19. BIG-BANG COSMOLOGY

Written July 2001 by K.A. Olive (University of Minnesota) and J.A. Peacock (University of Edinburgh). Revised November 2003.

19.1. Introduction to Standard Big-Bang Model

The observed expansion of the Universe [1,2,3] is a natural (almost inevitable) result of any homogeneous and isotropic cosmological model based on general relativity. However, by itself, the Hubble expansion does not provide sufficient evidence for what we generally refer to as the Big-Bang model of cosmology. While general relativity is in principle capable of describing the cosmology of any given distribution of matter, it is extremely fortunate that our Universe appears to be homogeneous and isotropic on large scales. Together, homogeneity and isotropy allow us to extend the Copernican Principle to the Cosmological Principle, stating that all spatial positions in the Universe are essentially equivalent.

The formulation of the Big-Bang model began in the 1940s with the work of George Gamow and his collaborators, Alpher and Herman. In order to account for the possibility that the abundances of the elements had a cosmological origin, they proposed that the early Universe which was once very hot and dense (enough so as to allow for the nucleosynthetic processing of hydrogen), and has expanded and cooled to its present state [4,5]. In 1948, Alpher and Herman predicted that a direct consequence of this model is the presence of a relic background radiation with a temperature of order a few K [6,7]. Of course this radiation was observed 16 years later as the microwave background radiation [8]. Indeed, it was the observation of the 3 K background radiation that singled out the Big-Bang model as the prime candidate to describe our Universe. Subsequent work on Big-Bang nucleosynthesis further confirmed the necessity of our hot and dense past. (See the review on BBN—Sec. 20 of this *Review* for a detailed discussion of BBN.) These relativistic cosmological models face severe problems with their initial conditions, to which the best modern solution is inflationary cosmology, discussed in Sec. 19.3.5. If correct, these ideas would strictly render the term 'Big Bang' redundant, since it was first coined by Hoyle to represent a criticism of the lack of understanding of the initial conditions.

19.1.1. *The Robertson-Walker Universe*:

The observed homogeneity and isotropy enable us to describe the overall geometry and evolution of the Universe in terms of two cosmological parameters accounting for the spatial curvature and the overall expansion (or contraction) of the Universe. These two quantities appear in the most general expression for a space-time metric which has a (3D) maximally symmetric subspace of a 4D space-time, known as the Robertson-Walker metric:

$$ds^2 = dt^2 - R^2(t)\left[\frac{dr^2}{1-kr^2} + r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\right] . \tag{19.1}$$

Note that we adopt $c = 1$ throughout. By rescaling the radial coordinate, we can choose the curvature constant k to take only the discrete values $+1$, -1, or 0 corresponding to closed, open, or spatially flat geometries. In this case, it is often more convenient to re-express the metric as

$$ds^2 = dt^2 - R^2(t)\left[d\chi^2 + S_k^2(\chi)\,(d\theta^2 + \sin^2\theta\, d\phi^2)\right] , \tag{19.2}$$

where the function $S_k(\chi)$ is $(\sin\chi, \chi, \sinh\chi)$ for $k = (+1, 0, -1)$. The coordinate r (in Eq. (19.1)) and the 'angle' χ (in Eq. (19.2)) are both dimensionless; the dimensions are carried by $R(t)$, which is the cosmological scale factor which determines proper distances in terms of the comoving coordinates. A common alternative is to define a dimensionless scale factor, $a(t) = R(t)/R_0$, where $R_0 \equiv R(t_0)$ is R at the present epoch. It is also sometimes convenient to define a dimensionless or conformal time coordinate, η, by $d\eta = dt/R(t)$. Along constant spatial sections, the proper time is defined by the time coordinate, t. Similarly, for $dt = d\theta = d\phi = 0$, the proper distance is given by $R(t)\chi$. For standard texts on cosmological models see *e.g.*, Refs. [9–14].

19.1.2. *The redshift*:

The cosmological redshift is a direct consequence of the Hubble expansion, determined by $R(t)$. A local observer detecting light from a distant emitter sees a redshift in frequency. We can define the redshift as

$$z \equiv \frac{\nu_1 - \nu_2}{\nu_2} \simeq \frac{v_{12}}{c} , \tag{19.3}$$

where ν_1 is the frequency of the emitted light, ν_2 is the observed frequency and v_{12} is the relative velocity between the emitter and the observer. While the definition, $z = (\nu_1 - \nu_2)/\nu_2$ is valid on all distance scales, relating the redshift to the relative velocity in this simple way is only true on small scales (*i.e.*, less than cosmological scales) such that the expansion velocity is non-relativistic. For light signals, we can use the metric given by Eq. (19.1) and $ds^2 = 0$ to write

$$\frac{v_{12}}{c} = \dot{R}\,\delta r = \frac{\dot{R}}{R}\,\delta t = \frac{\delta R}{R} = \frac{R_2 - R_1}{R_1} , \tag{19.4}$$

where $\delta r(\delta t)$ is the radial coordinate (temporal) separation between the emitter and observer. Thus, we obtain the simple relation between the redshift and the scale factor

$$1 + z = \frac{\nu_1}{\nu_2} = \frac{R_2}{R_1} . \tag{19.5}$$

This result does not depend on the non-relativistic approximation.

19.1.3. ***The Friedmann-Lemaître equations of motion*:**

The cosmological equations of motion are derived from Einstein's equations

$$\mathcal{R}_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} \mathcal{R} = 8\pi G_{\rm N} T_{\mu\nu} + \Lambda g_{\mu\nu} \ . \tag{19.6}$$

Gliner [15] and Zeldovich [16] seem to have pioneered the modern view, in which the Λ term is taken to the rhs and interpreted as particle-physics processes yielding an effective energy–momentum tensor $T_{\mu\nu}$ for the vacuum of $\Lambda g_{\mu\nu}/8\pi G_{\rm N}$. It is common to assume that the matter content of the Universe is a perfect fluid, for which

$$T_{\mu\nu} = -p g_{\mu\nu} + (p+\rho)\, u_\mu u_\nu \ , \tag{19.7}$$

where $g_{\mu\nu}$ is the space-time metric described by Eq. (19.1), p is the isotropic pressure, ρ is the energy density and $u = (1, 0, 0, 0)$ is the velocity vector for the isotropic fluid in co-moving coordinates. With the perfect fluid source, Einstein's equations lead to the Friedmann-Lemaître equations

$$H^2 \equiv \left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi\ G_{\rm N}\ \rho}{3} - \frac{k}{R^2} + \frac{\Lambda}{3} \ , \tag{19.8}$$

and

$$\frac{\ddot{R}}{R} = \frac{\Lambda}{3} - \frac{4\pi G_{\rm N}}{3}\ (\rho + 3p) \ , \tag{19.9}$$

where $H(t)$ is the Hubble parameter and Λ is the cosmological constant. The first of these is sometimes called the Friedmann equation. Energy conservation via $T^{\mu\nu}{}_{;\mu} = 0$, leads to a third useful equation [which can also be derived from Eq. (19.8) and Eq. (19.9)]

$$\dot{\rho} = -3H\,(\rho + p) \ . \tag{19.10}$$

Eq. (19.10) can also be simply derived as a consequence of the first law of thermodynamics.

Eq. (19.8) has a simple classical mechanical analog if we neglect (for the moment) the cosmological term Λ. By interpreting $-k/R^2$ as a "total energy", then we see that the evolution of the Universe is governed by a competition between the potential energy, $8\pi G_{\rm N}\rho/3$ and the kinetic term $(\dot{R}/R)^2$. For $\Lambda = 0$, it is clear that the Universe must be expanding or contracting (except at the turning point prior to collapse in a closed Universe). The ultimate fate of the Universe is determined by the curvature constant k. For $k = +1$, the Universe will recollapse in a finite time, whereas for $k = 0, -1$, the Universe will expand indefinitely. These simple conclusions can be altered when $\Lambda \neq 0$ or more generally with some component with $(\rho + 3p) < 0$.

19.1.4. *Definition of cosmological parameters*:

In addition to the Hubble parameter, it is useful to define several other measurable cosmological parameters. The Friedmann equation can be used to define a critical density such that $k = 0$ when $\Lambda = 0$,

$$\begin{aligned} \rho_c \equiv \frac{3H^2}{8\pi\, G_{\rm N}} &= 1.88 \times 10^{-26}\, h^2 \text{ kg m}^{-3} \\ &= 1.05 \times 10^{-5}\, h^2 \text{ GeV cm}^{-3} , \end{aligned} \tag{19.11}$$

where the scaled Hubble parameter, h, is defined by

$$\begin{aligned} H &\equiv 100\, h \text{ km s}^{-1} \text{ Mpc}^{-1} \\ \Rightarrow H^{-1} &= 9.78\, h^{-1} \text{ Gyr} \\ &= 2998\, h^{-1} \text{ Mpc} . \end{aligned} \tag{19.12}$$

The cosmological density parameter $\Omega_{\rm tot}$ is defined as the energy density relative to the critical density,

$$\Omega_{\rm tot} = \rho/\rho_c . \tag{19.13}$$

Note that one can now rewrite the Friedmann equation as

$$k/R^2 = H^2(\Omega_{\rm tot} - 1) , \tag{19.14}$$

From Eq. (19.14), one can see that when $\Omega_{\rm tot} > 1$, $k = +1$ and the Universe is closed, when $\Omega_{\rm tot} < 1$, $k = -1$ and the Universe is open, and when $\Omega_{\rm tot} = 1$, $k = 0$, and the Universe is spatially flat.

It is often necessary to distinguish different contributions to the density. It is therefore convenient to define present-day density parameters for pressureless matter (Ω_m) and relativistic particles (Ω_r), plus the quantity $\Omega_\Lambda = \Lambda/3H^2$. In more general models, we may wish to drop the assumption that the vacuum energy density is constant, and we therefore denote the present-day density parameter of the vacuum by Ω_v. The Friedmann equation then becomes

$$k/R_0^2 = H_0^2(\Omega_m + \Omega_r + \Omega_v - 1) , \tag{19.15}$$

where the subscript 0 indicates present-day values. Thus, it is the sum of the densities in matter, relativistic particles and vacuum that determines the overall sign of the curvature. Note that the quantity $-k/R_0^2 H_0^2$ is sometimes referred to as Ω_k. This usage is unfortunate: it encourages one to think of curvature as a contribution to the energy density of the Universe, which is not correct.

19.1.5. *Standard Model solutions*:

Much of the history of the Universe in the standard Big-Bang model can be easily described by assuming that either matter or radiation dominates the total energy density. During inflation or perhaps even today if we are living in an accelerating Universe, domination by a cosmological constant or some other form of dark energy should be considered. In the following, we shall delineate the solutions to the Friedmann equation when a single component dominates the energy density. Each component is distinguished by an equation of state parameter $w = p/\rho$.

