oldest objects in it, and the formation of structure via gravitational collapse of initially small inhomogeneities. Today, these observations are confirmed to within a few percent accuracy, and have helped establish the hot Big Bang as the preferred model of the universe. All the physics involved in the above observations is routinely tested in the laboratory (atomic and nuclear physics experiments) or in the solar system (general relativity).

However, this theory leaves a range of crucial questions unanswered, most of which are initial conditions' problems. There is the reasonable assumption that these cosmological problems will be solved or explained by *new physical principles* at high energies, in the early universe. This assumption leads to the natural conclusion that accurate observations of the present state of the universe may shed light onto processes and physical laws at energies above those reachable by particle accelerators, present or future. We will see that this is a very optimistic approach indeed, and that there are many unresolved issues related to those problems. However, there might be in the near future reasons to be optimistic.

4.1 Shortcomings of Big Bang Cosmology

The Big Bang theory could not explain the origin of matter and structure in the universe; that is, the origin of the matter–antimatter asymmetry, without which the universe today would be filled by a uniform radiation continuously expanding and cooling, with no traces of matter, and thus without the possibility to form gravitationally bound systems like galaxies, stars and planets that could sustain life. Moreover,

⁹This mixture is generic, unless all the fields thermalize simultaneously at reheating, just after inflation, in which case the entropy perturbations that would give rise to the isocurvature modes disappear.

the standard Big Bang theory assumes, but cannot explain, the origin of the extraordinary smoothness and flatness of the universe on the very large scales seen by the microwave background probes and the largest galaxy catalogs. It cannot explain the origin of the primordial density perturbations that gave rise to cosmic structures like galaxies, clusters and superclusters, via gravitational collapse; the quantity and nature of the dark matter that we believe holds the universe together; nor the origin of the Big Bang itself.

A summary [10] of the problems that the Big Bang theory cannot explain is:

- The global structure of the universe.
 - Why is the universe so close to spatial flatness?
 - Why is matter so homogeneously distributed on large scales?
- The origin of structure in the universe.
 - How did the primordial spectrum of density perturbations originate?
- The origin of matter and radiation.
 - Where does all the energy in the universe come from?
 - What is the nature of the dark matter in the universe?
 - How did the matter-antimatter asymmetry arise?
- The initial singularity.
 - Did the universe have a beginning?
 - What is the global structure of the universe beyond our observable patch?

Let me discuss one by one the different issues:

4.11 The Flatness Problem

The Big Bang theory assumes but cannot explain the extraordinary spatial flatness of our local patch of the universe. In the general FRW metric (2) the parameter K that characterizes spatial curvature is a free parameter. There is nothing in the theory that determines this parameter *a priori*. However, it is directly related, via the Friedmann equation (8), to the dynamics, and thus the matter content, of the universe,

$$K = \frac{8\pi G}{3}\rho a^2 - H^2 a^2 = \frac{8\pi G}{3}\rho a^2 \left(\frac{\Omega - 1}{\Omega} \right). \quad (119)$$

We can therefore define a new variable,

$$x \equiv \frac{\Omega - 1}{\Omega} = \frac{\text{const.}}{\rho a^2}, \quad (120)$$

whose time evolution is given by

$$x' = \frac{dx}{dN} = (1 + 3\omega)x, \quad (121)$$

where $N = \ln(a/a_i)$ characterizes the *number of e-folds* of universe expansion ($dN = Hdt$) and where we have used Eq. (7) for the time evolution of the total energy, ρa^3 , which only depends on the barotropic ratio ω . It is clear from Eq. (121) that the phase-space diagram (x, x') presents an unstable critical (saddle) point at $x = 0$ for $\omega > -1/3$, i.e. for the radiation ($\omega = 1/3$) and matter ($\omega = 0$) eras. A small perturbation from $x = 0$ will drive the system towards $x = \pm\infty$. Since we know the universe went through both the radiation era (because of primordial nucleosynthesis) and the matter era (because of structure formation), tiny deviations from $\Omega = 1$ would have grown since then, such that today

$$x_0 = \frac{\Omega_0 - 1}{\Omega_0} = x_{\text{in}} \left(\frac{T_{\text{in}}}{T_{\text{eq}}} \right)^2 (1 + z_{\text{eq}}). \quad (122)$$

In order that today's value be in the range $0.1 < \Omega_0 < 1.2$, or $x_0 \approx \mathcal{O}(1)$, it is required that at, say, primordial nucleosynthesis ($T_{\text{NS}} \simeq 10^6 T_{\text{eq}}$) its value be

$$\Omega(t_{\text{NS}}) = 1 \pm 10^{-15}, \quad (123)$$

which represents a tremendous finetuning. Perhaps the universe indeed started with such a peculiar initial condition, but it is epistemologically more satisfying if we give a fundamental dynamical reason for the universe to have started so close to spatial flatness. These arguments were first used by Robert Dicke in the 1960s, much before inflation. He argued that the most natural initial condition for the spatial curvature should have been the Planck scale curvature, $(^3R = 6K/l_P^2)$, where the Planck length is $l_P = (\hbar G/c^3)^{1/2} = 1.62 \times 10^{-33}$ cm, that is, 60 orders of magnitude smaller than the present size of the universe, $a_0 = 1.38 \times 10^{28}$ cm. A universe with this immense curvature would have collapsed within a Planck time, $t_P = (\hbar G/c^5)^{1/2} = 5.39 \times 10^{-44}$ s, again 60 orders of magnitude smaller than the present age of the universe, $t_0 = 4.1 \times 10^{17}$ s. Therefore, the flatness problem is also related to the Age Problem, why is it that the universe is so old and flat when, under ordinary circumstances (based on the fundamental scale of gravity) it should have lasted only a Planck time and reached a size of order the Planck length? As we will see, inflation gives a dynamical reason to such a peculiar initial condition.

4.12 The Homogeneity Problem

An expanding universe has *particle horizons*, that is, spatial regions beyond which causal communication cannot occur. The horizon distance can be defined as the maximum distance that light could have travelled since the origin of the universe [15],

$$d_H(t) \equiv a(t) \int_0^t \frac{dt'}{a(t')} \sim H^{-1}(t), \quad (124)$$

which is proportional to the Hubble scale.¹⁰ For instance, at the beginning of nucleosynthesis the horizon distance is a few light-seconds, but grows *linearly* with time and by the end of nucleosynthesis it is a few light-minutes, i.e. a factor 100 larger, while the scale factor has increased *only* a factor of 10. The fact that the causal horizon increases faster, $d_H \sim t$, than the scale factor, $a \sim t^{1/2}$, implies that at any given time the universe contains regions within itself that, according to the Big Bang theory, were *never* in causal contact before. For instance, the number of causally disconnected regions at a given redshift z present in our causal volume today, $d_H(t_0) \equiv a_0$, is

$$N_{CD}(z) \sim \left(\frac{a(t)}{d_H(t)} \right)^3 \simeq (1+z)^{3/2}, \quad (125)$$

which, for the time of decoupling, is of order $N_{CD}(z_{dec}) \sim 10^5 \gg 1$.

This phenomenon is particularly acute in the case of the observed microwave background. Information cannot travel faster than the speed of light, so the causal region at the time of photon decoupling could not be larger than $d_H(t_{dec}) \sim 3 \times 10^5$ light years across, or about 1° projected in the sky today. So why should regions that are separated by more than 1° in the sky today have exactly the same temperature, to within 10 ppm, when the photons that come from those two distant regions could not have been in causal contact when they were emitted? This constitutes the so-called horizon problem, see Fig. 26, and was first discussed by Robert Dicke in the 1970s as a profound inconsistency of the Big Bang theory.

4.2 Cosmological Inflation

In the 1980s, a new paradigm, deeply rooted in fundamental physics, was put forward by Alan H. Guth [71], Andrei D. Linde [72] and others [73, 74, 75], to address these fundamental questions. According to the inflationary paradigm, the early universe went through a period of exponential expansion, driven by the approximately constant energy density of a scalar field called the inflaton. In modern physics, elementary particles are represented by quantum fields, which resemble the familiar electric, magnetic and gravitational fields. A field is simply a function of space and time whose quantum oscillations are interpreted as particles. In our case, the inflaton field has, associated with it, a large potential

¹⁰For the radiation era, the horizon distance is equal to the Hubble scale. For the matter era it is twice the Hubble scale.

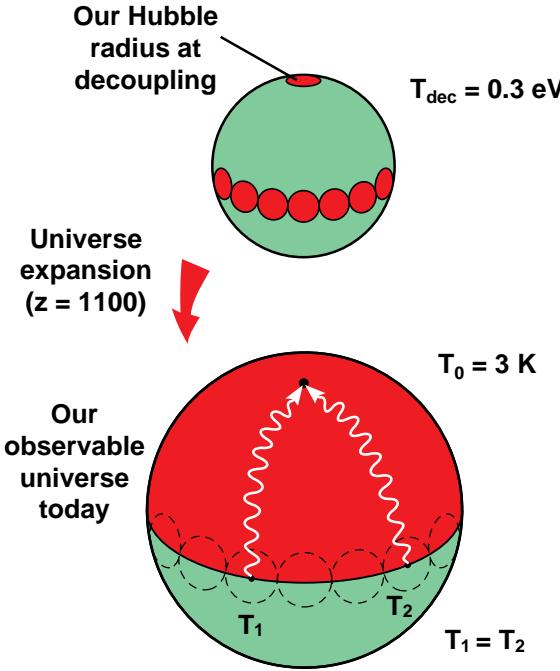


Fig. 26: Perhaps the most acute problem of the Big Bang theory is explaining the extraordinary homogeneity and isotropy of the microwave background, see Fig. 10. At the time of decoupling, the volume that gave rise to our present universe contained many causally disconnected regions (top figure). Today we observe a blackbody spectrum of photons coming from those regions and they appear to have the same temperature, $T_1 = T_2$, to one part in 10^5 . Why is the universe so homogeneous? This constitutes the so-called horizon problem, which is spectacularly solved by inflation. From Ref. [70].

energy density, which drives the exponential expansion during inflation, see Fig. 27. We know from general relativity that the density of matter determines the expansion of the universe, but a constant energy density acts in a very peculiar way: as a repulsive force that makes any two points in space separate at exponentially large speeds. (This does not violate the laws of causality because there is no information carried along in the expansion, it is simply the stretching of space-time.)

This superluminal expansion is capable of explaining the large scale homogeneity of our observable universe and, in particular, why the microwave background looks so isotropic: regions separated today by more than 1° in the sky were, in fact, in causal contact before inflation, but were stretched to cosmological distances by the expansion. Any inhomogeneities present before the tremendous expansion would be washed out. This explains why photons from supposedly causally disconnected regions have actually the same spectral distribution with the same temperature, see Fig. 26.

Moreover, in the usual Big Bang scenario a flat universe, one in which the gravitational attraction of matter is exactly balanced by the cosmic expansion, is unstable under perturbations: a small deviation from flatness is amplified and soon produces either an empty universe or a collapsed one. As we discussed above, for the universe to be nearly flat today, it must have been extremely flat at nucleosynthesis, deviations not exceeding more than one part in 10^{15} . This extreme fine tuning of initial conditions was also solved by the inflationary paradigm, see Fig. 28. Thus inflation is an extremely elegant hypothesis that explains how a region much, much greater than our own observable universe could have become smooth and flat without recourse to *ad hoc* initial conditions. Furthermore, inflation dilutes away any “unwanted” relic species that could have remained from early universe phase transitions, like monopoles, cosmic strings, etc., which are predicted in grand unified theories and whose energy density could be so large that the universe would have become unstable, and collapsed, long ago. These relics are diluted by

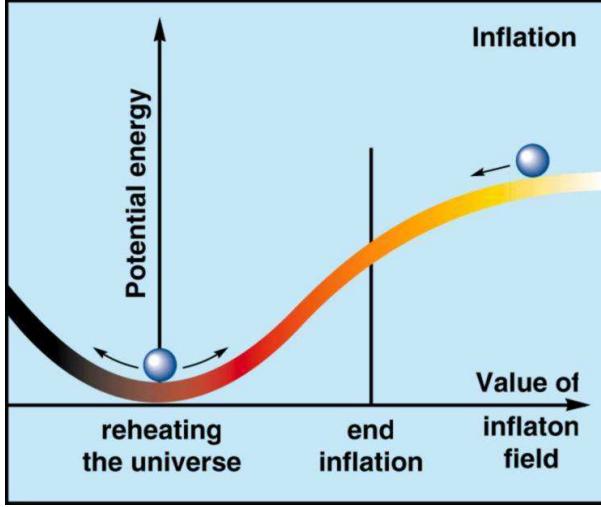


Fig. 27: The inflaton field can be represented as a ball rolling down a hill. During inflation, the energy density is approximately constant, driving the tremendous expansion of the universe. When the ball starts to oscillate around the bottom of the hill, inflation ends and the inflaton energy decays into particles. In certain cases, the coherent oscillations of the inflaton could generate a resonant production of particles which soon thermalize, reheating the universe. From Ref. [70].

the superluminal expansion, which leaves at most one of these particles per causal horizon, making them harmless to the subsequent evolution of the universe.

The only thing we know about this peculiar scalar field, the *inflaton*, is that it has a mass and a self-interaction potential $V(\phi)$ but we ignore everything else, even the scale at which its dynamics determines the superluminal expansion. In particular, we still do not know the nature of the inflaton field itself, is it some new *fundamental* scalar field in the electroweak symmetry breaking sector, or is it just some *effective* description of a more fundamental high energy interaction? Hopefully, in the near future, experiments in particle physics might give us a clue to its nature. Inflation had its original inspiration in the Higgs field, the scalar field supposed to be responsible for the masses of elementary particles (quarks and leptons) and the breaking of the electroweak symmetry. Such a field has not been found yet, and its discovery at the future particle colliders would help understand one of the truly fundamental problems in physics, the origin of masses. If the experiments discover something completely new and unexpected, it would automatically affect the idea of inflation at a fundamental level.

4.21 Homogeneous scalar field dynamics

In this subsection I will describe the theoretical basis for the phenomenon of inflation. Consider a scalar field ϕ , a singlet under any given interaction, with an effective potential $V(\phi)$. The Lagrangian for such a field in a curved background is

$$\mathcal{L}_{\text{inf}} = \frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi), \quad (126)$$

whose evolution equation in a Friedmann-Robertson-Walker metric (2) and for a *homogeneous* field $\phi(t)$ is given by

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0, \quad (127)$$

where H is the rate of expansion, together with the Einstein equations,

$$H^2 = \frac{\kappa^2}{3} \left(\frac{1}{2} \dot{\phi}^2 + V(\phi) \right), \quad (128)$$

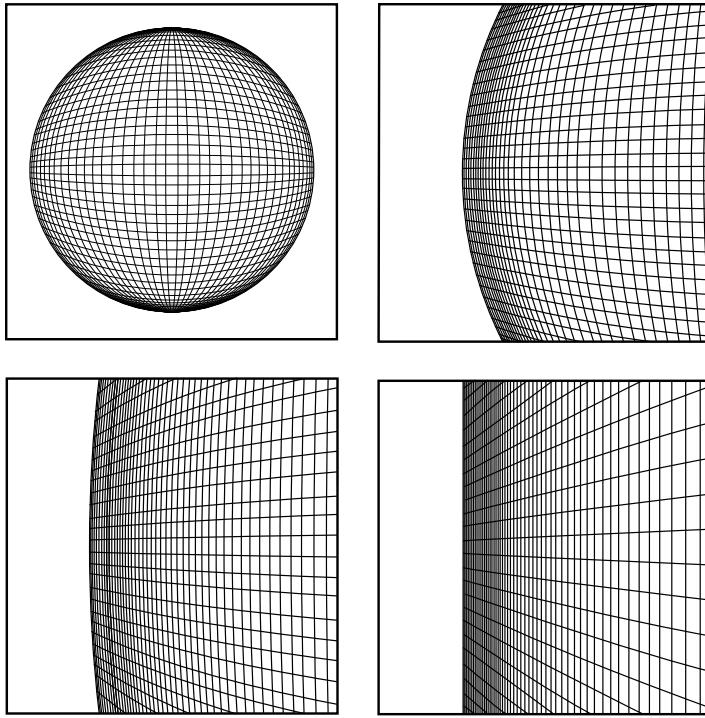


Fig. 28: The exponential expansion during inflation made the radius of curvature of the universe so large that our observable patch of the universe today appears essentially flat, analogous (in three dimensions) to how the surface of a balloon appears flatter and flatter as we inflate it to enormous sizes. This is a crucial prediction of cosmological inflation that will be tested to extraordinary accuracy in the next few years. From Ref. [74, 70].

$$\dot{H} = -\frac{\kappa^2}{2} \dot{\phi}^2, \quad (129)$$

where $\kappa^2 \equiv 8\pi G$. The dynamics of inflation can be described as a perfect fluid (5) with a time dependent pressure and energy density given by

$$\rho = \frac{1}{2} \dot{\phi}^2 + V(\phi), \quad (130)$$

$$p = \frac{1}{2} \dot{\phi}^2 - V(\phi). \quad (131)$$

The field evolution equation (127) can then be written as the energy conservation equation,

$$\dot{\rho} + 3H(\rho + p) = 0. \quad (132)$$

If the potential energy density of the scalar field dominates the kinetic energy, $V(\phi) \gg \dot{\phi}^2$, then we see that

$$p \simeq -\rho \quad \Rightarrow \quad \rho \simeq \text{const.} \quad \Rightarrow \quad H(\phi) \simeq \text{const.}, \quad (133)$$

which leads to the solution

$$a(t) \sim \exp(HT) \quad \Rightarrow \quad \frac{\ddot{a}}{a} > 0 \quad \text{accelerated expansion.} \quad (134)$$

Using the definition of the number of *e*-folds, $N = \ln(a/a_i)$, we see that the scale factor grows exponentially, $a(N) = a_i \exp(N)$. This solution of the Einstein equations solves immediately the flatness

problem. Recall that the problem with the radiation and matter eras is that $\Omega = 1$ ($x = 0$) is an unstable critical point in phase-space. However, during inflation, with $p \simeq -\rho \Rightarrow \omega \simeq -1$, we have that $1 + 3\omega \geq 0$ and therefore $x = 0$ is a stable *attractor* of the equations of motion, see Eq. (121). As a consequence, what seemed an *ad hoc* initial condition, becomes a natural *prediction* of inflation. Suppose that during inflation the scale factor increased N *e-folds*, then

$$x_0 = x_{\text{in}} e^{-2N} \left(\frac{T_{\text{rh}}}{T_{\text{eq}}} \right)^2 (1 + z_{\text{eq}}) \simeq e^{-2N} 10^{56} \leq 1 \quad \Rightarrow \quad N \geq 65, \quad (135)$$

where we have assumed that inflation ended at the scale V_{end} , and the transfer of the inflaton energy density to thermal radiation at reheating occurred almost instantaneously¹¹ at the temperature $T_{\text{rh}} \sim V_{\text{end}}^{1/4} \sim 10^{15}$ GeV. Note that we can now have initial conditions with a large uncertainty, $x_{\text{in}} \simeq 1$, and still have today $x_0 \simeq 1$, thanks to the inflationary attractor towards $\Omega = 1$. This can be understood very easily by realizing that the three curvature evolves during inflation as

$${}^{(3)}R = \frac{6K}{a^2} = {}^{(3)}R_{\text{in}} e^{-2N} \quad \longrightarrow \quad 0, \quad \text{for } N \gg 1. \quad (136)$$

Therefore, if cosmological inflation lasted over 65 *e-folds*, as most models predict, then today the universe (or at least our local patch) should be exactly flat, see Fig. 28, a prediction that can be tested with great accuracy in the near future and for which already seems to be some evidence from observations of the microwave background [87].

Furthermore, inflation also solves the homogeneity problem in a spectacular way. First of all, due to the superluminal expansion, any inhomogeneity existing prior to inflation will be washed out,

$$\delta_k \sim \left(\frac{k}{aH} \right)^2 \Phi_k \propto e^{-2N} \quad \longrightarrow \quad 0, \quad \text{for } N \gg 1. \quad (137)$$

Moreover, since the scale factor grows exponentially, while the horizon distance remains essentially constant, $d_H(t) \simeq H^{-1} = \text{const.}$, any scale within the horizon during inflation will be stretched by the superluminal expansion to enormous distances, in such a way that at photon decoupling all the causally disconnected regions that encompass our present horizon actually come from a single region during inflation, about 65 *e-folds* before the end. This is the reason why two points separated more than 1° in the sky have the same backbody temperature, as observed by the COBE satellite: they were actually in causal contact during inflation. There is at present no other proposal known that could solve the homogeneity problem without invoking an acausal mechanism like inflation.

Finally, any relic particle species (relativistic or not) existing prior to inflation will be diluted by the expansion,

$$\rho_M \propto a^{-3} \sim e^{-3N} \quad \longrightarrow \quad 0, \quad \text{for } N \gg 1, \quad (138)$$

$$\rho_R \propto a^{-4} \sim e^{-4N} \quad \longrightarrow \quad 0, \quad \text{for } N \gg 1. \quad (139)$$

Note that the vacuum energy density ρ_v remains constant under the expansion, and therefore, very soon it is the only energy density remaining to drive the expansion of the universe.

4.22 The slow-roll approximation

In order to simplify the evolution equations during inflation, we will consider the slow-roll approximation (SRA). Suppose that, during inflation, the scalar field evolves very slowly down its effective potential,

¹¹There could be a small delay in thermalization, due to the intrinsic inefficiency of reheating, but this does not change significantly the required number of *e-folds*.

then we can define the slow-roll parameters [76],

$$\epsilon \equiv -\frac{\dot{H}}{H^2} = \frac{\kappa^2}{2} \frac{\dot{\phi}^2}{H^2} \ll 1, \quad (140)$$

$$\delta \equiv -\frac{\ddot{\phi}}{H\dot{\phi}} \ll 1, \quad (141)$$

$$\xi \equiv \frac{\phi}{H^2\dot{\phi}} - \delta^2 \ll 1. \quad (142)$$

It is easy to see that the condition

$$\epsilon < 1 \iff \frac{\ddot{a}}{a} > 0 \quad (143)$$

characterizes inflation: it is all you need for superluminal expansion, i.e. for the horizon distance to grow more slowly than the scale factor, in order to solve the homogeneity problem, as well as for the spatial curvature to decay faster than usual, in order to solve the flatness problem.

The number of e -folds during inflation can be written with the help of Eq. (140) as

$$N = \ln \frac{a_{\text{end}}}{a_i} = \int_{t_i}^{t_e} H dt = \int_{\phi_i}^{\phi_e} \frac{\kappa d\phi}{\sqrt{2\epsilon(\phi)}}, \quad (144)$$

which is an exact expression in terms of $\epsilon(\phi)$.

In the limit given by Eqs. (140), the evolution equations (127) and (128) become

$$H^2 \left(1 - \frac{\epsilon}{3}\right) \simeq H^2 = \frac{\kappa^2}{3} V(\phi), \quad (145)$$

$$3H\dot{\phi} \left(1 - \frac{\delta}{3}\right) \simeq 3H\dot{\phi} = -V'(\phi). \quad (146)$$

Note that this corresponds to a reduction of the dimensionality of phase-space from two to one dimensions, $H(\phi, \dot{\phi}) \rightarrow H(\phi)$. In fact, it is possible to prove a theorem, for single-field inflation, which states that the slow-roll approximation is an attractor of the equations of motion, and thus we can always evaluate the inflationary trajectory in phase-space within the SRA, therefore reducing the number of initial conditions to just one, the initial value of the scalar field. If $H(\phi)$ only depends on ϕ , then $H'(\phi) = -\kappa^2 \dot{\phi}/2$ and we can rewrite the slow-roll parameters (140) as

$$\epsilon = \frac{2}{\kappa^2} \left(\frac{H'(\phi)}{H(\phi)} \right)^2 \simeq \frac{1}{2\kappa^2} \left(\frac{V'(\phi)}{V(\phi)} \right)^2 \equiv \epsilon_V \ll 1, \quad (147)$$

$$\delta = \frac{2}{\kappa^2} \frac{H''(\phi)}{H(\phi)} \simeq \frac{1}{\kappa^2} \frac{V''(\phi)}{V(\phi)} - \frac{1}{2\kappa^2} \left(\frac{V'(\phi)}{V(\phi)} \right)^2 \equiv \eta_V - \epsilon_V \ll 1, \quad (148)$$

$$\begin{aligned} \xi &= \frac{4}{\kappa^4} \frac{H'(\phi)H'''(\phi)}{H^2(\phi)} \simeq \frac{1}{\kappa^4} \frac{V'(\phi)V''''(\phi)}{V^2(\phi)} - \frac{3}{2\kappa^4} \frac{V''(\phi)}{V(\phi)} \left(\frac{V'(\phi)}{V(\phi)} \right)^2 \\ &+ \frac{3}{4\kappa^4} \left(\frac{V'(\phi)}{V(\phi)} \right)^4 \equiv \xi_V - 3\eta_V \epsilon_V + 3\epsilon_V^2 \ll 1. \end{aligned} \quad (149)$$

These expressions define the new slow-roll parameters ϵ_V , η_V and ξ_V . The number of e -folds can also be rewritten in this approximation as

$$N \simeq \int_{\phi_i}^{\phi_e} \frac{\kappa d\phi}{\sqrt{2\epsilon_V(\phi)}} = \kappa^2 \int_{\phi_i}^{\phi_e} \frac{V(\phi) d\phi}{V'(\phi)}, \quad (150)$$

a very useful expression for evaluating N for a given effective scalar potential $V(\phi)$.