19.1.5.1. *Solutions for a general equation of state:*

Let us first assume a general equation of state parameter for a single component, w which is constant. In this case, Eq. (19.10) can be written as $\dot{\rho} = -3(1+w)\rho\dot{R}/R$ and is easily integrated to yield

$$\rho \propto R^{-3(1+w)} . \tag{19.16}$$

Note that at early times when R is small, the less singular curvature term k/R^2 in the Friedmann equation can be neglected so long as $w > -1/3$. Curvature domination occurs at rather late times (if a cosmological constant term does not dominate sooner). For $w \neq -1$, one can insert this result into the Friedmann equation Eq. (19.8) and if one neglects the curvature and cosmological constant terms, it is easy to integrate the equation to obtain,

$$R(t) \propto t^{2/[3(1+w)]} . \tag{19.17}$$

19.1.5.2. *A Radiation-dominated Universe:*

In the early hot and dense Universe, it is appropriate to assume an equation of state corresponding to a gas of radiation (or relativistic particles) for which $w = 1/3$. In this case, Eq. (19.16) becomes $\rho \propto R^{-4}$. The "extra" factor of $1/R$ is due to the cosmological redshift; not only is the number density of particles in the radiation background decreasing as R^{-3} since volumes scales as R^3, but in addition, each particle's energy is decreasing as $E \propto \nu \propto R^{-1}$. Similarly, one can substitute $w = 1/3$ into Eq. (19.17) to obtain

$$R(t) \propto t^{1/2} ; \qquad H = 1/2t . \tag{19.18}$$

19.1.5.3. *A Matter-dominated Universe:*

At relatively late times, non-relativistic matter eventually dominates the energy density over radiation (see Sec. 19.3.8). A pressureless gas ($w = 0$) leads to the expected dependence $\rho \propto R^{-3}$ from Eq. (19.16) and, if $k = 0$, we get

$$R(t) \propto t^{2/3} ; \qquad H = 2/3t . \tag{19.19}$$

19.1.5.4. *A Universe dominated by vacuum energy:*

If there is a dominant source of vacuum energy, V_0, it would act as a cosmological constant with $\Lambda = 8\pi G_{\rm N} V_0$ and equation of state $w = -1$. In this case, the solution to the Friedmann equation is particularly simple and leads to an exponential expansion of the Universe

$$R(t) \propto e^{\sqrt{\Lambda/3}t} . \tag{19.20}$$

A key parameter is the equation of state of the vacuum, $w \equiv p/\rho$: this need not be the $w = -1$ of Λ, and may not even be constant [17,18,19]. It is now common to use w to stand for this vacuum equation of state, rather than of any other constituent of the Universe, and we use the symbol in this sense hereafter. We generally assume w to be independent of time, and where results relating to the vacuum are quoted without an explicit w dependence, we have adopted $w = -1$.

The presence of vacuum energy can dramatically alter the fate of the Universe. For example, if $\Lambda < 0$, the Universe will eventually recollapse independent of the sign of k. For large values of Λ (larger than the Einstein static value needed to halt any cosmological expansion or contraction), even a closed Universe will expand forever. One way to quantify this is the deceleration parameter, q_0, defined as

$$q_0 = -\left.\frac{R\ddot{R}}{\dot{R}^2}\right|_0 = \frac{1}{2}\Omega_m + \Omega_r + \frac{(1+3w)}{2}\Omega_v . \tag{19.21}$$

This equation shows us that $w < -1/3$ for the vacuum may lead to an accelerating expansion. Astonishingly, it appears that such an effect has been observed in the Supernova Hubble diagram [20–23] (see Fig. 19.1 below); current data indicate that vacuum energy is indeed the largest contributor to the cosmological density budget, with $\Omega_v = 0.72 \pm 0.05$ and $\Omega_m = 0.28 \pm 0.05$ if $k = 0$ is assumed [23].

The nature of this dominant term is presently uncertain, but much effort is being invested in dynamical models (*e.g.*, rolling scalar fields), under the catch-all heading of "quintessence."

19.2. Introduction to Observational Cosmology

19.2.1. ***Fluxes, luminosities, and distances***:

The key quantities for observational cosmology can be deduced quite directly from the metric.

(1) The *proper* transverse size of an object seen by us to subtend an angle $d\psi$ is its comoving size $d\psi\, S_k(\chi)$ times the scale factor at the time of emission:

$$d\ell = d\psi\ R_0 S_k(\chi)/(1+z) . \tag{19.22}$$

(2) The apparent flux density of an object is deduced by allowing its photons to flow through a sphere of current radius $R_0 S_k(\chi)$; but photon energies and arrival rates are redshifted, and the bandwidth $d\nu$ is reduced. The observed photons at frequency ν_0 were

emitted at frequency $\nu_0(1+z)$, so the flux density is the luminosity at this frequency, divided by the total area, divided by $1+z$:

$$S_\nu(\nu_0) = \frac{L_\nu([1+z]\nu_0)}{4\pi R_0^2 S_k^2(\chi)(1+z)} \ . \tag{19.23}$$

These relations lead to the following common definitions:

$$\begin{aligned} \text{angular-diameter distance:} \quad & D_{\rm A} = (1+z)^{-1} R_0 S_k(\chi) \\ \text{luminosity distance:} \quad & D_{\rm L} = (1+z)\ R_0 S_k(\chi) \end{aligned} \tag{19.24}$$

These distance-redshift relations are expressed in terms of observables by using the equation of a null radial geodesic ($R(t)d\chi = dt$) plus the Friedmann equation:

$$\begin{aligned} R_0 d\chi = & \frac{1}{H(z)}\, dz = \frac{1}{H_0}\ \Big[(1-\Omega_m - \Omega_v - \Omega_r)(1+z)^2 \\ & + \ \Omega_v(1+z)^{3+3w} \ + \ \Omega_m(1+z)^3 + \Omega_r(1+z)^4\Big]^{-1/2} \ dz \ . \end{aligned} \tag{19.25}$$

The main scale for the distance here is the Hubble length, $1/H_0$.

The flux density is the product of the specific intensity I_ν and the solid angle $d\Omega$ subtended by the source: $S_\nu = I_\nu \ d\Omega$. Combining the angular size and flux-density relations thus gives the relativistic version of surface-brightness conservation:

$$I_\nu(\nu_0) = \frac{B_\nu([1+z]\nu_0)}{(1+z)^3} \ , \tag{19.26}$$

where B_ν is surface brightness (luminosity emitted into unit solid angle per unit area of source). We can integrate over ν_0 to obtain the corresponding total or bolometric formula:

$$I_{\rm tot} = \frac{B_{\rm tot}}{(1+z)^4} \ . \tag{19.27}$$

This cosmology-independent form expresses Liouville's Theorem: photon phase-space density is conserved along rays.

19.2.2. *Distance data and geometrical tests of cosmology*:

In order to confront these theoretical predictions with data, we have to bridge the divide between two extremes. Nearby objects may have their distances measured quite easily, but their radial velocities are dominated by deviations from the ideal Hubble flow, which typically have a magnitude of several hundred $\mathrm{km\,s^{-1}}$. On the other hand, objects at redshifts $z \gtrsim 0.01$ will have observed recessional velocities that differ from their ideal values by $\lesssim 10\%$, but absolute distances are much harder to supply in this case. The traditional solution to this problem is the construction of the distance ladder: an interlocking set of methods for obtaining relative distances between various classes of object, which begins with absolute distances at the 10 to 100 pc level and terminates with galaxies at significant redshifts. This is reviewed in the review on Global cosmological parameters—Sec. 21 of this *Review*.

By far the most exciting development in this area has been the use of type Ia Supernovae (SNe), which now allow measurement of relative distances with 5% precision. In combination with Cepheid data from the HST key project on the distance scale, SNe results are the dominant contributor to the best modern value for H_0: $72\,\mathrm{km\,s^{-1}Mpc^{-1}} \pm 10\%$ [24]. Better still, the analysis of high-z SNe has allowed the first meaningful test of cosmological geometry to be carried out: as shown in Fig. 19.1 and Fig. 19.2, a combination of supernova data and measurements of microwave-background anisotropies strongly favors a $k = 0$ model dominated by vacuum energy. (See the review on Global cosmological parameters—Sec. 21 of this *Review* for a more comprehensive review of Hubble parameter determinations.)

19.2.3. *Age of the Universe*:

The most striking conclusion of relativistic cosmology is that the Universe has not existed forever. The dynamical result for the age of the Universe may be written as

$$\begin{aligned} H_0 t_0 &= \int_0^\infty \frac{dz}{(1+z)H(z)} \\ &= \int_0^\infty \frac{dz}{(1+z)\,\left[(1+z)^2(1+\Omega_m z) - z(2+z)\Omega_v\right]^{1/2}} \ , \end{aligned} \tag{19.28}$$

where we have neglected Ω_r and chosen $w = -1$. Over the range of interest ($0.1 \lesssim \Omega_m \lesssim 1$, $|\Omega_v| \lesssim 1$), this exact answer may be approximated to a few % accuracy by

$$H_0 t_0 \simeq \tfrac{2}{3}\,(0.7\Omega_m + 0.3 - 0.3\Omega_v)^{-0.3} \ . \tag{19.29}$$

For the special case that $\Omega_m + \Omega_v = 1$, the integral in Eq. (19.28) can be expressed analytically as

$$H_0 t_0 = \frac{2}{3\sqrt{\Omega_v}} \ln \frac{1+\sqrt{\Omega_v}}{\sqrt{1-\Omega_v}} \quad (\Omega_m < 1) \ . \tag{19.30}$$

The most accurate means of obtaining ages for astronomical objects is based on the natural clocks provided by radioactive decay. The use of these clocks is complicated by a lack of knowledge of the initial conditions of the decay. In the Solar System, chemical fractionation of different elements helps pin down a precise age for the pre-Solar nebula of 4.6 Gyr, but for stars it is necessary to attempt an a priori calculation of the relative abundances of nuclei that result from supernova explosions. In this way, a lower limit for the age of stars in the local part of the Milky Way of about 11 Gyr is obtained [25].