4.3 The origin of density perturbations

If cosmological inflation made the universe so extremely flat and homogeneous, where did the galaxies and clusters of galaxies come from? One of the most astonishing predictions of inflation, one that was not even expected, is that quantum fluctuations of the inflaton field are stretched by the exponential expansion and generate large-scale perturbations in the metric. Inflaton fluctuations are small wave packets of energy that, according to general relativity, modify the space-time fabric, creating a whole spectrum of curvature perturbations. The use of the word spectrum here is closely related to the case of light waves propagating in a medium: a spectrum characterizes the amplitude of each given wavelength. In the case of inflation, the inflaton fluctuations induce waves in the space-time metric that can be decomposed into different wavelengths, all with approximately the same amplitude, that is, corresponding to a scale-invariant spectrum. These patterns of perturbations in the metric are like fingerprints that unequivocally characterize a period of inflation. When matter fell in the troughs of these waves, it created density perturbations that collapsed gravitationally to form galaxies, clusters and superclusters of galaxies, with a spectrum that is also scale invariant. Such a type of spectrum was proposed in the early 1970s (before inflation) by Harrison and Zel'dovich [26], to explain the distribution of galaxies and clusters of galaxies on very large scales in our observable universe. Perhaps the most interesting aspect of structure formation is the possibility that the detailed knowledge of what seeded galaxies and clusters of galaxies will allow us to test the idea of inflation.

4.3.1 Reparametrization invariant perturbation theory

Until now we have considered only the unperturbed FRW metric described by a scale factor $a(t)$ and a homogeneous scalar field $\phi(t)$,

$$ds^2 = a^2(\eta)[-d\eta^2 + \gamma_{ij} dx^i dx^j], \quad (151)$$

$$\phi = \phi(\eta), \quad (152)$$

where $\eta = \int dt/a(t)$ is the conformal time, under which the background equations of motion can be written as

$$\mathcal{H}^2 = \frac{\kappa^2}{3} \left(\frac{1}{2}\phi'^2 + a^2 V(\phi) \right), \quad (153)$$

$$\mathcal{H}' - \mathcal{H}^2 = -\frac{\kappa^2}{2}\phi'^2, \quad (154)$$

$$\phi'' + 2\mathcal{H}\phi' + a^2 V'(\phi) = 0, \quad (155)$$

where $\mathcal{H} = aH$ and $\phi' = a\dot{\phi}$.

During inflation, the quantum fluctuations of the scalar field will induce metric perturbations which will backreact on the scalar field. Let us consider, in linear perturbation theory, the most general line element with both scalar and tensor metric perturbations [77],¹² together with the scalar field perturbations

$$ds^2 = a^2(\eta) \left[-(1+2A)d\eta^2 + 2B_{|i}dx^i d\eta + \{(1+2\mathcal{R})\gamma_{ij} + 2E_{|ij} + 2h_{ij}\} dx^i dx^j \right], \quad (156)$$

$$\phi = \phi(\eta) + \delta\phi(\eta, x^i). \quad (157)$$

The indices $\{i, j\}$ label the three-dimensional spatial coordinates with metric γ_{ij} , and the $|i$ denotes covariant derivative with respect to that metric. The gauge invariant tensor perturbation h_{ij} corresponds to a transverse traceless gravitational wave, $\nabla^i h_{ij} = h_i^i = 0$. The four scalar perturbations (A, B, \mathcal{R}, E) are *gauge dependent* functions of (η, x^i) . Under a general coordinate (gauge) transformation [77, 78]

$$\tilde{\eta} = \eta + \xi^0(\eta, x^i), \quad (158)$$

$$\tilde{x}^i = x^i + \gamma^{ij}\xi_{|j}(\eta, x^i), \quad (159)$$

¹²Note that inflation cannot generate, to linear order, a vector perturbation.

with arbitrary functions (ξ^0, ξ) , the scalar and tensor perturbations transform, to linear order, as

$$\tilde{A} = A - \xi^{0'} - \mathcal{H}\xi^0, \quad \tilde{B} = B + \xi^0 - \xi', \quad (160)$$

$$\tilde{\mathcal{R}} = \mathcal{R} - \mathcal{H}\xi^0, \quad \tilde{E} = E - \xi, \quad (161)$$

$$\tilde{h}_{ij} = h_{ij}, \quad (162)$$

where a prime denotes derivative with respect to conformal time. It is possible to construct, however, two gauge-invariant gravitational potentials [77, 78],

$$\Phi = A + (B - E')' + \mathcal{H}(B - E'), \quad (163)$$

$$\Psi = \mathcal{R} + \mathcal{H}(B - E'), \quad (164)$$

which are related through the perturbed Einstein equations,

$$\Phi = \Psi, \quad (165)$$

$$\frac{k^2 - 3K}{a^2} \Psi = \frac{\kappa^2}{2} \delta\rho, \quad (166)$$

where $\delta\rho$ is the gauge-invariant density perturbation, and the latter expression is nothing but the Poisson equation for the gravitational potential, written in relativistic form.

During inflation, the energy density is given in terms of a scalar field, and thus the gauge-invariant equations for the perturbations on comoving hypersurfaces (constant energy density hypersurfaces) are

$$\Phi'' + 3\mathcal{H}\Phi' + (\mathcal{H}' + 2\mathcal{H}^2)\Phi = \frac{\kappa^2}{2}[\phi'\delta\phi' - a^2V'(\phi)\delta\phi], \quad (167)$$

$$-\nabla^2\Phi + 3\mathcal{H}\Phi' + (\mathcal{H}' + 2\mathcal{H}^2)\Phi = -\frac{\kappa^2}{2}[\phi'\delta\phi' + a^2V'(\phi)\delta\phi], \quad (168)$$

$$\Phi' + \mathcal{H}\Phi = \frac{\kappa^2}{2}\phi'\delta\phi, \quad (169)$$

$$\delta\phi'' + 2\mathcal{H}\delta\phi' - \nabla^2\delta\phi = 4\phi'\Phi' - 2a^2V'(\phi)\Phi - a^2V''(\phi)\delta\phi. \quad (170)$$

This system of equations seem too difficult to solve at first sight. However, there is a gauge invariant combination of variables that allows one to find exact solutions. Let us define [78]

$$u \equiv a\delta\phi + z\Phi, \quad (171)$$

$$z \equiv a\frac{\phi'}{\mathcal{H}}. \quad (172)$$

Under this redefinition, the above equations simplify enormously to just three independent equations,

$$u'' - \nabla^2u - \frac{z''}{z}u = 0, \quad (173)$$

$$\nabla^2\Phi = \frac{\kappa^2}{2}\frac{\mathcal{H}}{a^2}(zu' - z'u), \quad (174)$$

$$\left(\frac{a^2\Phi}{\mathcal{H}}\right)' = \frac{\kappa^2}{2}zu. \quad (175)$$

From Equation (173) we can find a solution $u(z)$, which substituted into (175) can be integrated to give $\Phi(z)$, and together with $u(z)$ allow us to obtain $\delta\phi(z)$.

4.32 Quantum Field Theory in curved space-time

Until now we have treated the perturbations as classical, but we should in fact consider the perturbations Φ and $\delta\phi$ as quantum fields. Note that the perturbed action for the scalar mode u can be written as

$$\delta S = \frac{1}{2} \int d^3x d\eta \left[(u')^2 - (\nabla u)^2 + \frac{z''}{z} u^2 \right]. \quad (176)$$

In order to quantize the field u in the curved background defined by the metric (151), we can write the operator

$$\hat{u}(\eta, \mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \left[u_k(\eta) \hat{a}_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} + u_k^*(\eta) \hat{a}_{\mathbf{k}}^\dagger e^{-i\mathbf{k}\cdot\mathbf{x}} \right], \quad (177)$$

where the creation and annihilation operators satisfy the commutation relation of bosonic fields, and the scalar field's Fock space is defined through the vacuum condition,

$$[\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}^\dagger] = \delta^3(\mathbf{k} - \mathbf{k}'), \quad (178)$$

$$\hat{a}_{\mathbf{k}} |0\rangle = 0. \quad (179)$$

Note that we are not assuming that the inflaton is a fundamental scalar field, but that is can be written as a quantum field with its commutation relations (as much as a pion can be described as a quantum field).

The equations of motion for each mode $u_k(\eta)$ are decoupled in linear perturbation theory,

$$u_k'' + \left(k^2 - \frac{z''}{z} \right) u_k = 0. \quad (180)$$

The ratio z''/z acts like a time-dependent potential for this Schrödinger like equation. In order to find exact solutions to the mode equation, we will use the slow-roll parameters (140), see Ref. [76]

$$\epsilon = 1 - \frac{\mathcal{H}'}{\mathcal{H}^2} = \frac{\kappa^2}{2} \frac{z^2}{a^2}, \quad (181)$$

$$\delta = 1 - \frac{\phi''}{\mathcal{H}\phi'} = 1 + \epsilon - \frac{z'}{\mathcal{H}z}, \quad (182)$$

$$\xi = - \left(2 - \epsilon - 3\delta + \delta^2 - \frac{\phi'''}{\mathcal{H}^2\phi'} \right). \quad (183)$$

In terms of these parameters, the conformal time and the effective potential for the u_k mode can be written as

$$\eta = \frac{-1}{\mathcal{H}} + \int \frac{eda}{a\mathcal{H}}, \quad (184)$$

$$\frac{z''}{z} = \mathcal{H}^2 \left[(1 + \epsilon - \delta)(2 - \delta) + \mathcal{H}^{-1}(\epsilon' - \delta') \right]. \quad (185)$$

Note that the slow-roll parameters, (181) and (182), can be taken as *constant*,¹³ to order ϵ^2 ,

$$\begin{aligned} \epsilon' &= 2\mathcal{H}(\epsilon^2 - \epsilon\delta) = \mathcal{O}(\epsilon^2), \\ \delta' &= \mathcal{H}(\epsilon\delta - \xi) = \mathcal{O}(\epsilon^2). \end{aligned} \quad (186)$$

In that case, for constant slow-roll parameters, we can write

$$\eta = \frac{-1}{\mathcal{H}} \frac{1}{1 - \epsilon}, \quad (187)$$

$$\frac{z''}{z} = \frac{1}{\eta^2} \left(\nu^2 - \frac{1}{4} \right), \quad \text{where} \quad \nu = \frac{1 + \epsilon - \delta}{1 - \epsilon} + \frac{1}{2}. \quad (188)$$

¹³For instance, there are models of inflation, like power-law inflation, $a(t) \sim t^p$, where $\epsilon = \delta = 1/p < 1$, that give constant slow-roll parameters.