The other major means of obtaining cosmological age estimates is based on the theory of stellar evolution. In principle, the main-sequence turnoff point in the color-magnitude diagram of a globular cluster should yield a reliable age. However, these have been controversial owing to theoretical uncertainties in the evolution model, as well as

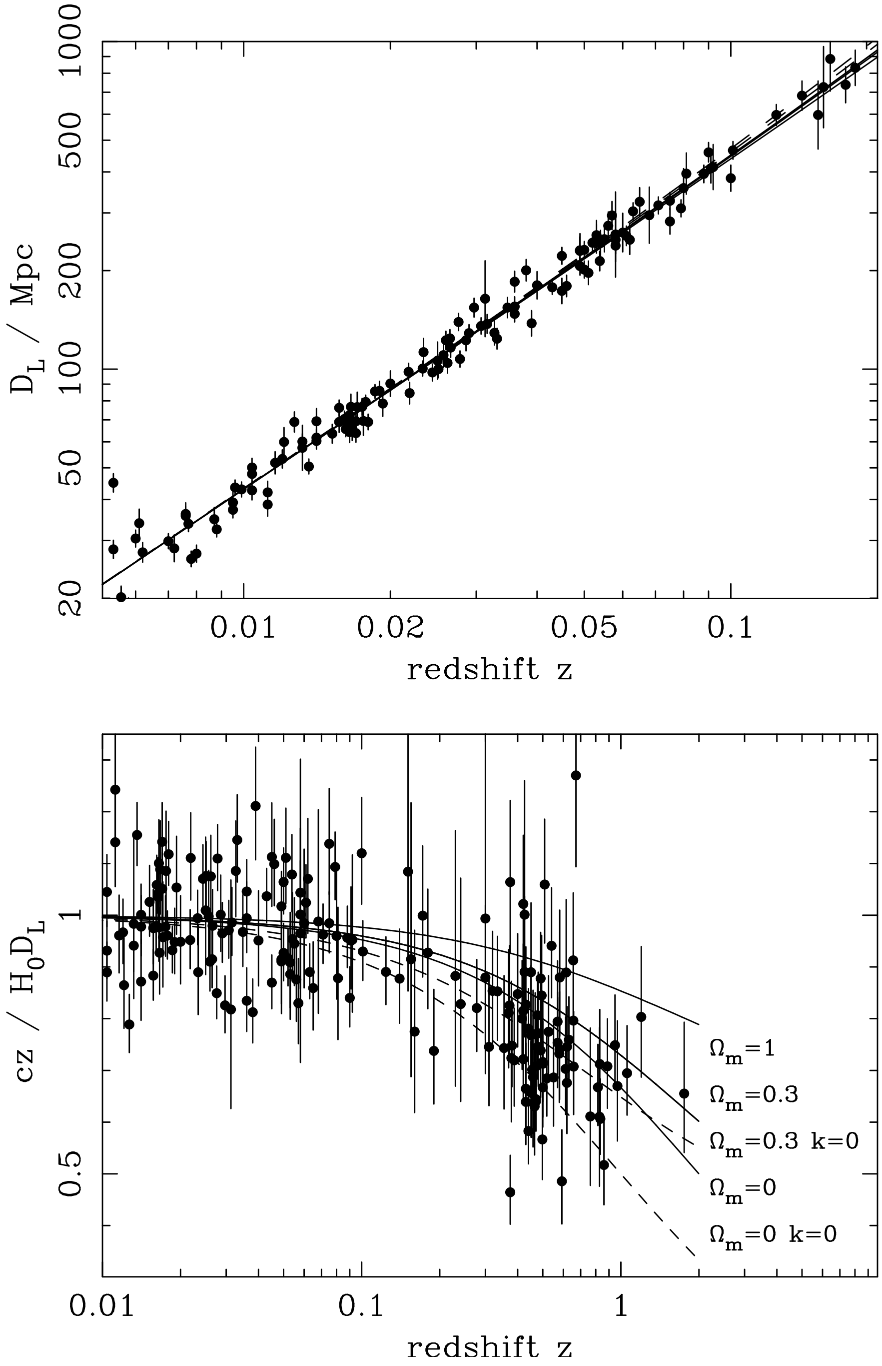

Figure 19.1: The type Ia supernova Hubble diagram [20–22]. The first panel shows that for $z \ll 1$ the large-scale Hubble flow is indeed linear and uniform; the second panel shows an expanded scale, with the linear trend divided out, and with the redshift range extended to show how the Hubble law becomes nonlinear. ($\Omega_r = 0$ is assumed.) Comparison with the prediction of Friedmann-Lemaître models appears to favor a vacuum-dominated Universe.

observational uncertainties in the distance, dust extinction and metallicity of clusters. The present consensus favors ages for the oldest clusters of about 12 Gyr [26,27].

These methods are all consistent with the age deduced from studies of structure formation, using the microwave background and large-scale structure: $t_0 = 13.7 \pm 0.2$ Gyr [28], where the extra accuracy comes at the price of assuming the Cold Dark Matter model to be true.

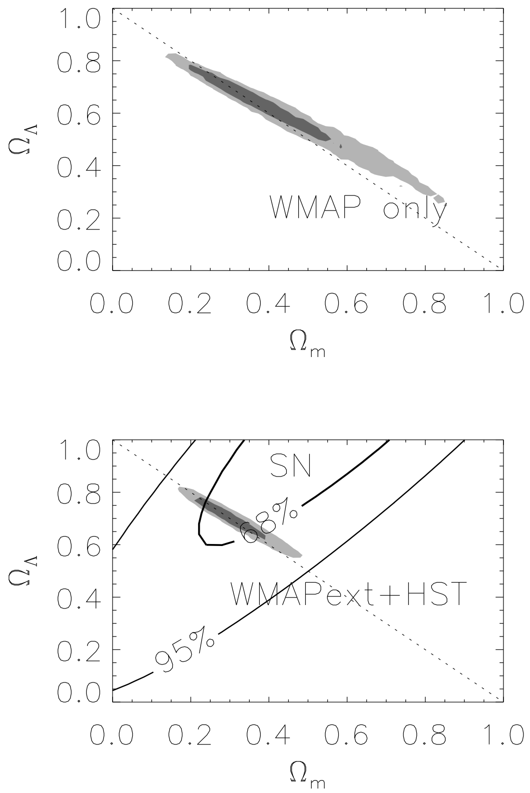

Figure 19.2: Likelihood-based confidence contours [28] over the plane Ω_Λ (*i.e.* Ω_v assuming $w = -1$) vs Ω_m. The SNe Ia results very nearly constrain $\Omega_v - \Omega_m$, whereas the results of CMB anisotropies (from the first-year WMAP data) favor a flat model with $\Omega_v + \Omega_m \simeq 1$. The intersection of these constraints is the most direct (but far from the only) piece of evidence favoring a flat model with $\Omega_m \simeq 0.3$.

19.2.4. *Horizon, isotropy, flatness problems*:

For photons, the radial equation of motion is just $c\,dt = R\,d\chi$. How far can a photon get in a given time? The answer is clearly

$$\Delta\chi = \int_{t_1}^{t_2} \frac{dt}{R(t)} \equiv \Delta\eta \;, \tag{19.31}$$

i.e., just the interval of conformal time. We can replace dt by $dR/\dot{R}$, which the Friedmann equation says is $\propto dR/\sqrt{\rho R^2}$ at early times. Thus, this integral converges if $\rho R^2 \to \infty$ as $t_1 \to 0$, otherwise it diverges. Provided the equation of state is such that ρ changes faster than R^{-2}, light signals can only propagate a finite distance between the Big Bang and the present; there is then said to be a particle horizon. Such a horizon therefore exists in conventional Big-Bang models, which are dominated by radiation ($\rho \propto R^{-4}$) at early times.

At late times, the integral for the horizon is largely determined by the matter-dominated phase, for which

$$D_{\rm H} = R_0 \, \chi_{\rm H} \equiv R_0 \int_0^{t(z)} \frac{dt}{R(t)} \simeq \frac{6000}{\sqrt{\Omega z}} \, h^{-1} \, {\rm Mpc} \quad (z \gg 1) \; . \tag{19.32}$$

The horizon at the time of formation of the microwave background ('last scattering': $z \simeq 1100$) was thus of order 100 Mpc in size, subtending an angle of about 1°. Why then are the large number of causally disconnected regions we see on the microwave sky all at the same temperature? The Universe is very nearly isotropic and homogeneous, even though the initial conditions appear not to permit such a state to be constructed.

A related problem is that the $\Omega = 1$ Universe is unstable:

$$\Omega(a) - 1 = \frac{\Omega - 1}{1 - \Omega + \Omega_v a^2 + \Omega_m a^{-1} + \Omega_r a^{-2}} \; , \tag{19.33}$$

where Ω with no subscript is the total density parameter, and $a(t) = R(t)/R_0$. This requires $\Omega(t)$ to be unity to arbitrary precision as the initial time tends to zero; a universe of non-zero curvature today requires very finely tuned initial conditions.

19.3. The Hot Thermal Universe

19.3.1. *Thermodynamics of the early Universe*:

As alluded to above, we expect that much of the early Universe can be described by a radiation-dominated equation of state. In addition, through much of the radiation-dominated period, thermal equilibrium is established by the rapid rate of particle interactions relative to the expansion rate of the Universe (see Sec. 19.3.3 below). In equilibrium, it is straightforward to compute the thermodynamic quantities, ρ, p, and the entropy density, s. In general, the energy density for a given particle type i can be written as

$$\rho_i = \int E_i \, dn_{q_i} \; , \tag{19.34}$$

with the density of states given by

$$dn_{q_i} = \frac{g_i}{2\pi^2} \left(\exp[(E_{q_i} - \mu_i)/T_i] \pm 1 \right)^{-1} \; q_i^2 \, dq_i \; , \tag{19.35}$$

where g_i counts the number of degrees of freedom for particle type i, $E_{q_i}^2 = m_i^2 + q_i^2$, μ_i is the chemical potential, and the $\pm$ corresponds to either Fermi or Bose statistics. Similarly, we can define the pressure of a perfect gas as

$$p_i = \frac{1}{3} \int \frac{q_i^2}{E_i} \, dn_{q_i} \; . \tag{19.36}$$

The number density of species i is simply

$$n_i = \int dn_{q_i} \ , \qquad (19.37)$$

and the entropy density is

$$s_i = \frac{\rho_i + p_i - \mu_i n_i}{T_i} \ , \qquad (19.38)$$

In the Standard Model, a chemical potential is often associated with baryon number, and since the net baryon density relative to the photon density is known to be very small (of order 10^{-10}), we can neglect the chemical potential.

For photons, we can compute all of the thermodynamic quantities rather easily. Taking $g_i = 2$ for the 2 photon polarization states, we have

$$\rho_\gamma = \frac{\pi^2}{15} T^4 \ ; \quad p_\gamma = \frac{1}{3}\rho_\gamma \ ; \quad s_\gamma = \frac{4\rho_\gamma}{3T} \ ; \quad n_\gamma = \frac{2\zeta(3)}{\pi^2} T^3 \ , \qquad (19.39)$$

with $2\zeta(3)/\pi^2 \simeq 0.2436$. Note that Eq. (19.10) can be converted into an equation for entropy conservation. Recognizing that $\dot{p} = s\dot{T}$, Eq. (19.10) becomes

$$d(sR^3)/dt = 0 \ . \qquad (19.40)$$

For radiation, this corresponds to the relationship between expansion and cooling, $T \propto R^{-1}$ in an adiabatically expanding Universe. Note also that both s and n_γ scale as T^3.

19.3.2. *Radiation content of the Early Universe*:

At the very high temperatures associated with the early Universe, massive particles are pair produced, and are part of the thermal bath. If for a given particle species i we have $T \gg m_i$, then we can neglect the mass in Eq. (19.34) to Eq. (19.38), and the thermodynamic quantities are easily computed as in Eq. (19.39). In general, we can approximate the energy density (at high temperatures) by including only those particles with $m_i \ll T$. In this case, we have

$$\rho = \left(\sum_B g_B + \frac{7}{8} \sum_F g_F \right) \frac{\pi^2}{30} T^4 \equiv \frac{\pi^2}{30} N(T) T^4 \ , \qquad (19.41)$$

where $g_{B(F)}$ is the number of degrees of freedom of each boson (fermion) and the sum runs over all boson and fermion states with $m \ll T$. The factor of 7/8 is due to the difference between the Fermi and Bose integrals. Eq. (19.41) defines the effective number of degrees of freedom, $N(T)$, by taking into account new particle degrees of freedom as the temperature is raised. This quantity is plotted in Fig. 19.3 [29].