We are now going to search for approximate solutions of the mode equation (180), where the effective potential (185) is of order $z''/z \simeq 2\mathcal{H}^2$ in the slow-roll approximation. In quasi-de Sitter there is a characteristic scale given by the (event) horizon size or Hubble scale during inflation, H^{-1} . There will be modes u_k with physical wavelengths much smaller than this scale, $k/a \gg H$, that are well within the de Sitter horizon and therefore do not feel the curvature of space-time. On the other hand, there will be modes with physical wavelengths much greater than the Hubble scale, $k/a \ll H$. In these two asymptotic regimes, the solutions can be written as

$$u_k = \frac{1}{\sqrt{2k}} e^{-ik\eta} \quad k \gg aH, \quad (189)$$

$$u_k = C_1 z \quad k \ll aH. \quad (190)$$

In the limit $k \gg aH$ the modes behave like ordinary quantum modes in Minkowsky space-time, appropriately normalized, while in the opposite limit, u/z becomes constant on superhorizon scales. For approximately constant slow-roll parameters one can find exact solutions to (180), with the effective potential given by (188), that interpolate between the two asymptotic solutions,

$$u_k(\eta) = \frac{\sqrt{\pi}}{2} e^{i(\nu+\frac{1}{2})\frac{\pi}{2}} (-\eta)^{1/2} H_\nu^{(1)}(-k\eta), \quad (191)$$

where $H_\nu^{(1)}(z)$ is the Hankel function of the first kind [79], and ν is given by (188) in terms of the slow-roll parameters. In the limit $k\eta \rightarrow 0$, the solution becomes

$$|u_k| = \frac{2^{\nu-\frac{3}{2}}}{\sqrt{2k}} \frac{\Gamma(\nu)}{\Gamma(\frac{3}{2})} (-k\eta)^{\frac{1}{2}-\nu} \equiv \frac{C(\nu)}{\sqrt{2k}} \left(\frac{k}{aH}\right)^{\frac{1}{2}-\nu}, \quad (192)$$

$$C(\nu) = 2^{\nu-\frac{3}{2}} \frac{\Gamma(\nu)}{\Gamma(\frac{3}{2})} (1-\epsilon)^{\nu-\frac{1}{2}} \simeq 1 \quad \text{for } \epsilon, \delta \ll 1. \quad (193)$$

We can now compute Φ and $\delta\phi$ from the super-Hubble-scale mode solution (190), for $k \ll aH$. Substituting into Eq. (175), we find

$$\Phi = C_1 \left(1 - \frac{\mathcal{H}}{a^2} \int a^2 d\eta\right) + C_2 \frac{\mathcal{H}}{a^2}, \quad (194)$$

$$\delta\phi = \frac{C_1}{a^2} \int a^2 d\eta - \frac{C_2}{a^2}. \quad (195)$$

The term proportional to C_1 corresponds to the growing solution, while that proportional to C_2 corresponds to the decaying solution, which can soon be ignored. These quantities are gauge invariant but evolve with time outside the horizon, during inflation, and before entering again the horizon during the radiation or matter eras. We would like to write an expression for a gauge invariant quantity that is also *constant* for superhorizon modes. Fortunately, in the case of adiabatic perturbations, there is such a quantity:

$$\zeta \equiv \Phi + \frac{1}{\epsilon\mathcal{H}} (\Phi' + \mathcal{H}\Phi) = \frac{u}{z}, \quad (196)$$

which is constant, see Eq. (190), for $k \ll aH$. In fact, this quantity ζ is identical, for superhorizon modes, to the gauge invariant curvature metric perturbation \mathcal{R}_c on comoving (constant energy density) hypersurfaces, see Ref. [77, 80],

$$\zeta = \mathcal{R}_c + \frac{1}{\epsilon\mathcal{H}^2} \nabla^2 \Phi. \quad (197)$$

Using Eq. (174) we can write the evolution equation for $\zeta = \frac{u}{z}$ as $\zeta' = \frac{1}{\epsilon\mathcal{H}} \nabla^2 \Phi$, which confirms that ζ is constant for (adiabatic¹⁴) superhorizon modes, $k \ll aH$. Therefore, we can evaluate the Newtonian

¹⁴This conservation fails for entropy or isocurvature perturbations, see Ref. [80].

potential Φ_k when the perturbation reenters the horizon during radiation/matter eras in terms of the curvature perturbation \mathcal{R}_k when it left the Hubble scale during inflation,

$$\Phi_k = \left(1 - \frac{\mathcal{H}}{a^2} \int a^2 d\eta\right) \mathcal{R}_k = \frac{3+3\omega}{5+3\omega} \mathcal{R}_k = \begin{cases} \frac{2}{3} \mathcal{R}_k & \text{radiation era,} \\ \frac{3}{5} \mathcal{R}_k & \text{matter era.} \end{cases} \quad (198)$$

Let us now compute the tensor or gravitational wave metric perturbations generated during inflation. The perturbed action for the tensor mode can be written as

$$\delta S = \frac{1}{2} \int d^3x d\eta \frac{a^2}{2\kappa^2} \left[(h'_{ij})^2 - (\nabla h_{ij})^2 \right], \quad (199)$$

with the tensor field h_{ij} considered as a quantum field,

$$\hat{h}_{ij}(\eta, \mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \sum_{\lambda=1,2} \left[h_k(\eta) e_{ij}(\mathbf{k}, \lambda) \hat{a}_{\mathbf{k}, \lambda} e^{i\mathbf{k}\cdot\mathbf{x}} + h.c. \right], \quad (200)$$

where $e_{ij}(\mathbf{k}, \lambda)$ are the two polarization tensors, satisfying symmetric, transverse and traceless conditions

$$e_{ij} = e_{ji}, \quad k^i e_{ij} = 0, \quad e_{ii} = 0, \quad (201)$$

$$e_{ij}(-\mathbf{k}, \lambda) = e_{ij}^*(\mathbf{k}, \lambda), \quad \sum_{\lambda} e_{ij}^*(\mathbf{k}, \lambda) e^{ij}(\mathbf{k}, \lambda) = 4, \quad (202)$$

while the creation and annihilation operators satisfy the usual commutation relation of bosonic fields, Eq. (178). We can now redefine our gauge invariant tensor amplitude as

$$v_k(\eta) = \frac{a}{\sqrt{2}\kappa} h_k(\eta), \quad (203)$$

which satisfies the following evolution equation, decoupled for each mode $v_k(\eta)$ in linear perturbation theory,

$$v_k'' + \left(k^2 - \frac{a''}{a} \right) v_k = 0. \quad (204)$$

The ratio a''/a acts like a time-dependent potential for this Schrödinger like equation, analogous to the term z''/z for the scalar metric perturbation. For constant slow-roll parameters, the potential becomes

$$\frac{a''}{a} = 2\mathcal{H}^2 \left(1 - \frac{\epsilon}{2}\right) = \frac{1}{\eta^2} \left(\mu^2 - \frac{1}{4}\right), \quad (205)$$

$$\mu = \frac{1}{1-\epsilon} + \frac{1}{2}. \quad (206)$$

We can solve equation (204) in the two asymptotic regimes,

$$v_k = \frac{1}{\sqrt{2k}} e^{-ik\eta} \quad k \gg aH, \quad (207)$$

$$v_k = C a \quad k \ll aH. \quad (208)$$

In the limit $k \gg aH$ the modes behave like ordinary quantum modes in Minkowsky space-time, appropriately normalized, while in the opposite limit, the metric perturbation h_k becomes *constant* on superhorizon scales. For constant slow-roll parameters one can find exact solutions to (204), with effective potential given by (205), that interpolate between the two asymptotic solutions. These are identical to Eq. (191) except for the substitution $\nu \rightarrow \mu$. In the limit $k\eta \rightarrow 0$, the solution becomes

$$|v_k| = \frac{C(\mu)}{\sqrt{2k}} \left(\frac{k}{aH} \right)^{\frac{1}{2}-\mu}. \quad (209)$$

Since the mode h_k becomes constant on superhorizon scales, we can evaluate the tensor metric perturbation when it reentered during the radiation or matter era directly in terms of its value during inflation.

4.33 Power spectrum of scalar and tensor metric perturbations

Not only do we expect to measure the amplitude of the metric perturbations generated during inflation and responsible for the anisotropies in the CMB and density fluctuations in LSS, but we should also be able to measure its power spectrum, or two-point correlation function in Fourier space. Let us consider first the scalar metric perturbations \mathcal{R}_k , which enter the horizon at $a = k/H$. Its correlator is given by [76]

$$\langle 0 | \mathcal{R}_k^* \mathcal{R}_{k'} | 0 \rangle = \frac{|u_k|^2}{z^2} \delta^3(\mathbf{k} - \mathbf{k}') \equiv \frac{\mathcal{P}_{\mathcal{R}}(k)}{4\pi k^3} (2\pi)^3 \delta^3(\mathbf{k} - \mathbf{k}'), \quad (210)$$

$$\mathcal{P}_{\mathcal{R}}(k) = \frac{k^3}{2\pi^2} \frac{|u_k|^2}{z^2} = \frac{\kappa^2}{2\epsilon} \left(\frac{H}{2\pi}\right)^2 \left(\frac{k}{aH}\right)^{3-2\nu} \equiv A_S^2 \left(\frac{k}{aH}\right)^{n_s-1}, \quad (211)$$

where we have used $\mathcal{R}_k = \zeta_k = \frac{u_k}{z}$ and Eq. (192). This last equation determines the power spectrum in terms of its amplitude at horizon-crossing, A_S , and a tilt,

$$n_s - 1 \equiv \frac{d \ln \mathcal{P}_{\mathcal{R}}(k)}{d \ln k} = 3 - 2\nu = 2 \left(\frac{\delta - 2\epsilon}{1 - \epsilon} \right) \simeq 2\eta_V - 6\epsilon_V, \quad (212)$$

see Eqs. (147), (148). Note from this equation that it is possible, in principle, to obtain from inflation a scalar tilt which is either positive ($n > 1$) or negative ($n < 1$). Furthermore, depending on the particular inflationary model [81], we can have significant departures from scale invariance.

Note that at horizon entry $k\eta = -1$, and thus we can alternatively evaluate the tilt as

$$n_s - 1 \equiv -\frac{d \ln \mathcal{P}_{\mathcal{R}}}{d \ln \eta} = -2\eta \mathcal{H} \left[(1 - \epsilon) - (\epsilon - \delta) - 1 \right] = 2 \left(\frac{\delta - 2\epsilon}{1 - \epsilon} \right) \simeq 2\eta_V - 6\epsilon_V, \quad (213)$$

and the running of the tilt

$$\frac{dn_s}{d \ln k} = -\frac{dn_s}{d \ln \eta} = -\eta \mathcal{H} \left(2\xi + 8\epsilon^2 - 10\epsilon\delta \right) \simeq 2\xi_V + 24\epsilon_V^2 - 16\eta_V\epsilon_V, \quad (214)$$

where we have used Eqs. (186).

Let us consider now the tensor (gravitational wave) metric perturbation, which enter the horizon at $a = k/H$,

$$\sum_{\lambda} \langle 0 | h_{k,\lambda}^* h_{k',\lambda} | 0 \rangle = 4 \frac{2\kappa^2}{a^2} |v_k|^2 \delta^3(\mathbf{k} - \mathbf{k}') \equiv \frac{\mathcal{P}_g(k)}{4\pi k^3} (2\pi)^3 \delta^3(\mathbf{k} - \mathbf{k}'), \quad (215)$$

$$\mathcal{P}_g(k) = 8\kappa^2 \left(\frac{H}{2\pi}\right)^2 \left(\frac{k}{aH}\right)^{3-2\mu} \equiv A_T^2 \left(\frac{k}{aH}\right)^{n_T}, \quad (216)$$

where we have used Eqs. (203) and (209). Therefore, the power spectrum can be approximated by a power-law expression, with amplitude A_T and tilt

$$n_T \equiv \frac{d \ln \mathcal{P}_g(k)}{d \ln k} = 3 - 2\mu = \frac{-2\epsilon}{1 - \epsilon} \simeq -2\epsilon_V < 0, \quad (217)$$

which is always negative. In the slow-roll approximation, $\epsilon \ll 1$, the tensor power spectrum is scale invariant.

Alternatively, we can evaluate the tensor tilt by

$$n_T \equiv -\frac{d \ln \mathcal{P}_g}{d \ln \eta} = -2\eta \mathcal{H} \left[(1 - \epsilon) - 1 \right] = \frac{-2\epsilon}{1 - \epsilon} \simeq -2\epsilon_V, \quad (218)$$

and its running by

$$\frac{dn_T}{d \ln k} = -\frac{dn_T}{d \ln \eta} = -\eta \mathcal{H} \left(4\epsilon^2 - 4\epsilon\delta \right) \simeq 8\epsilon_V^2 - 4\eta_V\epsilon_V, \quad (219)$$

where we have used Eqs. (186).

4.4 The anisotropies of the microwave background

The metric fluctuations generated during inflation are not only responsible for the density perturbations that gave rise to galaxies via gravitational collapse, but one should also expect to see such ripples in the metric as temperature anisotropies in the cosmic microwave background, that is, minute deviations in the temperature of the blackbody spectrum when we look at different directions in the sky. Such anisotropies had been looked for ever since Penzias and Wilson's discovery of the CMB, but had eluded all detection, until COBE satellite discovered them in 1992, see Fig. 10. The reason why they took so long to be discovered was that they appear as perturbations in temperature of only one part in 10^5 . Soon after COBE, other groups quickly confirmed the detection of temperature anisotropies at around $30 \mu\text{K}$, at higher multipole numbers or smaller angular scales.

4.4.1 The Sachs-Wolfe effect

The anisotropies corresponding to large angular scales are only generated via gravitational red-shift and density perturbations through the Einstein equations, $\delta\rho/\rho = -2\Phi$ for adiabatic perturbations; we can ignore the Doppler contribution, since the perturbation is non-causal. In that case, the temperature anisotropy in the sky today is given by [82]

$$\frac{\delta T}{T}(\theta, \phi) = \frac{1}{3}\Phi(\eta_{\text{LS}})Q(\eta_0, \theta, \phi) + 2 \int_{\eta_{\text{LS}}}^{\eta_0} dr \Phi'(\eta_0 - r)Q(r, \theta, \phi), \quad (220)$$

where η_0 is the *coordinate distance* to the last scattering surface, i.e. the present conformal time, while $\eta_{\text{LS}} \simeq 0$ determines that comoving hypersurface. The above expression is known as the Sachs-Wolfe effect [82], and contains two parts, the intrinsic and the Integrated Sachs-Wolfe (ISW) effect, due to integration along the line of sight of time variations in the gravitational potential.

In linear perturbation theory, the scalar metric perturbations can be separated into $\Phi(\eta, \mathbf{x}) \equiv \Phi(\eta)Q(\mathbf{x})$, where $Q(\mathbf{x})$ are the scalar harmonics, eigenfunctions of the Laplacian in three dimensions, $\nabla^2 Q_{klm}(r, \theta, \phi) = -k^2 Q_{klm}(r, \theta, \phi)$. These functions have the general form [83]

$$Q_{klm}(r, \theta, \phi) = \Pi_{kl}(r)Y_{lm}(\theta, \phi), \quad (221)$$

where $Y_{lm}(\theta, \phi)$ are the usual spherical harmonics [79].

In order to compute the temperature anisotropy associated with the Sachs-Wolfe effect, we have to know the evolution of the metric perturbation during the matter era,

$$\Phi'' + 3\mathcal{H}\Phi' + a^2\Lambda\Phi - 2K\Phi = 0. \quad (222)$$

In the case of a flat universe without cosmological constant, the Newtonian potential remains constant during the matter era and only the intrinsic SW effect contributes to $\delta T/T$. In case of a non-vanishing Λ , since its contribution is negligible in the past, most of the photon's trajectory towards us is unperturbed, and the only difference with respect to the $\Lambda = 0$ case is an overall factor [86]. We will consider here the approximation $\Phi = \text{constant}$ during the matter era and ignore that factor, see Ref. [84].

In a flat universe, the radial part of the eigenfunctions (221) can be written as [83]

$$\Pi_{kl}(r) = \sqrt{\frac{2}{\pi}}k j_l(kr), \quad (223)$$

where $j_l(z)$ are the spherical Bessel functions [79]. The growing mode solution of the metric perturbation that left the Hubble scale during inflation contributes to the temperature anisotropies on large scales (220) as

$$\frac{\delta T}{T}(\theta, \phi) = \frac{1}{3}\Phi(\eta_{\text{LS}})Q = \frac{1}{5}\mathcal{R}Q(\eta_0, \theta, \phi) \equiv \sum_{l=2}^{\infty} \sum_{m=-l}^l a_{lm} Y_{lm}(\theta, \phi), \quad (224)$$

where we have used the fact that at reentry (at the surface of last scattering) the gauge invariant Newtonian potential Φ is related to the curvature perturbation \mathcal{R} at Hubble-crossing during inflation, see Eq. (198); and we have expanded $\delta T/T$ in spherical harmonics.

We can now compute the two-point correlation function or angular power spectrum, $C(\theta)$, of the CMB anisotropies on large scales, defined as an expansion in multipole number,

$$C(\theta) = \left\langle \frac{\delta T^*}{T}(\mathbf{n}) \frac{\delta T}{T}(\mathbf{n}') \right\rangle_{\mathbf{n} \cdot \mathbf{n}' = \cos \theta} = \frac{1}{4\pi} \sum_{l=2}^{\infty} (2l+1) C_l P_l(\cos \theta), \quad (225)$$

where $P_l(z)$ are the Legendre polynomials [79], and we have averaged over different universe realizations. Since the coefficients a_{lm} are isotropic (to first order), we can compute the $C_l = \langle |a_{lm}|^2 \rangle$ as

$$C_l^{(S)} = \frac{4\pi}{25} \int_0^\infty \frac{dk}{k} \mathcal{P}_{\mathcal{R}}(k) j_l^2(k\eta_0), \quad (226)$$

where we have used Eqs. (224) and (210). In the case of scalar metric perturbation produced during inflation, the scalar power spectrum at reentry is given by $\mathcal{P}_{\mathcal{R}}(k) = A_S^2 (k\eta_0)^{n-1}$, in the power-law approximation, see Eq. (211). In that case, one can integrate (226) to give

$$C_l^{(S)} = \frac{2\pi}{25} A_S^2 \frac{\Gamma[\frac{3}{2}] \Gamma[1 - \frac{n-1}{2}] \Gamma[l + \frac{n-1}{2}]}{\Gamma[\frac{3}{2} - \frac{n-1}{2}] \Gamma[l + 2 - \frac{n-1}{2}]}, \quad (227)$$

$$\frac{l(l+1) C_l^{(S)}}{2\pi} = \frac{A_S^2}{25} = \text{constant}, \quad \text{for } n = 1. \quad (228)$$

This last expression corresponds to what is known as the Sachs-Wolfe plateau, and is the reason why the coefficients C_l are always plotted multiplied by $l(l+1)$, see Fig. 3.4.

Tensor metric perturbations also contribute with an approximately constant angular power spectrum, $l(l+1)C_l$. The Sachs-Wolfe effect for a gauge invariant tensor perturbation is given by [82]

$$\frac{\delta T}{T}(\theta, \phi) = \int_{\eta_{\text{LS}}}^{\eta_0} dr h'(\eta_0 - r) Q_{rr}(r, \theta, \phi), \quad (229)$$

where Q_{rr} is the rr -component of the tensor harmonic along the line of sight [83]. The tensor perturbation h during the matter era satisfies the following evolution equation

$$h_k'' + 3\mathcal{H} h_k' + (k^2 + 2K) h_k = 0, \quad (230)$$

which depends on the wavenumber k , contrary to what happens with the scalar modes, see Eq. (222). For a flat ($K = 0$) universe, the solution to this equation is $h_k(\eta) = h G_k(\eta)$, where h is the constant tensor metric perturbation at horizon crossing and $G_k(\eta) = 3 j_1(k\eta)/k\eta$, normalized so that $G_k(0) = 1$ at the surface of last scattering. The radial part of the tensor harmonic Q_{rr} in a flat universe can be written as [83]

$$Q_{kl}^{rr}(r) = \left[\frac{(l-1)l(l+1)(l+2)}{\pi k^2} \right]^{1/2} \frac{j_l(kr)}{r^2}. \quad (231)$$

The tensor angular power spectrum can finally be expressed as

$$C_l^{(T)} = \frac{9\pi}{4} (l-1)l(l+1)(l+2) \int_0^\infty \frac{dk}{k} \mathcal{P}_g(k) I_{kl}^2, \quad (232)$$

$$I_{kl} = \int_0^{x_0} dx \frac{j_2(x_0 - x) j_l(x)}{(x_0 - x)x^2}, \quad (233)$$

where $x \equiv k\eta$, and $\mathcal{P}_g(k)$ is the primordial tensor spectrum (216). For a scale invariant spectrum, $n_T = 0$, we can integrate (232) to give [85]

$$l(l+1)C_l^{(T)} = \frac{\pi}{36} \left(1 + \frac{48\pi^2}{385}\right) A_T^2 B_l, \quad (234)$$

with $B_l = (1.1184, 0.8789, \dots, 1.00)$ for $l = 2, 3, \dots, 30$. Therefore, $l(l+1)C_l^{(T)}$ also becomes constant for large l . Beyond $l \sim 30$, the Sachs-Wolfe expression is not a good approximation and the tensor angular power spectrum decays very quickly at large l , see Fig. 31.

4.42 The consistency relation

In spite of the success of inflation in predicting a homogeneous and isotropic background on which to imprint a scale-invariant spectrum of inhomogeneities, it is difficult to test the idea of inflation. A CMB cosmologist before the 1980s would have argued that *ad hoc* initial conditions could have been at the origin of the homogeneity and flatness of the universe on large scales, while a LSS cosmologist would have agreed with Harrison and Zel'dovich that the most natural spectrum needed to explain the formation of structure was a scale-invariant spectrum. The surprise was that inflation incorporated an understanding of *both* the globally homogeneous and spatially flat background, and the approximately scale-invariant spectrum of perturbations in the same formalism. But that could have been just a coincidence.

What is *unique* to inflation is the fact that inflation determines not just one but *two* primordial spectra, corresponding to the scalar (density) and tensor (gravitational waves) metric perturbations, from a single continuous function, the inflaton potential $V(\phi)$. In the slow-roll approximation, one determines, from $V(\phi)$, two continuous functions, $\mathcal{P}_R(k)$ and $\mathcal{P}_g(k)$, that in the power-law approximation reduces to two amplitudes, A_S and A_T , and two tilts, n and n_T . It is clear that there must be a relation between the four parameters. Indeed, one can see from Eqs. (234) and (228) that the ratio of the tensor to scalar contribution to the angular power spectrum is proportional to the tensor tilt [76],

$$R \equiv \frac{C_l^{(T)}}{C_l^{(S)}} = \frac{25}{9} \left(1 + \frac{48\pi^2}{385}\right) 2\epsilon \simeq -2\pi n_T. \quad (235)$$

This is a unique prediction of inflation, which could not have been postulated a priori by any cosmologist. If we finally observe a tensor spectrum of anisotropies in the CMB, or a stochastic gravitational wave background in laser interferometers like LIGO or LISA, with sufficient accuracy to determine their spectral tilt, one might have some chance to test the idea of inflation, via the consistency relation (235). For the moment, observations of the microwave background anisotropies suggest that the Sachs-Wolfe plateau exists, see Fig. 3.4, but it is still premature to determine the tensor contribution. Perhaps in the near future, from the analysis of polarization as well as temperature anisotropies, with the CMB satellites MAP and Planck, we might have a chance of determining the validity of the consistency relation.

Assuming that the scalar contribution dominates over the tensor on large scales, i.e. $R \ll 1$, one can actually give a measure of the amplitude of the scalar metric perturbation from the observations of the Sachs-Wolfe plateau in the angular power spectrum [20],

$$\left[\frac{l(l+1)C_l^{(S)}}{2\pi} \right]^{1/2} = \frac{A_S}{5} = (1.03 \pm 0.07) \times 10^{-5}, \quad (236)$$

$$n = 0.97 \pm 0.03. \quad (237)$$

These measurements can be used to normalize the primordial spectrum and determine the parameters of the model of inflation [81]. In the near future these parameters will be determined with much better accuracy, as described in Section 4.4.5.

4.43 The acoustic peaks

The Sachs-Wolfe plateau is a distinctive feature of Fig. 24. These observations confirm the existence of a primordial spectrum of scalar (density) perturbations on all scales, otherwise the power spectrum would have started from zero at $l = 2$. However, we see that the spectrum starts to rise around $l = 20$ towards the first acoustic peak, where the SW approximation breaks down and the above formulae are no longer valid.

As mentioned above, the first peak in the photon distribution corresponds to overdensities that have undergone half an oscillation, that is, a compression, and appear at a scale associated with the size of the horizon at last scattering, about 1° projected in the sky today. Since photons scatter off baryons, they will also feel the acoustic wave and create a peak in the correlation function. The height of the peak is proportional to the amount of baryons: the larger the baryon content of the universe, the higher the peak. The position of the peak in the power spectrum depends on the geometrical size of the particle horizon at last scattering. Since photons travel along geodesics, the projected size of the causal horizon at decoupling depends on whether the universe is flat, open or closed. In a flat universe the geodesics are straight lines and, by looking at the angular scale of the first acoustic peak, we would be measuring the actual size of the horizon at last scattering. In an open universe, the geodesics are inward-curved trajectories, and therefore the projected size on the sky appears smaller. In this case, the first acoustic peak should occur at higher multipoles or smaller angular scales. On the other hand, for a closed universe, the first peak occurs at smaller multipoles or larger angular scales. The dependence of the position of the first acoustic peak on the spatial curvature can be approximately given by $l_{\text{peak}} \simeq 220 \Omega_0^{-1/2}$, where $\Omega_0 = \Omega_M + \Omega_\Lambda = 1 - \Omega_K$. Past observations from the balloon experiment BOOMERANG [87], suggested clearly a few years ago that the first peak was between $l = 180$ and 250 at 95% c.l., with an amplitude $\delta T = 80 \pm 10 \mu\text{K}$, and therefore the universe was most probably flat. However, with the high precision WMAP data we can now pinpoint the spatial curvature to a few percent,

$$\Omega_0 = 1.02 \pm 0.02 \quad (95\% \text{ c.l.}) \quad (238)$$

That is, the universe is spatially flat (i.e. Euclidean), within 2% uncertainty, which is much better than we could ever do before.