Temperature	New Particles	$4N(T)$
$T < m_e$	γ's + ν's	29
$m_e < T < m_\mu$	$e^\pm$	43
$m_\mu < T < m_\pi$	$\mu^\pm$	57
$m_\pi < T < T_c{}^\dagger$	π's	69
$T_c < T < m_{\rm strange}$	π's + $u, \bar{u}, d, \bar{d}$ + gluons	205
$m_s < T < m_{\rm charm}$	$s, \bar{s}$	247
$m_c < T < m_\tau$	$c, \bar{c}$	289
$m_\tau < T < m_{\rm bottom}$	$\tau^\pm$	303
$m_b < T < m_{\rm W,Z}$	$b, \bar{b}$	345
$m_{W,Z} < T < m_{\rm Higgs}$	$W^\pm, Z$	381
$m_H < T < m_{\rm top}$	H^0	385
$m_t < T$	$t, \bar{t}$	427

$^\dagger T_c$ corresponds to the confinement-deconfinement transition between quarks and hadrons.

The value of $N(T)$ at any given temperature depends on the particle physics model. In the standard SU(3) $\times$ SU(2) $\times$ U(1) model, we can specify $N(T)$ up to temperatures of O(100) GeV. The change in N (ignoring mass effects) can be seen in the following table.

At higher temperatures, $N(T)$ will be model dependent. For example, in the minimal SU(5) model, one needs to add 24 states to $N(T)$ for the X and Y gauge bosons, another 24 from the adjoint Higgs, and another 6 (in addition to the 4 already counted in $W^\pm, Z$, and H) from the $\overline{5}$ of Higgs. Hence for $T > m_{\rm X}$ in minimal SU(5), $N(T) = 160.75$. In a supersymmetric model this would at least double, with some changes possibly necessary in the table if the lightest supersymmetric particle has a mass below m_t.

In the radiation-dominated epoch, Eq. (19.10) can be integrated (neglecting the T-dependence of N) giving us a relationship between the age of the Universe and its temperature

$$t = \left(\frac{90}{32\pi^3 G_{\rm N} N(T)} \right)^{1/2} T^{-2} \ , \tag{19.42}$$

Put into a more convenient form

$$t\, T^2_{\rm MeV} = 2.4 [N(T)]^{-1/2} \ , \tag{19.43}$$

where t is measured in seconds and $T_{\rm MeV}$ in units of MeV.

19.3.3. *Neutrinos and equilibrium*:

Due to the expansion of the Universe, certain rates may be too slow to either establish or maintain equilibrium. Quantitatively, for each particle i, as a minimal condition for equilibrium, we will require that some rate Γ_i involving that type be larger than the expansion rate of the Universe or

$$\Gamma_i > H \ . \tag{19.44}$$

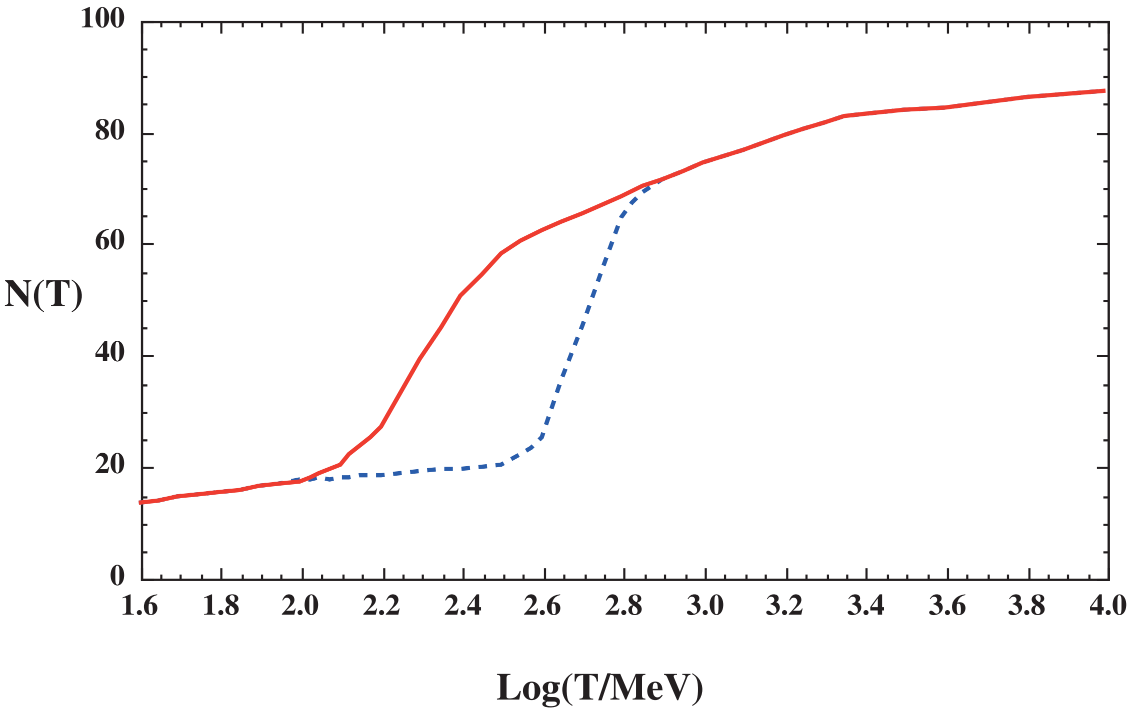

Figure 19.3: The effective numbers of relativistic degrees of freedom as a function of temperature. The sharp drop corresponds to the quark-hadron transition. The solid curve assume a QCD scale of 150 MeV, while the dashed curve assumes 450 MeV.

Recalling that the age of the Universe is determined by H^{-1}, this condition is equivalent to requiring that on average, at least one interaction has occurred over the lifetime of the Universe.

A good example for a process which goes in and out of equilibrium is the weak interactions of neutrinos. On dimensional grounds, one can estimate the thermally averaged scattering cross section

$$\langle \sigma v \rangle \sim O(10^{-2}) T^2 / m_{\mathrm{W}}^4 \tag{19.45}$$

for $T \lesssim m_{\mathrm{W}}$. Recalling that the number density of leptons is $n \propto T^3$, we can compare the weak interaction rate, $\Gamma \sim n \langle \sigma v \rangle$, with the expansion rate,

$$\begin{aligned} H = \left(\frac{8\pi G_{\mathrm{N}} \rho}{3} \right)^{1/2} &= \left(\frac{8\pi^3}{90} N(T) \right)^{1/2} T^2 / M_{\mathrm{P}} \\ &\sim 1.66 N(T)^{1/2} T^2 / M_{\mathrm{P}} . \end{aligned} \tag{19.46}$$

The Planck mass $M_{\mathrm{P}} = G_{\mathrm{N}}^{-1/2} = 1.22 \times 10^{19}$ GeV.

Neutrinos will be in equilibrium when $\Gamma_{\mathrm{wk}} > H$ or

$$T > (500\, m_{\mathrm{W}}^4 / M_{\mathrm{P}})^{1/3} \sim 1 \ \mathrm{MeV} \ . \tag{19.47}$$

The temperature at which these rates are equal is commonly referred to as the neutrino decoupling or freeze-out temperature and is defined by $\Gamma(T_d) = H(T_d)$.

At very high temperatures, the Universe is too young for equilibrium to have been established. For $T \gg m_{\rm W}$, we should write $\langle \sigma v \rangle \sim O(10^{-2})/T^2$, so that $\Gamma \sim 10^{-2}T$. Thus at temperatures $T \gtrsim 10^{-2} M_{\rm P}/\sqrt{N}$, equilibrium will not have been established.

For $T < T_d$, neutrinos drop out of equilibrium. The Universe becomes transparent to neutrinos and their momenta simply redshift with the cosmic expansion. The effective neutrino temperature will simply fall with $T \sim 1/R$.

Soon after decoupling, $e^{\pm}$ pairs in the thermal background begin to annihilate (when $T \lesssim m_e$). Because the neutrinos are decoupled, the energy released due to annihilation heats up the photon background relative to the neutrinos. The change in the photon temperature can be easily computed from entropy conservation. The neutrino entropy must be conserved separately from the entropy of interacting particles. A straightforward computation yields

$$T_\nu = (4/11)^{1/3}\, T_\gamma \simeq 1.9 \text{ K} . \tag{19.48}$$

Today, the total entropy density is therefore given by

$$s = \frac{4}{3}\frac{\pi^2}{30}\left(2 + \frac{21}{4}(T_\nu/T_\gamma)^3\right) T_\gamma^3 = \frac{4}{3}\frac{\pi^2}{30}\left(2 + \frac{21}{11}\right) T_\gamma^3 = 7.04\, n_\gamma \ . \tag{19.49}$$

Similarly, the total relativistic energy density today is given by

$$\rho_r = \frac{\pi^2}{30}\left[2 + \frac{21}{4}(T_\nu/T_\gamma)^4\right] T_\gamma^4 \simeq 1.68 \rho_\gamma \ . \tag{19.50}$$

In practice, a small correction is needed to this, since neutrinos are not totally decoupled at $e^{\pm}$ annihilation: the effective number of massless neutrino species is 3.04, rather than 3 [30].

This expression ignores neutrino rest masses, but current oscillation data require at least one neutrino eigenstate to have a mass exceeding 0.05 eV. In this minimal case, $\Omega_\nu h^2 = 5 \times 10^{-4}$, so the neutrino contribution to the matter budget would be negligibly small (which is our normal assumption). However, a nearly degenerate pattern of mass eigenstates could allow larger densities, since oscillation experiments only measure differences in m^2 values. Note that a 0.05-eV neutrino becomes non-relativistic at redshift $z = 2.2$, so the above expression for the total present relativistic density is really only an extrapolation. However, neutrinos are almost certainly relativistic at all epochs where the radiation content of the universe is dynamically significant.

19.3.4. *Field Theory and Phase transitions*:

It is very likely that the Universe has undergone one or more phase transitions during the course of its evolution [31–34]. Our current vacuum state is described by $\mathrm{SU}(3)_c\times$ $\mathrm{U}(1)_{\mathrm{em}}$ which in the Standard Model is a remnant of an unbroken $\mathrm{SU}(3)_c\times$ $\mathrm{SU}(2)_L\times$ $\mathrm{U}(1)_Y$ gauge symmetry. Symmetry breaking occurs when a non-singlet gauge field (the Higgs field in the Standard Model) picks up a non-vanishing vacuum expectation value, determined by a scalar potential. For example, a simple (non-gauged) potential describing symmetry breaking is $V(\phi) = \frac{1}{4}\lambda\phi^4 - \frac{1}{2}\mu^2\phi^2 + V(0)$. The resulting expectation value is simply $\langle\phi\rangle = \mu/\sqrt{\lambda}$.

In the early Universe, finite temperature radiative corrections typically add terms to the potential of the form $\phi^2 T^2$. Thus, at very high temperatures, the symmetry is restored and $\langle\phi\rangle = 0$. As the Universe cools, depending on the details of the potential, symmetry breaking will occur via a first order phase transition in which the field tunnels through a potential barrier, or via a second order transition in which the field evolves smoothly from one state to another (as would be the case for the above example potential).