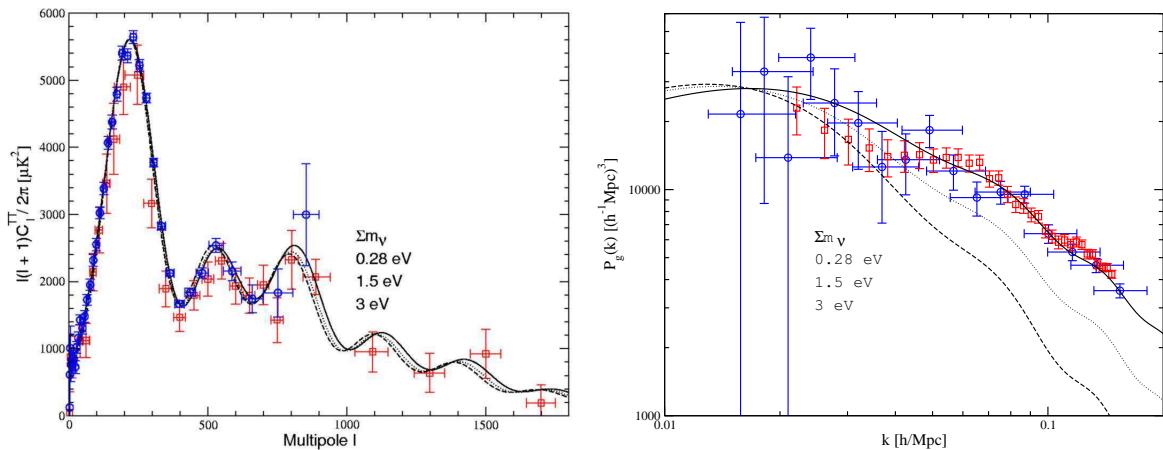


Fig. 29: The dependence of CMB anisotropies and LSS power spectrum on the sum of the mass of all neutrino species. The blue(red) data corresponds to WMAP(Boomerang, etc.) and SDSS(2dFGRS), for the CMB and LSS respectively.

With Boomerang, CBI, VSA, and specially with WMAP, we have evidence of at least three distinct acoustic peaks. In the near future, even before Planck, we may be able to distinguish another two. These peaks should occur at harmonics of the first one, but are typically much lower because of Silk damping. Since the amplitude and position of the primary and secondary peaks are directly determined by the

sound speed (and, hence, the equation of state) and by the geometry and expansion of the universe, they can be used as a powerful test of the density of baryons and dark matter, and other cosmological parameters.

By looking at these patterns in the anisotropies of the microwave background, cosmologists can determine not only the cosmological parameters, but also the primordial spectrum of density perturbations produced during inflation. It turns out that the observed temperature anisotropies are compatible with a scale-invariant spectrum, see Eq. (237), as predicted by inflation. This is remarkable, and gives very strong support to the idea that inflation may indeed be responsible for both the CMB anisotropies and the large-scale structure of the universe. Different models of inflation have different specific predictions for the fine details associated with the spectrum generated during inflation. It is these minute differences that will allow cosmologists to differentiate between alternative models of inflation and discard those that do not agree with observations. However, most importantly, perhaps, the pattern of anisotropies predicted by inflation is completely different from those predicted by alternative models of structure formation, like cosmic defects: strings, vortices, textures, etc. These are complicated networks of energy density concentrations left over from an early universe phase transition, analogous to the defects formed in the laboratory in certain kinds of liquid crystals when they go through a phase transition. The cosmological defects have spectral properties very different from those generated by inflation. That is why it is so important to launch more sensitive instruments, and with better angular resolution, to determine the properties of the CMB anisotropies.

4.44 The new microwave anisotropy satellites, WMAP and Planck

The large amount of information encoded in the anisotropies of the microwave background is the reason why both NASA and the European Space Agency have decided to launch two independent satellites to measure the CMB temperature and polarization anisotropies to unprecedented accuracy. The Wilkinson Microwave Anisotropy Probe [88] was launched by NASA at the end of 2000, and has fulfilled most of our expectation, while Planck [89] is expected to be launched by ESA in 2007. There are at the moment other large proposals like CMB Pol [95], ACT [96], etc. which will see the light in the next few years, see Ref. [90].

As we have emphasized before, the fact that these anisotropies have such a small amplitude allow for an accurate calculation of the predicted anisotropies in linear perturbation theory. A particular cosmological model is characterized by a dozen or so parameters: the rate of expansion, the spatial curvature, the baryon content, the cold dark matter and neutrino contribution, the cosmological constant (vacuum energy), the reionization parameter (optical depth to the last scattering surface), and various primordial spectrum parameters like the amplitude and tilt of the adiabatic and isocurvature spectra, the amount of gravitational waves, non-Gaussian effects, etc. All these parameters can now be fed into very fast CMB codes called CMBFAST [93] and CAMB [94], that compute the predicted temperature and polarization anisotropies to better than 1% accuracy, and thus can be used to compare with observations.

These two satellites will improve both the sensitivity, down to μK , and the resolution, down to arc minutes, with respect to the previous COBE satellite, thanks to large numbers of microwave horns of various sizes, positioned at specific angles, and also thanks to recent advances in detector technology, with high electron mobility transistor amplifiers (HEMTs) for frequencies below 100 GHz and bolometers for higher frequencies. The primary advantage of HEMTs is their ease of use and speed, with a typical sensitivity of $0.5 \text{ mKs}^{1/2}$, while the advantage of bolometers is their tremendous sensitivity, better than $0.1 \text{ mKs}^{1/2}$, see Ref. [97]. This will allow cosmologists to extract information from around 3000 multipoles! Since most of the cosmological parameters have specific signatures in the height and position of the first few acoustic peaks, the higher the resolution, the more peaks one is expected to see, and thus the better the accuracy with which one will be able to measure those parameters, see Table 2.

Although the satellite probes were designed for the accurate measurement of the CMB temperature anisotropies, there are other experiments, like balloon-borne and ground interferometers [90]. Prob-

ably the most important objective of the future satellites (beyond WMAP) will be the measurement of the CMB polarization anisotropies, discovered by DASI in November 2002 [98], and confirmed a few months later by WMAP with greater accuracy [20], see Fig. 24. These anisotropies were predicted by models of structure formation and indeed found at the level of microKelvin sensitivities, where the new satellites were aiming at. The complementary information contained in the polarization anisotropies already provides much more stringent constraints on the cosmological parameters than from the temperature anisotropies alone. However, in the future, Planck and CMB pol will have much better sensitivities. In particular, the curl-curl component of the polarization power spectra is nowadays the only means we have to determine the tensor (gravitational wave) contribution to the metric perturbations responsible for temperature anisotropies, see Fig. 30. If such a component is found, one could constraint very precisely the model of inflation from its spectral properties, specially the tilt [91].

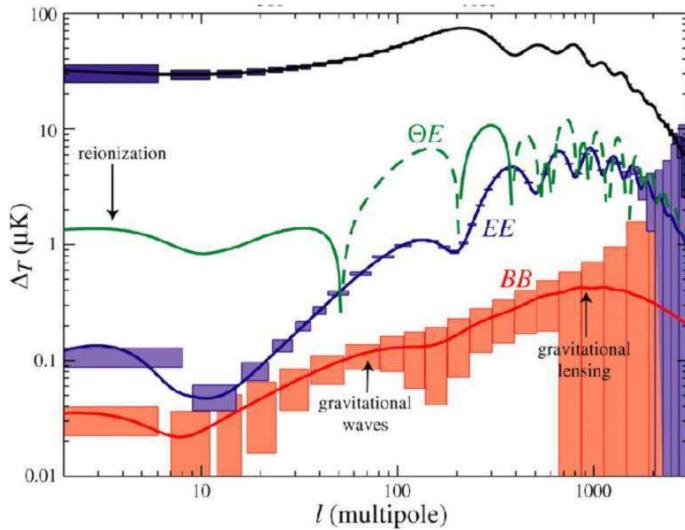


Fig. 30: Theoretical predictions for the four non-zero CMB temperature-polarization spectra as a function of multipole moment, together with the expectations from Planck. From Ref. [92].

4.5 From metric perturbations to large scale structure

If inflation is responsible for the metric perturbations that gave rise to the temperature anisotropies observed in the microwave background, then the primordial spectrum of density inhomogeneities induced by the same metric perturbations should also be responsible for the present large scale structure [99]. This simple connection allows for more stringent tests on the inflationary paradigm for the generation of metric perturbations, since it relates the large scales (of order the present horizon) with the smallest scales (on galaxy scales). This provides a very large lever arm for the determination of primordial spectra parameters like the tilt, the nature of the perturbations, whether adiabatic or isocurvature, the geometry of the universe, as well as its matter and energy content, whether CDM, HDM or mixed CHDM.

4.51 The galaxy power spectrum

As metric perturbations enter the causal horizon during the radiation or matter era, they create density fluctuations via gravitational attraction of the potential wells. The density contrast δ can be deduced from the Einstein equations in linear perturbation theory, see Eq. (166),

$$\delta_k \equiv \frac{\delta\rho_k}{\rho} = \left(\frac{k}{aH}\right)^2 \frac{2}{3} \Phi_k = \left(\frac{k}{aH}\right)^2 \frac{2+2\omega}{5+3\omega} \mathcal{R}_k, \quad (239)$$

where we have assumed $K = 0$, and used Eq. (198). From this expression one can compute the power spectrum, at horizon crossing, of matter density perturbations induced by inflation, see Eq. (210),

$$P(k) = \langle |\delta_k|^2 \rangle = A \left(\frac{k}{aH} \right)^n, \quad (240)$$

with n given by the scalar tilt (212), $n = 1 + 2\eta - 6\epsilon$. This spectrum reduces to a Harrison-Zel'dovich spectrum (100) in the slow-roll approximation: $\eta, \epsilon \ll 1$.

Since perturbations evolve after entering the horizon, the power spectrum will not remain constant. For scales entering the horizon well after matter domination ($k^{-1} \gg k_{\text{eq}}^{-1} \simeq 81 \text{ Mpc}$), the metric perturbation has not changed significantly, so that $\mathcal{R}_k(\text{final}) = \mathcal{R}_k(\text{initial})$. Then Eq. (239) determines the final density contrast in terms of the initial one. On smaller scales, there is a linear transfer function $T(k)$, which may be defined as [76]

$$\mathcal{R}_k(\text{final}) = T(k) \mathcal{R}_k(\text{initial}). \quad (241)$$

To calculate the transfer function one has to specify the initial condition with the relative abundance of photons, neutrinos, baryons and cold dark matter long before horizon crossing. The most natural condition is that the abundances of all particle species are uniform on comoving hypersurfaces (with constant total energy density). This is called the *adiabatic* condition, because entropy is conserved independently for each particle species X , i.e. $\delta\rho_X = \dot{\rho}_X \delta t$, given a perturbation in time from a comoving hypersurface, so

$$\frac{\delta\rho_X}{\rho_X + p_X} = \frac{\delta\rho_Y}{\rho_Y + p_Y}, \quad (242)$$

where we have used the energy conservation equation for each species, $\dot{\rho}_X = -3H(\rho_X + p_X)$, valid to first order in perturbations. It follows that each species of radiation has a common density contrast δ_r , and each species of matter has also a common density contrast δ_m , with the relation $\delta_m = \frac{3}{4}\delta_r$.

Given the adiabatic condition, the transfer function is determined by the physical processes occurring between horizon entry and matter domination. If the radiation behaves like a perfect fluid, its density perturbation oscillates during this era, with decreasing amplitude. The matter density contrast living in this background does not grow appreciably before matter domination because it has negligible self-gravity. The transfer function is therefore given roughly by, see Eq. (103),

$$T(k) = \begin{cases} 1, & k \ll k_{\text{eq}} \\ (k/k_{\text{eq}})^2, & k \gg k_{\text{eq}} \end{cases} \quad (243)$$

The perfect fluid description of the radiation is far from being correct after horizon entry, because roughly half of the radiation consists of neutrinos whose perturbation rapidly disappears through free streaming. The photons are also not a perfect fluid because they diffuse significantly, for scales below the Silk scale, $k_S^{-1} \sim 1 \text{ Mpc}$. One might then consider the opposite assumption, that the radiation has zero perturbation after horizon entry. Then the matter density perturbation evolves according to

$$\ddot{\delta}_k + 2H\dot{\delta}_k + (c_s^2 k_{\text{ph}}^2 - 4\pi G\rho) \delta_k = 0, \quad (244)$$

which corresponds to the equation of a damped harmonic oscillator. The zero-frequency oscillator defines the Jeans wavenumber, $k_J = \sqrt{4\pi G\rho/c_s^2}$. For $k \ll k_J$, δ_k grows exponentially on the dynamical timescale, $\tau_{\text{dyn}} = \text{Im } \omega^{-1} = (4\pi G\rho)^{-1/2} = \tau_{\text{grav}}$, which is the time scale for gravitational collapse. One can also define the Jeans length,

$$\lambda_J = \frac{2\pi}{k_J} = c_s \sqrt{\frac{\pi}{G\rho}}, \quad (245)$$

which separates gravitationally stable from unstable modes. If we define the pressure response timescale as the size of the perturbation over the sound speed, $\tau_{\text{pres}} \sim \lambda/c_s$, then, if $\tau_{\text{pres}} > \tau_{\text{grav}}$, gravitational collapse of a perturbation can occur before pressure forces can respond to restore hydrostatic equilibrium (this occurs for $\lambda > \lambda_J$). On the other hand, if $\tau_{\text{pres}} < \tau_{\text{grav}}$, radiation pressure prevents gravitational collapse and there are damped acoustic oscillations (for $\lambda < \lambda_J$).