The evolution of scalar fields can have a profound impact on the early Universe. The equation of motion for a scalar field ϕ can be derived from the energy-momentum tensor

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - \frac{1}{2}g_{\mu\nu}\partial_\rho\phi\partial^\rho\phi - g_{\mu\nu}V(\phi)\ . \tag{19.51}$$

By associating $\rho = T_{00}$ and $p = R^{-2}(t)T_{ii}$ we have

$$\begin{aligned}\rho &= \frac{1}{2}\dot{\phi}^2 + \frac{1}{2}R^{-2}(t)(\nabla\phi)^2 + V(\phi)\\ p &= \frac{1}{2}\dot{\phi}^2 - \frac{1}{6}R^{-2}(t)(\nabla\phi)^2 - V(\phi)\end{aligned} \tag{19.52}$$

and from Eq. (19.10) we can write the equation of motion (by considering a homogeneous region, we can ignore the gradient terms)

$$\ddot{\phi} + 3H\dot{\phi} = -\partial V/\partial\phi\ . \tag{19.53}$$

19.3.5. *Inflation*:

In Sec. 19.2.4, we discussed some of the problems associated with the standard Big-Bang model. However, during a phase transition, our assumptions of an adiabatically expanding universe are generally not valid. If, for example, a phase transition occurred in the early Universe such that the field evolved slowly from the symmetric state to the global minimum, the Universe may have been dominated by the vacuum energy density associated with the potential near $\phi \approx 0$. During this period of slow evolution, the energy density due to radiation will fall below the vacuum energy density, $\rho \ll V(0)$. When this happens, the expansion rate will be dominated by the constant V(0) and we obtain the exponentially expanding solution given in Eq. (19.20). When the field evolves towards the global minimum it will begin to oscillate about the minimum, energy will be released during its decay and a hot thermal universe will be restored. If released fast enough, it will produce radiation at a temperature $N{T_R}^4 \lesssim V(0)$. In this reheating process entropy has been created and the final value of RT is greater than the initial value of RT. Thus, we see that, during a phase transition, the relation $RT \sim$ constant need not hold true. This is the basis of the inflationary Universe scenario [35–37].

If during the phase transition the value of RT changed by a factor of $O(10^{29})$, the cosmological problems discussed above would be solved. The observed isotropy would be generated by the immense expansion; one small causal region could get blown up and hence our entire visible Universe would have been in thermal contact some time in the past. In addition, the density parameter Ω would have been driven to 1 (with exponential precision). Density perturbations will be stretched by the expansion, $\lambda \sim R(t)$. Thus it will appear that $\lambda \gg H^{-1}$ or that the perturbations have left the horizon, where in fact the size of the causally connected region is now no longer simply H^{-1}. However, not only does inflation offer an explanation for large scale perturbations, it also offers a source for the perturbations themselves through quantum fluctuations.

Early models of inflation were based on a first order phase transition of a Grand Unified theory [38]. Although these models led to sufficient exponential expansion, completion of the transition through bubble percolation did not occur. Later models of inflation [39,40], also based on Grand Unified symmetry breaking, through second order transitions were also doomed. While they successfully inflated and reheated, and in fact produced density perturbations due to quantum fluctuations during the evolution of the scalar field, they predicted density perturbations many orders of magnitude too large. Most models today are based on an unknown symmetry breaking involving a new scalar field, the inflaton, ϕ.

19.3.6. *Baryogenesis*:

The Universe appears to be populated exclusively with matter rather than antimatter. Indeed antimatter is only detected in accelerators or in cosmic rays. However, the presence of antimatter in the latter is understood to be the result of collisions of primary particles in the interstellar medium. There is in fact strong evidence against primary forms of antimatter in the Universe. Furthermore, the density of baryons compared to the density of photons is extremely small, $\eta \sim 10^{-10}$.

The production of a net baryon asymmetry requires baryon number violating interactions, C and CP violation and a departure from thermal equilibrium [41]. The first two of these ingredients are expected to be contained in grand unified theories as well as in the non-perturbative sector of the standard model, the third can be realized in an expanding universe where as we have seen interactions come in and out of equilibrium.

There are several interesting and viable mechanisms for the production of the baryon asymmetry. While, we can not review any of them here in any detail, we mention some of the important scenarios. In all cases, all three ingredients listed above are incorporated. One of the first mechanisms was based on the out of equilibrium decay of a massive particle such as a superheavy GUT gauge of Higgs boson [42,43]. A novel mechanism involving the decay of flat directions in supersymmetric models is known as the Affleck-Dine scenario [44]. Recently, much attention has been focused on the possibility of generating the baryon asymmetry at the electro-weak scale using the non-perturbative interactions of sphalerons [45]. Because these interactions conserve the sum of baryon and lepton number, $B + L$, it is possible to first generate a lepton asymmetry (*e.g.*, by the out-of-equilibrium decay of a superheavy right-handed neutrino), which is converted to a baryon asymmetry at the electro-weak scale [46]. This mechanism is known as lepto-baryogenesis.

19.3.7. *Nucleosynthesis*:

An essential element of the standard cosmological model is Big-Bang nucleosynthesis (BBN), the theory which predicts the abundances of the light element isotopes D, ^{3}He, ^{4}He, and ^{7}Li. Nucleosynthesis takes place at a temperature scale of order 1 MeV. The nuclear processes lead primarily to ^{4}He, with a primordial mass fraction of about 24%. Lesser amounts of the other light elements are produced: about 10^{-5} of D and ^{3}He and about 10^{-10} of ^{7}Li by number relative to H. The abundances of the light elements depend almost solely on one key parameter, the baryon-to-photon ratio, η. The nucleosynthesis predictions can be compared with observational determinations of the abundances of the light elements. Consistency between theory and observations leads to a conservative range of

$$3.4 \times 10^{-10} < \eta < 6.9 \times 10^{-10} . \tag{19.54}$$

η is related to the fraction of Ω contained in baryons, Ω_b

$$\Omega_b = 3.66 \times 10^7 \eta \, h^{-2} , \tag{19.55}$$

or $10^{10}\eta = 274\Omega_b h^2$. The WMAP result [28] for $\Omega_b h^2$ of 0.0224 ± 0.0009 translates into a value of $\eta = 6.15 \pm 0.25$. This result can be used to 'predict' the light element abundance which can in turn be compared with observation [47]. The resulting D/H abundance is in excellent agreement with that found in quasar absorption systems. It is in reasonable agreement with the helium abundance observed in extra-galactic HII regions (once systematic uncertainties are accounted for) but is in poor agreement with the Li abundance observed in the atmospheres of halo dwarf stars. (See the review on BBN—Sec. 20 of this *Review* for a detailed discussion of BBN or references [48,49].)

19.3.8. *The transition to a matter-dominated Universe*:

In the Standard Model, the temperature (or redshift) at which the Universe undergoes a transition from a radiation dominated to a matter dominated Universe is determined by the amount of dark matter. Assuming three nearly massless neutrinos, the energy density in radiation at temperatures $T \ll 1$ MeV, is given by

$$\rho_r = \frac{\pi^2}{30} \left[2 + \frac{21}{4} \left(\frac{4}{11} \right)^{4/3} \right] T^4 . \tag{19.56}$$

In the absence of non-baryonic dark matter, the matter density can be written as

$$\rho_m = m_{\rm N} \eta \, n_\gamma , \tag{19.57}$$

where $m_{\rm N}$ is the nucleon mass. Recalling that $n_\gamma \propto T^3$ [cf. Eq. (19.39)], we can solve for the temperature or redshift at the matter-radiation equality when $\rho_r = \rho_m$,

$$T_{eq} = 0.22\, m_{\rm N}\, \eta \qquad \text{or} \qquad (1+z_{\rm eq}) = 0.22\, \eta \frac{m_{\rm N}}{T_0} \ , \tag{19.58}$$

where T_0 is the present temperature of the microwave background. For $\eta = 5 \times 10^{-10}$, this corresponds to a temperature $T_{\rm eq} \simeq 0.1$ eV or $(1+z_{\rm eq}) \simeq 425$. A transition this late is very problematic for structure formation (see Sec. 19.4.5).

The redshift of matter domination can be pushed back significantly if non-baryonic dark matter is present. If instead of Eq. (19.57), we write

$$\rho_m = \Omega_m \rho_c \left(\frac{T}{T_0}\right)^3 \ , \tag{19.59}$$

we find that

$$T_{eq} = 0.9 \frac{\Omega_m \rho_c}{T_0^3} \qquad \text{or} \qquad (1+z_{eq}) = 2.4 \times 10^4 \Omega_m h^2 \ . \tag{19.60}$$

19.4. The Universe at late times

19.4.1. *The CMB*:

One form of the infamous Olbers' paradox says that, in Euclidean space, surface brightness is independent of distance. Every line of sight will terminate on matter that is hot enough to be ionized and so scatter photons: $T \gtrsim 10^3$ K; the sky should therefore shine as brightly as the surface of the Sun. The reason the night sky is dark is entirely due to the expansion, which cools the radiation temperature to 2.73 K. This gives a Planck function peaking at around 1 mm to produce the microwave background (CMB).

The CMB spectrum is a very accurate match to a Planck function [50]. (See the review on CBR–Sec. 23 of this *Review*.) The COBE estimate of the temperature is [51]

$$T = 2.725 \pm 0.002\,\mathrm{K} \ . \tag{19.61}$$

The lack of any distortion of the Planck spectrum is a strong physical constraint. It is very difficult to account for in any expanding universe other than one that passes through a hot stage. Alternative schemes for generating the radiation, such as thermalization of starlight by dust grains, inevitably generate a superposition of temperatures. What is required in addition to thermal equilibrium is that $T \propto 1/R$, so that radiation from different parts of space appears identical.

Although it is common to speak of the CMB as originating at "recombination," a more accurate terminology is the era of "last scattering," In practice, this takes place at $z \simeq 1100$, almost independently of the main cosmological parameters, at which time the fractional ionization is very small. This occurred when the age of the Universe was a few hundred thousand years. (See the review on CBR–Sec. 23 of this *Review* for a full discussion of the CMB.)

19.4.2. *Matter in the Universe*:

One of the main tasks of cosmology is to measure the density of the Universe, and how this is divided between dark matter and baryons. The baryons consist partly of stars, with $0.002 \lesssim \Omega_* \lesssim 0.003$ [52] but mainly inhabit the IGM. One powerful way in which this can be studied is via the absorption of light from distant luminous objects such as quasars. Even very small amounts of neutral hydrogen can absorb rest-frame UV photons (the Gunn-Peterson effect), and should suppress the continuum by a factor $\exp(-\tau)$, where

$$\tau \simeq \left[\frac{n_{\rm HI}(z)}{(1+z)\sqrt{1+\Omega_m z}}\right] \Big/ \, 10^{-4.62} \, h\,{\rm m}^{-3}, \tag{19.62}$$

and this expression applies while the Universe is matter dominated ($z \gtrsim 1$ in the $\Omega_m = 0.3$ $\Omega_v = 0.7$ model). It is possible that this general absorption has now been seen at $z = 6.2$ [53]. In any case, the dominant effect on the spectrum is a 'forest' of narrow absorption lines, which produce a mean $\tau = 1$ in the Lyα forest at about $z = 3$, and so we have $\Omega_{\rm HI} \simeq 10^{-5.5} h^{-1}$. This is such a small number that clearly the IGM is very highly ionized at these redshifts.