We will consider now the behaviour of modes within the horizon during the transition from the radiation ($c_s^2 = 1/3$) to the matter era ($c_s^2 = 0$). The growing mode solution increases only by a factor of 2 between horizon entry and the epoch when matter starts to dominate, i.e. $y = 1$. The transfer function is therefore again roughly given by Eq. (243). Since the radiation consists roughly half of neutrinos, which free stream, and half of photons, which either form a perfect fluid or just diffuse, neither the perfect fluid nor the free-streaming approximation looks very sensible. A more precise calculation is needed, including: neutrino free streaming around the epoch of horizon entry; the diffusion of photons around the same time, for scales below Silk scale; the diffusion of baryons along with the photons, and the establishment after matter domination of a common matter density contrast, as the baryons fall into the potential wells of cold dark matter. All these effects apply separately, to first order in the perturbations, to each Fourier component, so that a linear transfer function is produced. There are several parametrizations in the literature, but the one which is more widely used is that of Ref. [100],

$$T(k) = \left[1 + \left(ak + (bk)^{3/2} + (ck)^2 \right)^\nu \right]^{-1/\nu}, \quad \nu = 1.13, \quad (246)$$

$$a = 6.4 (\Omega_M h)^{-1} h^{-1} \text{ Mpc}, \quad (247)$$

$$b = 3.0 (\Omega_M h)^{-1} h^{-1} \text{ Mpc}, \quad (248)$$

$$c = 1.7 (\Omega_M h)^{-1} h^{-1} \text{ Mpc}. \quad (249)$$

We see that the behaviour estimated in Eq. (243) is roughly correct, although the break at $k = k_{\text{eq}}$ is not at all sharp, see Fig. 31. The transfer function, which encodes the solution to linear equations, ceases to be valid when the density contrast becomes of order 1. After that, the highly nonlinear phenomenon of gravitational collapse takes place, see Fig. 31.

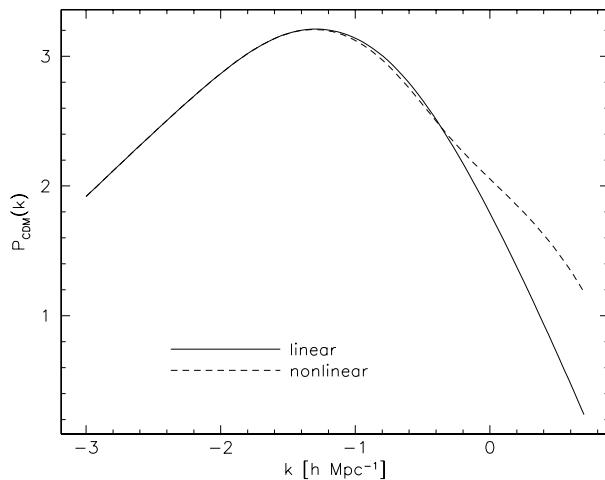


Fig. 31: The CDM power spectrum $P(k)$ as a function of wavenumber k , in logarithmic scale, normalized to the local abundance of galaxy clusters, for an Einstein-de Sitter universe with $h = 0.5$. The solid (dashed) curve shows the linear (non-linear) power spectrum. While the linear power spectrum falls off like k^{-3} , the non-linear power-spectrum illustrates the increased power on small scales due to non-linear effects, at the expense of the large-scale structures. From Ref. [41].

4.52 The new redshift catalogs, 2dF and Sloan Digital Sky Survey

Our view of the large-scale distribution of luminous objects in the universe has changed dramatically during the last 25 years: from the simple pre-1975 picture of a distribution of field and cluster galaxies, to the discovery of the first single superstructures and voids, to the most recent results showing an almost regular web-like network of interconnected clusters, filaments and walls, separating huge nearly empty volumes. The increased efficiency of redshift surveys, made possible by the development of spectrographs and – specially in the last decade – by an enormous increase in multiplexing gain (i.e. the ability to collect spectra of several galaxies at once, thanks to fibre-optic spectrographs), has allowed us not only to do *cartography* of the nearby universe, but also to statistically characterize some of its properties, see Ref. [101]. At the same time, advances in theoretical modeling of the development of structure, with large high-resolution gravitational simulations coupled to a deeper yet limited understanding of how to form galaxies within the dark matter halos, have provided a more realistic connection of the models to the observable quantities [102]. Despite the large uncertainties that still exist, this has transformed the study of cosmology and large-scale structure into a truly quantitative science, where theory and observations can progress side by side.

I will concentrate on two of the new catalogs, which are taking data at the moment and which have changed the field, the 2-degree-Field (2dF) Catalog and the Sloan Digital Sky Survey (SDSS). The advantages of multi-object fibre spectroscopy have been pushed to the extreme with the construction of the 2dF spectrograph for the prime focus of the Anglo-Australian Telescope [42]. This instrument is able to accommodate 400 automatically positioned fibres over a 2 degree in diameter field. This implies a density of fibres on the sky of approximately 130 deg^{-2} , and an optimal match to the galaxy counts for a magnitude $b_J \simeq 19.5$, similar to that of previous surveys like the ESP, with the difference that with such an area yield, the same number of redshifts as in the ESP survey can be collected in about 10 exposures, or slightly more than one night of telescope time with typical 1 hour exposures. This is the basis of the 2dF galaxy redshift survey. Its goal is to measure redshifts for more than 250,000 galaxies with $b_J < 19.5$. In addition, a faint redshift survey of 10,000 galaxies brighter than $R = 21$ will be done over selected fields within the two main strips of the South and North Galactic Caps. The survey has now finished, with a quarter of a million redshifts. The final result can be seen in Ref. [42].

The most ambitious and comprehensive galaxy survey currently in progress is without any doubt the Sloan Digital Sky Survey [43]. The aim of the project is, first of all, to observe photometrically the whole Northern Galactic Cap, 30° away from the galactic plane (about 10^4 deg^2) in five bands, at limiting magnitudes from 20.8 to 23.3. The expectation is to detect around 50 million galaxies and around 10^8 star-like sources. This has already led to the discovery of several high-redshift ($z > 4$) quasars, including the highest-redshift quasar known, at $z = 5.0$, see Ref. [43]. Using two fibre spectrographs carrying 320 fibres each, the spectroscopic part of the survey will then collect spectra from about 10^6 galaxies with $r' < 18$ and 10^5 AGNs with $r' < 19$. It will also select a sample of about 10^5 red luminous galaxies with $r' < 19.5$, which will be observed spectroscopically, providing a nearly volume-limited sample of early-type galaxies with a median redshift of $z \simeq 0.5$, that will be extremely valuable to study the evolution of clustering. The data that is coming from these catalogs is so outstanding that already cosmologists are using them for the determination of the cosmological parameters of the standard model of cosmology. The main outcome of these catalogs is the linear power spectrum of matter fluctuations that give rise to galaxies, and clusters of galaxies. It covers from the large scales of order Gigaparsecs, the realm of the unvirialised superclusters, to the small scales of hundreds of kiloparsecs, where the Lyman- α systems can help reconstruct the linear power spectrum, since they are less sensitive to the nonlinear growth of perturbations.

As often happens in particle physics, not always are observations from a single experiment sufficient to isolate and determine the precise value of the parameters of the standard model. We mentioned in the previous Section that some of the cosmological parameters created similar effects in the temperature anisotropies of the microwave background. We say that these parameters are *degenerate* with

respect to the observations. However, often one finds combinations of various experiments/observations which break the degeneracy, for example by depending on a different combination of parameters. This is precisely the case with the cosmological parameters, as measured by a combination of large-scale structure observations, microwave background anisotropies, Supernovae Ia observations and Hubble Space Telescope measurements. It is expected that in the near future we will be able to determine the parameters of the standard cosmological model with great precision from a combination of several different experiments.

5. CONCLUSION

In the last five years we have seen a true revolution in the quality and quantity of cosmological data that has allowed cosmologists to determine most of the cosmological parameters with a few percent accuracy and thus fix a Standard Model of Cosmology. The art of measuring the cosmos has developed so rapidly and efficiently that one may be tempted of renaming this science as Cosmonomy, leaving the word Cosmology for the theories of the Early Universe. In summary, we now know that the stuff we are made of – baryons – constitutes just about 4% of all the matter/energy in the Universe, while 25% is dark matter – perhaps a new particle species related to theories beyond the Standard Model of Particle Physics –, and the largest fraction, 70%, some form of diffuse tension also known as dark energy – perhaps a cosmological constant. The rest, about 1%, could be in the form of massive neutrinos.

Nowadays, a host of observations – from CMB anisotropies and large scale structure to the age and the acceleration of the universe – all converge towards these values, see Fig. 25. Fortunately, we will have, within this decade, new satellite experiments like Planck, CMBpol, SNAP as well as deep galaxy catalogs from Earth, to complement and precisely pin down the values of the Standard Model cosmological parameters below the percent level, see Table 1.

All these observations would not make much sense without the encompassing picture of the inflationary paradigm that determines the homogeneous and isotropic background on top of which it imprints an approximately scale invariant gaussian spectrum of adiabatic fluctuations. At present all observations are consistent with the predictions of inflation and hopefully in the near future we may have information, from the polarization anisotropies of the microwave background, about the scale of inflation, and thus about the physics responsible for the early universe dynamics.

ACKNOWLEDGEMENTS

I would like to thank the organizers of the CERN-JINR European School of High Energy Physics 2004, and very specially Matteo Cavalli-Sforza, without whom this wonderful school would not have been the success it was. This work was supported in part by a CICYT project FPA2003-04597.

References

- [1] A. Einstein, Sitz. Preuss. Akad. Wiss. Phys. **142** (1917) (§4); Ann. Phys. **69** (1922) 436.
- [2] A. Friedmann, Z. Phys. **10** (1922) 377.
- [3] E.P. Hubble, Publ. Nat. Acad. Sci. **15** (1929) 168.
- [4] G. Gamow, Phys. Rev. **70** (1946) 572; Phys. Rev. **74** (1948) 505.
- [5] A.A. Penzias and R.W. Wilson, Astrophys. J. **142** (1965) 419.
- [6] S. Weinberg, *Gravitation and Cosmology* (John Wiley & Sons, San Francisco, 1972).
- [7] S. Perlmutter *et al.* [Supernova Cosmology Project], Astrophys. J. **517** (1999) 565. Home Page
<http://scp.berkeley.edu/>

- [8] W. L. Freedman *et al.*, *Astrophys. J.* **553** (2001) 47.
- [9] A. G. Riess *et al.* [High-z Supernova Search], *Astron. J.* **116** (1998) 1009.
Home Page <http://cfa-www.harvard.edu/cfa/oir/Research/supernova/>
- [10] J. García-Bellido in *European School of High Energy Physics*, ed. A. Olchevski (CERN report 2000-007); e-print Archive: hep-ph/0004188.
- [11] R. A. Knop *et al.*, e-print Archive: astro-ph/0309368.
- [12] A. G. Riess *et al.*, e-print Archive: astro-ph/0402512.
- [13] S. Weinberg, *Rev. Mod. Phys.* **61** (1989) 1; S. M. Carroll, *Living Rev. Rel.* **4** (2001) 1; T. Padmanabhan, *Phys. Rept.* **380** (2003) 235; P. J. E. Peebles and B. Ratra, *Rev. Mod. Phys.* **75** (2003) 559.
- [14] The SuperNova/Acceleration Probe Home page: <http://snap.lbl.gov/>
- [15] E.W. Kolb and M.S. Turner, “The Early Universe”, Addison Wesley (1990).
- [16] R. Srianand, P. Petitjean and C. Ledoux, *Nature* **408** (2000) 931.
- [17] S. Burles, K.M. Nollett, J.N. Truran, M.S. Turner, *Phys. Rev. Lett.* **82** (1999) 4176; S. Burles, K.M. Nollett, M.S. Turner, “Big-Bang Nucleosynthesis: Linking Inner Space and Outer Space”, e-print Archive: astro-ph/9903300.
- [18] K. A. Olive, G. Steigman and T. P. Walker, *Phys. Rept.* **333** (2000) 389; J. P. Kneller and G. Steigman, “BBN For Pedestrians,” *New J. Phys.* **6** (2004) 117.
- [19] Particle Data Group Home Page, <http://pdg.web.cern.ch/pdg/>
- [20] D. N. Spergel *et al.*, *Astrophys. J. Suppl.* **148** (2003) 175.
- [21] J.C. Mather *et al.*, *Astrophys. J.* **512** (1999) 511.
- [22] R.H. Dicke, P.J.E. Peebles, P.G. Roll and D.T. Wilkinson, *Astrophys. J.* **142** (1965) 414.
- [23] C.L. Bennett *et al.*, *Astrophys. J.* **464** (1996) L1.
- [24] P.J.E. Peebles, “Principles of Physical Cosmology”, Princeton U.P. (1993).
- [25] T. Padmanabhan, “Structure Formation in the Universe”, Cambridge U.P. (1993).
- [26] E.R. Harrison, *Phys. Rev. D* **1** (1970) 2726; Ya. B. Zel'dovich, *Astron. Astrophys.* **5** (1970) 84.
- [27] The IRAS Point Source Catalog Web page:
<http://www-astro.physics.ox.ac.uk/~wjs/psc2.html>
- [28] P.J. Steinhardt, in *Particle and Nuclear Astrophysics and Cosmology in the Next Millennium*, ed. by E.W. kolb and R. Peccei (World Scientific, Singapore, 1995).
- [29] W.L. Freedman, “Determination of cosmological parameters”, Nobel Symposium (1998), e-print Archive: hep-ph/9905222.
- [30] S. Refsdal, *Mon. Not. R. Astr. Soc.* **128** (1964) 295; **132** (1966) 101.
- [31] R.D. Blandford and T. Kundić, “Gravitational Lensing and the Extragalactic Distance Scale”, e-print Archive: astro-ph/9611229.

- [32] N.A. Grogin and R. Narayan, *Astrophys. J.* **464** (1996) 92.
- [33] M. Birkinshaw, *Phys. Rep.* **310** (1999) 97.
- [34] The Chandra X-ray observatory Home Page: <http://chandra.harvard.edu/>
- [35] W. L. Freedman *et al.*, *Astrophys. J.* **553** (2001) 47
- [36] F. Zwicky, *Helv. Phys. Acata* **6** (1933) 110.
- [37] K.C. Freeman, *Astrophys. J.* **160** (1970) 811.
- [38] C.M. Baugh *et al.*, “Ab initio galaxy formation”, e-print Archive: astro-ph/9907056; *Astrophys. J.* **498** (1998) 405.
- [39] F. Prada *et al.*, *Astrophys. J.* **598** (2003) 260.
- [40] C.L. Sarazin, *Rev. Mod. Phys.* **58** (1986) 1.
- [41] M. Bartelmann *et al.*, *Astron. & Astrophys.* **330** (1998) 1; M. Bartelmann and P. Schneider, *Phys. Rept.* **340** (2001) 291
- [42] M. Colless *et al.* [2dFGRS Coll.], “The 2dF Galaxy Redshift Survey: Final Data Release,” Archive: astro-ph/0306581. The 2dFGRS Home Page: <http://www.mso.anu.edu.au/2dFGRS/>
- [43] M. Tegmark *et al.* [SDSS Collaboration], *Astrophys. J.* **606** (2004) 702; *Phys. Rev. D* **69** (2004) 103501. The SDSS Home Page: <http://www.sdss.org/sdss.html>
- [44] G.G. Raffelt, “Dark Matter: Motivation, Candidates and Searches”, European Summer School of High Energy Physics 1997. CERN Report pp. 235-278, e-print Archive: hep-ph/9712538.
- [45] P.J.E. Peebles, “Testing GR on the Scales of Cosmology,” e-print Archive: astro-ph/0410284.
- [46] M. C. Gonzalez-Garcia, “Global analysis of neutrino data,” e-print Archive: hep-ph/0410030.
- [47] S.D. Tremaine and J.E. Gunn, *Phys. Rev. Lett.* **42** (1979) 407; J. Madsen, *Phys. Rev. D* **44** (1991) 999.
- [48] J. Primack, D. Seckel and B. Sadoulet, *Ann. Rev. Nucl. Part. Sci.* **38** (1988) 751; N.E. Booth, B. Cabrera and E. Fiorini, *Ann. Rev. Nucl. Part. Sci.* **46** (1996) 471.
- [49] C. Kraus *et al.*, “Final results from phase II of the Mainz neutrino mass search in tritium beta decay,” e-print Archive: hep-ex/0412056.
- [50] H. V. Klapdor-Kleingrothaus *et al.*, *Mod. Phys. Lett. A* **16** (2001) 2409; *Mod. Phys. Lett. A* **18** (2003) 2243.
- [51] R. Bernabei *et al.*, “Dark matter search,” *Riv. Nuovo Cim.* **26N1** (2003) 1. DAMA Home Page, <http://www.lngs.infn.it/lngs/htexts/dama/welcome.html>
- [52] K. A. Olive, “Dark matter candidates in supersymmetric models,” e-print Archive: hep-ph/0412054.
- [53] D. S. Akerib *et al.* [CDMS Collaboration], *Phys. Rev. Lett.* **93** (2004) 211301; D. S. Akerib *et al.*, *Phys. Rev. D* **68** (2003) 082002.
- [54] B. Ahmed *et al.*, *Nucl. Phys. Proc. Suppl.* **124** (2003) 193; *Astropart. Phys.* **19** (2003) 691. UKDMC Home Page at <http://hepwww.rl.ac.uk/ukdmc/>

- [55] M. Bravin et al., *Astropart. Phys.* **12** (1999) 107.
- [56] J. I. Collar et al., *Phys. Rev. Lett.* **85** (2000) 3083.
- [57] G. Jungman, M. Kamionkowski and K. Griest, *Phys. Rep.* **267** (1996) 195.
- [58] The Alpha Magnetic Spectrometer Home Page: <http://ams.cern.ch/AMS/>
- [59] F. Halzen et al., *Phys. Rep.* **307** (1998) 243; M. Ackermann *et al.* [The AMANDA Collaboration], “Search for extraterrestrial point sources of high energy neutrinos with AMANDA-II using data collected in 2000-2002,” e-print Archive: astro-ph/0412347.
- [60] D.A. Vandenberg, M. Bolte and P.B. Stetson, *Ann. Rev. Astron. Astrophys.* **34** (1996) 461; e-print Archive: astro-ph/9605064.
- [61] L. M. Krauss, *Phys. Rept.* **333** (2000) 33.
- [62] B. Chaboyer, P. Demarque, P.J. Kernan and L.M. Krauss, *Science* **271** (1996) 957; *Astrophys. J.* **494** (1998) 96.
- [63] C.H. Lineweaver, *Science* **284** (1999) 1503.
- [64] D. Scott, J. Silk and M. White, *Science* **268** (1995) 829; W. Hu, N. Sugiyama and J. Silk, *Nature* **386** (1997) 37; E. Gawiser and J. Silk, *Phys. Rept.* **333** (2000) 245.
- [65] J. Silk, *Nature* **215** (1967) 1155;
- [66] A. Albrecht, R. A. Battye and J. Robinson, *Phys. Rev. Lett.* **79** (1997) 4736; N. Turok, U. L. Pen, U. Seljak, *Phys. Rev. D* **58** (1998) 023506; L. Pogosian, *Int. J. Mod. Phys. A* **16S1C** (2001) 1043.
- [67] U. Seljak et al., e-print Archive: astro-ph/0407372.
- [68] V. Barger, D. Marfatia and A. Tregre, *Phys. Lett. B* **595** (2004) 55; P. Crotty, J. Lesgourgues and S. Pastor, *Phys. Rev. D* **69** (2004) 123007; S. Hannestad, “Neutrino mass bounds from cosmology,” e-print Archive: hep-ph/0412181.
- [69] H. V. Peiris *et al.*, *Astrophys. J. Suppl.* **148** (2003) 213; P. Crotty, J. García-Bellido, J. Lesgourgues and A. Riazuelo, *Phys. Rev. Lett.* **91** (2003) 171301; J. Valiviita and V. Muhonen, *Phys. Rev. Lett.* **91** (2003) 131302; M. Beltrán, J. García-Bellido, J. Lesgourgues and A. Riazuelo, *Phys. Rev. D* **70** (2004) 103530; K. Moodley, M. Bucher, J. Dunkley, P. G. Ferreira and C. Skordis, *Phys. Rev. D* **70** (2004) 103520; C. Gordon and K. A. Malik, *Phys. Rev. D* **69** (2004) 063508; F. Ferrer, S. Rasanen and J. Valiviita, *JCAP* **0410** (2004) 010; H. Kurki-Suonio, V. Muhonen and J. Valiviita, e-print Archive: astro-ph/0412439; M. Beltrán, J. García-Bellido, J. Lesgourgues, A. R. Liddle and A. Slosar, “Bayesian model selection and isocurvature perturbations,” e-print Archive: astro-ph/0501477.
- [70] J. García-Bellido, *Phil. Trans. R. Soc. Lond. A* **357** (1999) 3237.
- [71] A. Guth, *Phys. Rev. D* **23** (1981) 347.
- [72] A.D. Linde, *Phys. Lett.* **108B** (1982) 389.
- [73] A. Albrecht and P.J. Steinhardt, *Phys. Rev. Lett.* **48** (1982) 1220.
- [74] For a personal historical account, see A. Guth, “The Inflationary Universe”, Perseus Books (1997).
- [75] A.D. Linde, “Particle Physics and Inflationary Cosmology”, Harwood Academic Press (1990).

- [76] A.R. Liddle and D.H. Lyth, Phys. Rep. **231** (1993) 1.
- [77] J.M. Bardeen, Phys. Rev. D **22** (1980) 1882.
- [78] V.F. Mukhanov, H.A. Feldman and R.H. Brandenberger, Phys. Rep. **215** (1992) 203.
- [79] M. Abramowitz and I. Stegun, “Handbook of Mathematical Functions”, Dover (1972).
- [80] J. García-Bellido and D. Wands, Phys. Rev. D **53** (1996) 5437; D. Wands, K. A. Malik, D. H. Lyth and A. R. Liddle, Phys. Rev. D **62** (2000) 043527.
- [81] D.H. Lyth and A. Riotto, Phys. Rep. **314** (1999) 1.
- [82] R.K. Sachs and A.M. Wolfe, Astrophys. J. **147** (1967) 73.
- [83] E.R. Harrison, Rev. Mod. Phys. **39** (1967) 862; L.F. Abbott and R.K. Schaefer, Astrophys. J. **308** (1986) 546.
- [84] E.F. Bunn, A.R. Liddle and M. White, Phys. Rev. D **54** (1996) 5917.
- [85] A.A. Starobinsky, Sov. Astron. Lett. **11** (1985) 133.
- [86] S.M. Carroll, W.H. Press and E.L. Turner, Ann. Rev. Astron. Astrophys. **30** (1992) 499.
- [87] P. de Bernardis *et al.*, New Astron. Rev. **43** (1999) 289; P. D. Mauskopf *et al.* [Boomerang Coll.], Astrophys. J. **536** (2000) L59. Boomerang Home Page: <http://oberon.roma1.infn.it/boomerang/>
- [88] Microwave Anisotropy Probe Home Page: <http://map.gsfc.nasa.gov/>
- [89] Planck Surveyor Home Page: <http://astro.estec.esa.nl/Planck/>
- [90] M. Tegmark Home Page: <http://www.hep.upenn.edu/~max/cmb/experiments.html>
- [91] M. Kamionkowski and A. Kosowsky, Ann. Rev. Nucl. Part. Sci. **49** (1999) 77.
- [92] W. Hu and S. Dodelson, Ann. Rev. Astron. Astrophys. **40** (2002) 171.
- [93] U. Seljak and M. Zaldarriaga, CMBFAST code Home Page: <http://www.cmbfast.org/>
- [94] A. Lewis and A. Challinor, CAMB code Home Page: <http://camb.info/>
- [95] CMB Polarization experiment Home Page:
http://www.mssl.ucl.ac.uk/www_astro/submm/CMBpol1.html
- [96] ACT experiment Home Page: <http://www.hep.upenn.edu/act/>
- [97] L.A. Page, “Measuring the anisotropy in the CMB”, e-print Archive: astro-ph/9911199.
- [98] J. Kovac et al., Nature **420** (2002) 772; DASI Home Page: <http://astro.uchicago.edu/dasi/>
- [99] A.R. Liddle and D.H. Lyth, “Cosmological Inflation and Large Scale Structure”, Cambridge University Press (2000).
- [100] J.R. Bond and G. Efstathiou, Astrophys. J. **285** (1984) L45.
- [101] G. Efstathiou *et al.* (Eds.) “Large-scale structure in the universe”, Phil. Trans. R. Soc. Lond. A **357** (1999) 1-198.
- [102] B. Moore, Phil. Trans. R. Soc. Lond. A **357** (1999) 3259.