The Lyα forest is of great importance in pinning down the abundance of deuterium. Because electrons in deuterium differ in reduced mass by about 1 part in 4000 compared to Hydrogen, each absorption system in the Lyα forest is accompanied by an offset deuterium line. By careful selection of systems with an optimal HI column density, a measurement of the D/H ratio can be made. This has now been done in 5 quasars, with relatively consistent results [49]. Combining these determinations with the theory of primordial nucleosynthesis yields a baryon density of $\Omega_b h^2 = 0.021 \pm 0.004$ (95% confidence). (See also the review on BBN—Sec. 20 of this *Review*.)

Ionized IGM can also be detected in emission when it is densely clumped, via bremsstrahlung radiation. This generates the spectacular X-ray emission from rich clusters of galaxies. Studies of this phenomenon allow us to achieve an accounting of the total baryonic material in clusters. Within the central $\simeq 1$ Mpc, the masses in stars, X-ray emitting gas and total dark matter can be determined with reasonable accuracy (perhaps 20% rms), and this allows a minimum baryon fraction to be determined [54,55]:

$$\frac{M_{\rm baryons}}{M_{\rm total}} \gtrsim 0.009 + (0.066 \pm 0.003)\, h^{-3/2} \; . \tag{19.63}$$

Because clusters are the largest collapsed structures, it is reasonable to take this as applying to the Universe as a whole. This equation implies a minimum baryon fraction of perhaps 12% (for reasonable h), which is too high for $\Omega_m = 1$ if we take $\Omega_b h^2 \simeq 0.02$ from nucleosynthesis. This is therefore one of the more robust arguments in favor of $\Omega_m \simeq 0.3$. (See the review on Global cosmological parameters—Sec. 21 of this *Review*.) This argument is also consistent with the inference on Ω_m that can be made from Fig. 19.2.

This method is much more robust than the older classical technique for weighing the Universe: '$L \times M/L$'. The overall light density of the Universe is reasonably well determined from redshift surveys of galaxies, so that a good determination of mass M

and luminosity L for a single object suffices to determine Ω_m *if* the mass-to-light ratio is universal.

Galaxy redshift surveys allow us to deduce the galaxy luminosity function, ϕ, which is the comoving number density of galaxies; this may be described by the Schechter function, which is a power law with an exponential cutoff:

$$\phi = \phi^* \left(\frac{L}{L^*}\right)^{-\alpha} e^{-L/L^*} \frac{dL}{L^*} \tag{19.64}$$

The total luminosity density produced by integrating over the distribution is

$$\rho_L = \phi^* L^* \Gamma(2-\alpha) \ , \tag{19.65}$$

and this tells us the average mass-to-light ratio needed to close the Universe. Answers vary (principally owing to uncertainties in ϕ^*). In blue light, the total luminosity density is $\rho_{\rm L} = 2 \pm 0.2 \times 10^8 \ hL_\odot {\rm Mpc}^{-3}$ [56,57]. The critical density is $2.78 \times 10^{11} \ \Omega h^2 M_\odot {\rm Mpc}^{-3}$, so the critical M/L for closure is

$$(M/L)_{\rm crit,\ B} = 1390\, h \pm 10\% \ . \tag{19.66}$$

Dynamical determinations of mass on the largest accessible scales consistently yield blue M/L values of at least $300\, h$, but normally fall short of the closure value [58]. This was a long-standing argument against the $\Omega_m = 1$ model, but it was never conclusive because the stellar populations in objects such as rich clusters (where the masses can be determined) differ systematically from those in other regions.

19.4.3. *Gravitational lensing*:

A robust method for determining masses in cosmology is to use gravitational light deflection. Most systems can be treated as a geometrically thin gravitational lens, where the light bending is assumed to take place only at a single distance. Simple geometry then determines a mapping between the coordinates in the intrinsic source plane and the observed image plane:

$$\alpha(D_{\rm L}\theta_{\rm I}) = \frac{D_{\rm S}}{D_{\rm LS}}(\theta_{\rm I} - \theta_{\rm S}) \ , \tag{19.67}$$

where the angles $\theta_{\rm I}, \theta_{\rm S}$ and α are in general two-dimensional vectors on the sky. The distances $D_{\rm LS}$ etc. are given by an extension of the usual distance-redshift formula:

$$D_{\rm LS} = \frac{R_0 S_k(\chi_{\rm S} - \chi_{\rm L})}{1 + z_{\rm S}} \ . \tag{19.68}$$

This is the angular-diameter distance for objects on the source plane as perceived by an observer on the lens.

Solutions of this equation divide into weak lensing, where the mapping between source plane and image plane is one-to-one, and strong lensing, in which multiple imaging is possible. For circularly-symmetric lenses, an on-axis source is multiply imaged into a 'caustic' ring, whose radius is the Einstein radius:

$$\begin{aligned}\theta_{\rm E} &= \left(4GM\,\frac{D_{\rm LS}}{D_{\rm L}D_{\rm S}}\right)^{1/2} \\ &= \left(\frac{M}{10^{11.09}M_\odot}\right)^{1/2}\left(\frac{D_{\rm L}D_{\rm S}/D_{\rm LS}}{\rm Gpc}\right)^{-1/2} \quad \text{arcsec}\end{aligned} \tag{19.69}$$

The observation of 'arcs' (segments of near-perfect Einstein rings) in rich clusters of galaxies has thus given very accurate masses for the central parts of clusters—generally in good agreement with other indicators, such as analysis of X-ray emission from the cluster IGM [59].

19.4.4. *Density Fluctuations*:

The overall properties of the Universe are very close to being homogeneous; and yet telescopes reveal a wealth of detail on scales varying from single galaxies to large-scale structures of size exceeding 100 Mpc. The existence of these structures must be telling us something important about the initial conditions of the Big Bang, and about the physical processes that have operated subsequently. This motivates the study of the density perturbation field, defined as

$$\delta(\mathbf{x}) \equiv \frac{\rho(\mathbf{x}) - \langle\rho\rangle}{\langle\rho\rangle} . \tag{19.70}$$

A critical feature of the δ field is that it inhabits a universe that is isotropic and homogeneous in its large-scale properties. This suggests that the statistical properties of δ should also be statistically homogeneous—*i.e.*, it is a stationary random process.

It is often convenient to describe δ as a Fourier superposition:

$$\delta(\mathbf{x}) = \sum \delta_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{x}} . \tag{19.71}$$

We avoid difficulties with an infinite universe by applying periodic boundary conditions in a cube of some large volume V. The cross-terms vanish when we compute the variance in the field, which is just a sum over modes of the power spectrum

$$\langle\delta^2\rangle = \sum |\delta_{\mathbf{k}}|^2 \equiv \sum P(k) . \tag{19.72}$$

Note that the statistical nature of the fluctuations must be isotropic, so we write $P(k)$ rather than $P(\mathbf{k})$. The $\langle\ldots\rangle$ average here is a volume average. Cosmological density fields are an example of an ergodic process, in which the average over a large volume tends to the same answer as the average over a statistical ensemble.

The statistical properties of discrete objects sampled from the density field are often described in terms of N-point correlation functions, which represent the excess probability over random for finding one particle in each of N boxes in a given configuration.

For the 2-point case, the correlation function is readily shown to be identical to the autocorrelation function of the δ field: $\xi(r) = \langle\delta(x)\delta(x+r)\rangle$.

The power spectrum and correlation function are Fourier conjugates, and thus are equivalent descriptions of the density field (similarly, k-space equivalents exist for the higher-order correlations). It is convenient to take the limit $V \to \infty$ and use k-space integrals, defining a dimensionless power spectrum as $\Delta^2(k) = d\langle\delta^2\rangle/d\ln k = Vk^3P(k)/2\pi^2$:

$$\xi(r) = \int \Delta^2(k)\,\frac{\sin kr}{kr}\,d\ln k; \quad \Delta^2(k) = \frac{2}{\pi}k^3\int_0^\infty \xi(r)\,\frac{\sin kr}{kr}\,r^2\,dr\ . \tag{19.73}$$

For many years, an adequate approximation to observational data on galaxies was $\xi = (r/r_0)^{-\gamma}$, with $\gamma \simeq 1.8$ and $r_0 \simeq 5\,h^{-1}$ Mpc. Modern surveys are now able to probe into the large-scale linear regime where traces of the curved primordial spectrum can be detected [60].

19.4.5. *Formation of cosmological structure*:

The simplest model for the generation of cosmological structure is gravitational instability acting on some small initial fluctuations (for the origin of which a theory such as inflation is required). If the perturbations are adiabatic (*i.e.*, fractionally perturb number densities of photons and matter equally), the linear growth law for matter perturbations is simple:

$$\delta \propto \begin{cases} a(t)^2 & \text{(radiation domination; } \Omega_{\rm r} = 1) \\ a(t) & \text{(matter domination; } \Omega_{\rm m} = 1) \end{cases} \tag{19.74}$$

For low density universes, the present-day amplitude is suppressed by a factor $g(\Omega)$, where

$$g(\Omega) \simeq \frac{5}{2}\Omega_m\left[\Omega_m^{4/7} - \Omega_v + (1+\Omega_m/2)(1+\frac{1}{70}\Omega_v)\right]^{-1}\ , \tag{19.75}$$

is an accurate fit for models with matter plus cosmological constant. The alternative perturbation mode is isocurvature: only the equation of state changes, and the total density is initially unperturbed. These modes perturb the total entropy density, and thus induce additional large-scale CMB anisotropies [61]. Although the character of perturbations in the simplest inflationary theories are purely adiabatic, correlated adiabatic and isocurvature modes are predicted in many models; the simplest example is the curvaton, which is a scalar field that decays to yield a perturbed radiation density. If the matter content already exists at this time, the overall perturbation field will have a significant isocurvature component. Such a prediction is inconsistent with current CMB data [62], and most analyses of CMB and LSS data assume the adiabatic case to hold exactly.

Linear evolution preserves the shape of the power spectrum. However, a variety of processes mean that growth actually depends on the matter content:

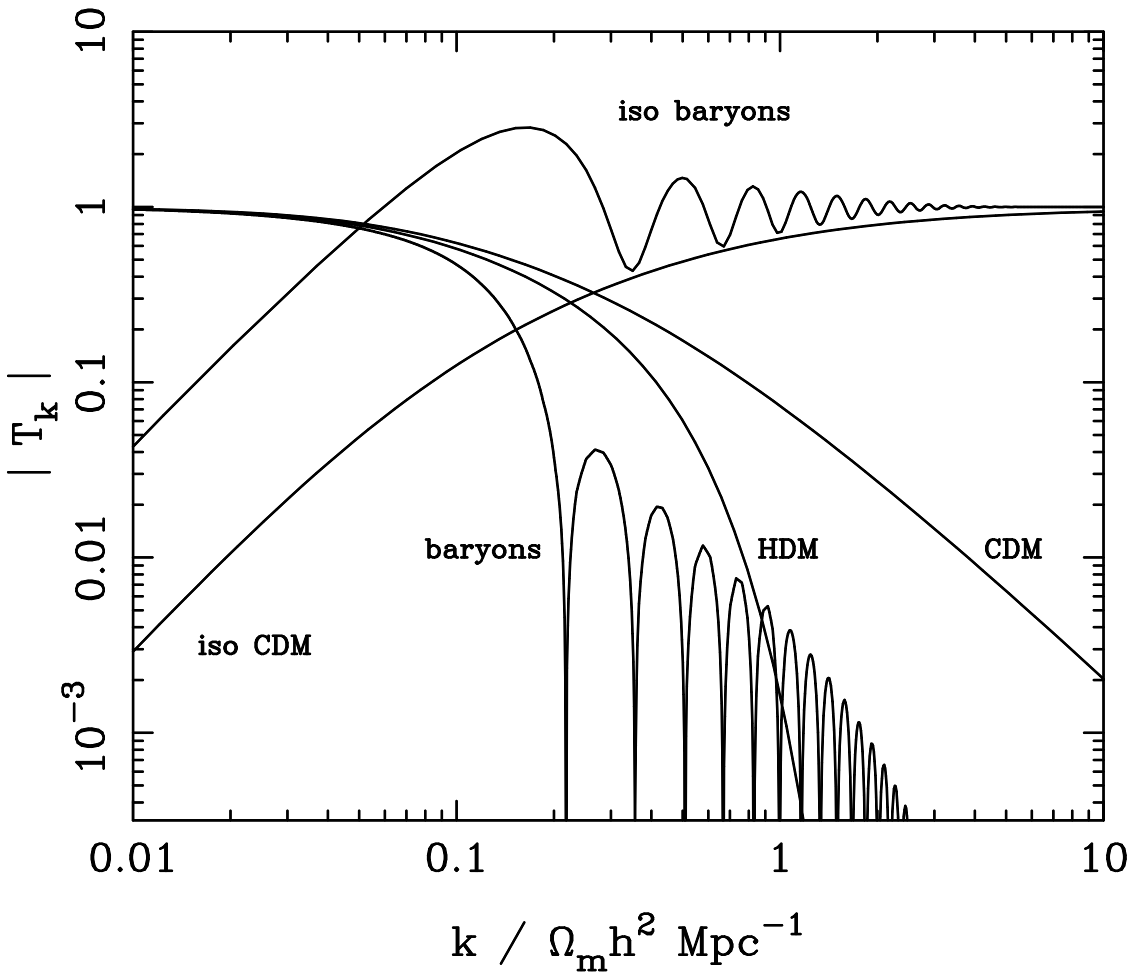

Figure 19.4: A plot of transfer functions for various models. For adiabatic models, $T_k \to 1$ at small k, whereas the opposite is true for isocurvature models. For dark-matter models, the characteristic wavenumber scales proportional to $\Omega_m h^2$. The scaling for baryonic models does not obey this exactly; the plotted cases correspond to $\Omega_m = 1$, $h = 0.5$.

(1) Pressure opposes gravity effectively for wavelengths below the horizon length while the Universe is radiation dominated. The *comoving* horizon size at $z_{\rm eq}$ is therefore an important scale:

$$D_{\rm H}(z_{\rm eq}) = \frac{2(\sqrt{2}-1)}{(\Omega_m z_{\rm eq})^{1/2} H_0} = \frac{16.0}{\Omega_m h^2}\,{\rm Mpc} \tag{19.76}$$

(2) At early times, dark matter particles will undergo free streaming at the speed of light, and so erase all scales up to the horizon—a process that only ceases when the particles go nonrelativistic. For light massive neutrinos, this happens at $z_{\rm eq}$; all structure up to the horizon-scale power-spectrum break is in fact erased. Hot(cold) dark matter models are thus sometimes dubbed large(small)-scale damping models.

(3) A further important scale arises where photon diffusion can erase perturbations in the matter–radiation fluid; this process is named Silk damping.

The overall effect is encapsulated in the transfer function, which gives the ratio of the late-time amplitude of a mode to its initial value (see Fig. 19.4). The overall power spectrum is thus the primordial power-law, times the square of the transfer function:

$$P(k) \propto k^n \, T_k^2 \ . \tag{19.77}$$

The most generic power-law index is $n = 1$: the 'Zeldovich' or 'scale-invariant' spectrum. Inflationary models tend to predict a small 'tilt': $|n-1| \lesssim 0.03$ [13,14]. On the assumption that the dark matter is cold, the power spectrum then depends on 5 parameters: n, h, Ω_b, $\Omega_{\rm cdm}$ ($\equiv \Omega_m - \Omega_b$) and an overall amplitude. The latter is often specified as σ_8, the linear-theory fractional rms in density when a spherical filter of radius $8\,h^{-1}$ Mpc is applied in linear theory. This scale can be probed directly via weak gravitational lensing, and also via its effect on the abundance of rich galaxy clusters. The favored value is approximately [63,64]

$$\sigma_8 = (0.7 \pm 15\%)\,(\Omega_m/0.3)^{-0.5} \ . \tag{19.78}$$

A direct measure of mass inhomogeneity is valuable, since the galaxies inevitably are biased with respect to the mass. This means that the fractional fluctuations in galaxy number, $\delta n/n$ may differ from the mass fluctuations, $\delta\rho/\rho$. It is commonly assumed that the two fields obey some proportionality on large scales where the fluctuations are small, $\delta n/n = b\delta\rho/\rho$, but even this is not guaranteed [65].

The main shape of the transfer function is a break around the horizon scale at $z_{\rm eq}$, which depends just on $\Omega_m h$ when wavenumbers are measured in observable units ($h\,{\rm Mpc}^{-1}$). In principle, accurate data over a wide range of k could determine both Ωh and n, but in practice there is a strong degeneracy between these. For reasonable baryon content, weak oscillations in the transfer function may be visible, giving an alternative means of fixing the baryon content. Current data [60] favor $\Omega_m h \simeq 0.20$ and a baryon fraction of about 0.15 for $n = 1$ (see Fig. 19.5). In order to constrain n itself, it is necessary to examine data on anisotropies in the CMB.

19.4.6. *CMB anisotropies*:

The CMB has a clear dipole anisotropy, of magnitude 1.23×10^{-3}. This is interpreted as being due to the Earth's motion, which is equivalent to a peculiar velocity for the Milky Way of

$$v_{\rm MW} \simeq 600\,{\rm km\,s^{-1}} \quad \text{towards} \quad (\ell, b) \simeq (270^\circ, 30^\circ) \ . \tag{19.79}$$

All higher-order multipole moments of the CMB are however much smaller (of order 10^{-5}), and interpreted as signatures of density fluctuations at last scattering ($\simeq 1100$). To analyze these, the sky is expanded in spherical harmonics as explained in the review on CBR–Sec. 23 of this *Review*. The dimensionless power per ln k or 'bandpower' for the CMB is defined as

$$\mathcal{T}^2(\ell) = \frac{\ell(\ell+1)}{2\pi}\, C_\ell \ . \tag{19.80}$$

This function encodes information from the three distinct mechanisms that cause CMB anisotropies:

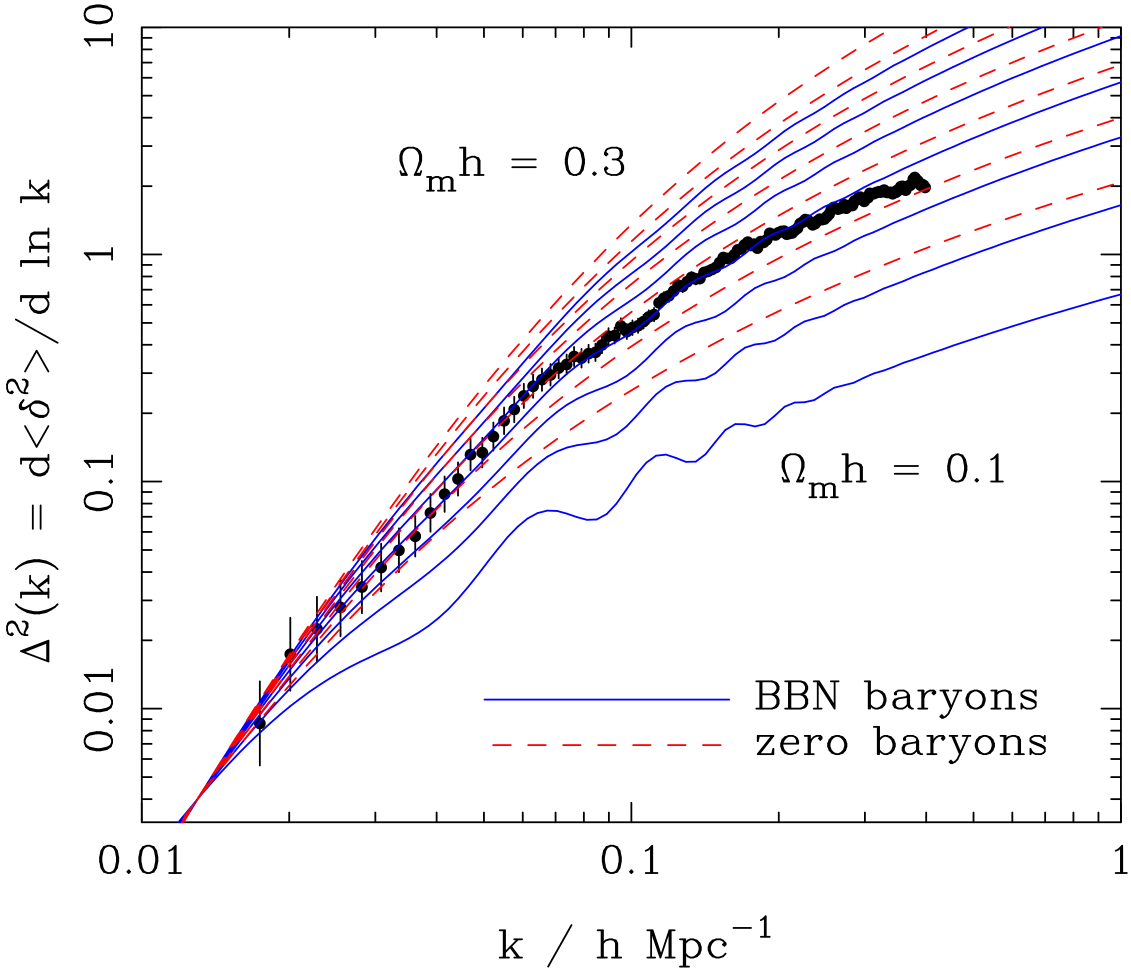

Figure 19.5: The galaxy power spectrum from the 2dFGRS, shown in dimensionless form, $\Delta^2(k) \propto k^3 P(k)$. The solid points with error bars show the power estimate. The window function correlates the results at different k values, and also distorts the large-scale shape of the power spectrum An approximate correction for the latter effect has been applied. The solid and dashed lines show various CDM models, all assuming $n = 1$. For the case with non-negligible baryon content, a big-bang nucleosynthesis value of $\Omega_b h^2 = 0.02$ is assumed, together with $h = 0.7$. A good fit is clearly obtained for $\Omega_m h \simeq 0.2$.

(1) Gravitational (Sachs–Wolfe) perturbations. Photons from high-density regions at last scattering have to climb out of potential wells, and are thus redshifted.

(2) Intrinsic (adiabatic) perturbations. In high-density regions, the coupling of matter and radiation can compress the radiation also, giving a higher temperature.

(3) Velocity (Doppler) perturbations. The plasma has a non-zero velocity at recombination, which leads to Doppler shifts in frequency and hence shifts in brightness temperature.

Because the potential fluctuations obey Poisson's equation, $\nabla^2\Phi = 4\pi G\rho\delta$, and the velocity field satisfies the continuity equation $\nabla \cdot \mathbf{u} = -\dot{\delta}$, the resulting different powers of k ensure that the Sachs-Wolfe effect dominates on large scales and adiabatic effects on small scales.

The relation between angle and comoving distance on the last-scattering sphere requires the comoving angular-diameter distance to the last-scattering sphere; because of its high redshift, this is effectively identical to the horizon size at the present epoch, $D_{\rm H}$:

$$
\begin{aligned}
D_{\rm H} &= \frac{2}{\Omega_m H_0} \quad (\Omega_v = 0) \\
D_{\rm H} &\simeq \frac{2}{\Omega_m^{0.4} H_0} \quad (\text{flat} : \Omega_m + \Omega_v = 1)
\end{aligned}
\tag{19.81}
$$

These relations show how the CMB is strongly sensitive to curvature: the horizon length at last scattering is $\propto 1/\sqrt{\Omega_m}$, so that this subtends an angle that is virtually independent of Ω_m for a flat model. Observations of a peak in the CMB power spectrum at relatively large scales ($\ell \simeq 225$) are thus strongly inconsistent with zero-Λ models with low density: current CMB data require $\Omega_m + \Omega_v \simeq 1 \pm 0.05$. (See *e.g.*, Fig. 19.2).

In addition to curvature, the CMB encodes information about several other key cosmological parameters. Within the compass of simple adiabatic CDM models, there are 9 of these:

$$
\omega_c,\ \omega_b,\ \Omega_t,\ h,\ \tau,\ n_s,\ n_t,\ r,\ Q\ . \tag{19.82}
$$

The symbol ω denotes the physical density, Ωh^2: the transfer function depends only on the densities of CDM (ω_c) and baryons (ω_b). Transcribing the power spectrum at last scattering into an angular power spectrum brings in the total density parameter ($\Omega_t \equiv \Omega_m + \Omega_v = \Omega_c + \Omega_b + \Omega_v$) and h: there is an exact geometrical degeneracy [66] between these that keeps the angular-diameter distance to last scattering invariant, so that models with substantial spatial curvature and large vacuum energy cannot be ruled out without prior knowledge of the Hubble parameter. Alternatively, the CMB alone cannot measure the Hubble parameter.

The other main parameter degeneracy involves the tensor contribution to the CMB anisotropies. These are important at large scales (up to the horizon scales); for smaller scales, only scalar fluctuations (density perturbations) are important. Each of these components is characterized by a spectral index, n, and a ratio between the power spectra of tensors and scalars (r). Finally, the overall amplitude of the spectrum must be specified (Q), together with the optical depth to Compton scattering owing to recent reionization (τ). The tensor degeneracy operates as follows: the main effect of adding a large tensor contribution is to reduce the contrast between low ℓ and the peak at $\ell \simeq 225$ (because the tensor spectrum has no acoustic component). The required height of the peak can be recovered by increasing n_s to increase the small-scale power in the scalar component; this in turn over-predicts the power at $\ell \sim 1000$, but this effect can be counteracted by raising the baryon density [67]. In order to break this degeneracy, additional data are required. For example, an excellent fit to the CMB data is obtained with a scalar-only model with zero curvature and $\omega_b = 0.023$, $\omega_c = 0.134$, $h = 0.73$, $n_s = 0.97$ [28], but this is indistinguishable from a model where tensors dominate at $\ell \lesssim 100$, but we raise ω_b to

0.03 and n_s to 1.2. This baryon density is too high for nucleosynthesis, which disfavors the high-tensor solution [68].

The reason the tensor component is introduced, and why it is so important, is that it is the only non-generic prediction of inflation. Slow-roll models of inflation involve two dimensionless parameters:

$$\begin{aligned} \epsilon &\equiv \frac{M_{\rm P}^2}{16\pi}\,(V'/V)^2 \\ \eta &\equiv \frac{M_{\rm P}^2}{8\pi}\,(V''/V) \end{aligned} \quad , \tag{19.83}$$

where V is the inflaton potential, and dashes denote derivatives with respect to the inflation field. In terms of these, the tensor-to-scalar ratio is $r \simeq 12\epsilon$, and the spectral indices are $n_s = 1 - 6\epsilon + 2\eta$ and $n_t = -2\epsilon$. The natural expectation of inflation is that the quasi-exponential phase ends once the slow-roll parameters become significantly non-zero, so that both $n_s \neq 1$ and a significant tensor component are expected. These prediction can be avoided in some models, but it is undeniable that observation of such features would be a great triumph for inflation. Much future effort in cosmology will therefore be directed towards the question of whether the Universe contains anything other than scale-invariant scalar fluctuations.

References:

1. V.M. Slipher, Pop. Astr. **23**, 21 (1915).
2. K. Lundmark, MNRAS **84**, 747 (1924).
3. E. Hubble and M.L. Humason, Ap. J. **74**, 43 (1931).
4. G. Gamow, Phys. Rev. **70**, 572 (1946).
5. R.A. Alpher, H. Bethe, and G. Gamow, Phys. Rev. **73**, 803 (1948).
6. R.A. Alpher and R.C. Herman, Phys. Rev. **74**, 1737 (1948).
7. R.A. Alpher and R.C. Herman, Phys. Rev. **75**, 1089 (1949).
8. A.A. Penzias and R.W. Wilson, Ap. J. **142**, 419 (1965).
9. S. Weinberg, *Gravitation and Cosmology*, John Wiley and Sons (1972).
10. P.J.E. Peebles, *Principles of Physical Cosmology* Princeton University Press (1993).
11. G. Börner, *The Early Universe: Facts and Fiction*, Springer-Verlag (1988).
12. E.W. Kolb and M.S. Turner, *The Early Universe*, Addison-Wesley (1990).
13. J.A. Peacock, *Cosmological Physics*, Cambridge University Press (1999).
14. A.R. Liddle and D. Lyth, *Cosmological Inflation and Large-Scale Structure*, Cambridge University Press (2000).
15. E.B. Gliner, Sov. Phys. JETP 22, 378 (1966).
16. Y.B. Zeldovich, (1967), Sov. Phys. Uspekhi, **11**, 381 (1968).
17. P.M. Garnavich *et al.*, Ap. J. **507**, 74 (1998).
18. S. Perlmutter *et al.*, Phys. Rev. Lett. **83**, 670 (1999).

19. I. Maor *et al.*, Phys. Rev. **D65**, 123003 (2002).
20. A.G. Riess *et al.*, A. J. **116**, 1009 (1998).
21. S. Perlmutter *et al.*, Ap. J. **517**, 565 (1999).
22. A.G. Riess, PASP, **112**, 1284 (2000).
23. J.L. Tonry *et al.*, `astro-ph/0305008` (2003).
24. W.L. Freedman *et al.*, ApJ **553**, 47 (2001).
25. J.A. Johnson and M. Bolte, ApJ **554**, 888 (2001).
26. R. Jimenez and P. Padoan, ApJ **498**, 704 (1998).
27. E. Carretta *et al.*, ApJ **533**, 215 (2000).
28. D.N. Spergel *et al.*, `astro-ph/0302209` (2003).
29. M. Srednicki, R. Watkins and K. A. Olive, Nucl. Phys. B **310**, 693 (1988).
30. G. Mangano *et al.*, Phys. Lett. **B534**, 8 (2002).
31. A. Linde, Phys. Rev. **D14**, 3345 (1976).
32. A. Linde, Rep. Prog. Phys. **42**, 389 (1979).
33. C.E. Vayonakis, Surveys High Energ. Phys. **5**, 87 (1986).
34. S.A. Bonometto and A. Masiero, Riv. Nuovo Cim. **9N5**, 1 (1986).
35. A. Linde, A.D. Linde, *Particle Physics And Inflationary Cosmology*, Harwood (1990).
36. K.A. Olive, Phys. Rep. **190**, 3345 (1990).
37. D. Lyth and A. Riotto, Phys. Rep. **314**, 1 (1999).
38. A.H. Guth, Phys. Rev. **D23**, 347 (1981).
39. A.D. Linde, Phys. Lett. **108B**, 389 (1982).
40. A. Albrecht and P.J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).
41. A.D. Sakharov, Sov. Phys. JETP Lett. **5**, 24 (1967).
42. S. Weinberg, Phys. Rev. Lett. **42**, 850 (1979).
43. D. Toussaint *et al.*, Phys. Rev. **D19**, 1036 (1979).
44. I. Affleck and M. Dine, Nucl. Phys. **B249**, 361 (1985).
45. V. Kuzmin, V. Rubakov, and M. Shaposhnikov, Phys. Lett. **B155**, 36 (1985).
46. M. Fukugita and T. Yanagida, Phys. Lett. **B174**, 45 (1986).
47. R.H. Cyburt, B.D. Fields, and K.A. Olive, Phys. Lett. **B567**, 227 (2003).
48. K.A. Olive, G. Steigman, and T.P. Walker, Phys. Rept. **333**, 389 (2000).
49. D. Kirkman *et al.*, `astro-ph/0302006` (2003).
50. D.J. Fixsen *et al.*, ApJ **473**, 576 (1996).
51. J.C. Mather *et al.*, ApJ **512**, 511 (1999).
52. S.M. Cole *et al.*, MNRAS **326**, 255 (2001).
53. R.H. Becker *et al.*, Astr. J. **122**, 2850 (2001).
54. S.D.M. White *et al.*, Nature **366**, 429 (1993).

55. S.W. Allen, R.W. Schmidt, A.C. Fabian, MNRAS, **334**, L11 (2002).
56. S. Folkes *et al.*, MNRAS **308**, 459 (1999).
57. M.L. Blanton *et al.*, Astr. J. **121**, 2358 (2001).
58. H. Hoekstra *et al.*, ApJ, **548**, L5 (2001).
59. S.W. Allen, MNRAS **296**, 392 (1998).
60. W.J. Percival *et al.*, MNRAS **327**, 1297 (2001).
61. G. Efstathiou and J.R. Bond, MNRAS **218**, 103 (1986).
62. C. Gordon and A. Lewis, Phys. Rev. **D67**, 123513 (2003).
63. M.L. Brown *et al.*, MNRAS, **341**, 1 (2003).
64. P.T.P. Viana, R.C. Nichol, A.R. Liddle, ApJ, **569**, L75 (2002).
65. A. Dekel and O. Lahav, ApJ **520**, 24 (1999).
66. G. Efstathiou and J.R. Bond, MNRAS, **304**, 75 (1999).
67. G.P. Efstathiou *et al.*, MNRAS **330**, L29 (2002).
68. G.P. Efstathiou, MNRAS **332**, 193 (2002).