### 2.5.5 Evidence for vacuum energy at late times

The idea of inflation is audacious, but undeniably speculative. However, once we accept the idea that quantum fields can generate an equation of state resembling a cosmological constant, we need not confine this mechanism to GUT-scale energies. There is no known mechanism that requires the minimum of $V(\phi)$ to lie exactly at zero energy, so it is quite plausible that there remains in the universe today some non-zero vacuum energy.

The most direct way of detecting vacuum energy has been the immense recent progress in the use of supernovae as standard candles. Type Ia SNe have been used as standard objects for around two decades, with an rms scatter in luminosity of 40%, and so a distance error of 20%. The big breakthrough came when it was realized that the intrinsic timescale of the SNe correlates with luminosity (a brighter SNe lasts longer). Taking out this effect produces corrected standard candles that are capable of measuring distances to about 5% accuracy. Large search campaigns have made it possible to find of the order of 100 SNe over the range $0.1 \lesssim z \lesssim 1$, and two teams have used this strategy to make an empirical estimate of the cosmological distance–redshift relation.

The results of the *Supernova Cosmology Project* (e.g. Perlmutter *et al* 1998) and the *High-z Supernova Search* (e.g. Riess *et al* 1998) are highly consistent. Figure 2.5 shows the Hubble diagram from the latter team. The SNe magnitudes are $K$-corrected, so that their variation with redshift should be a direct measure of luminosity distance as a function of redshift.

We have seen earlier that this is written as the following integral, which must usually be evaluated numerically:

$$D_{\rm L}(z) = (1+z)R_0 S_k(r) = (1+z)\frac{c}{H_0}|1-\Omega|^{-1/2} \times S_k\left[\int_0^z \frac{|1-\Omega|^{1/2}\,{\rm d}z'}{\sqrt{(1-\Omega)(1+z')^2+\Omega_{\rm v}+\Omega_{\rm m}(1+z')^3}}\right],$$

where $\Omega = \Omega_{\rm m} + \Omega_{\rm v}$, and $S_k$ is sinh if $\Omega < 1$, otherwise sin. It is clear from figure 2.5 that the empirical distance–redshift relation is very different from the simplest inflationary prediction, which is the $\Omega = 1$ Einstein–de Sitter model; by redshift 0.6, the SNe are fainter than expected in this model by about 0.5 magnitudes. If this model fails, we can try adjusting $\Omega_{\rm m}$ and $\Omega_{\rm v}$ in an attempt to do better. Comparing each such model to the data yields the likelihood contours shown in figure 2.6, which can be used in the standard way to set confidence limits on the cosmological parameters. The results very clearly require a low-density universe. For $\Lambda = 0$, a very low density is just barely acceptable, with $\Omega_{\rm m} \lesssim 0.1$. However, the discussion of the CMB later shows that such a heavily open model is hard to sustain. The preferred model has $\Omega_{\rm v} \simeq 1$; if we restrict ourselves to the inflationary $k = 0$, then the required parameters are very close to $(\Omega_{\rm m}, \Omega_{\rm v}) = (0.3, 0.7)$.

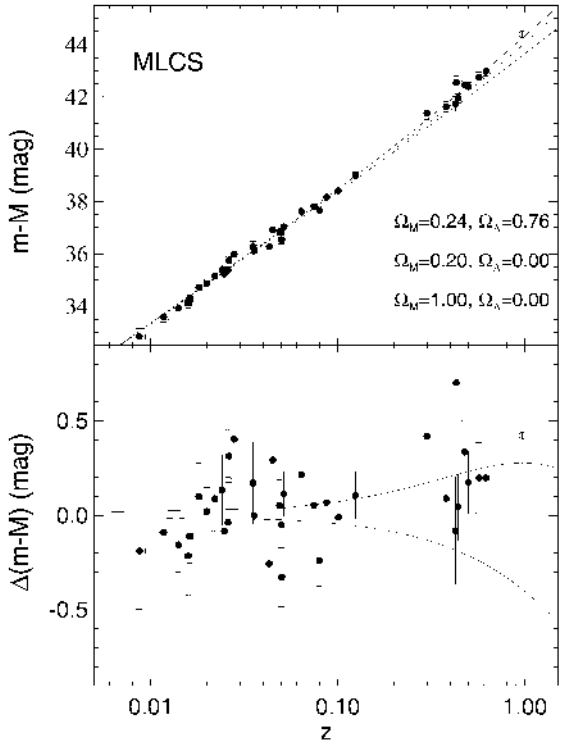


**Figure 2.5.** The Hubble diagram produced by the High-$z$ Supernova search team (Riess *et al* 1998). The lower panel shows the data divided by a default model ($\Omega_{\rm m} = 0.2$, $\Omega_{\rm v} = 0$). The results lie clearly above this model, favouring a non-zero $\Lambda$. The lowest line is the Einstein–de Sitter model, which is in gross disagreement with observation.

### 2.5.6 Cosmic coincidence

This is an astonishing result—an observational detection of the physical reality of vacuum energy. The error bars continue to shrink, and no convincing systematic error has been suggested that could yield this result spuriously; this is one of the most important achievements of 20th century physics.

And yet, accepting the reality of vacuum energy raises a difficult question. If the universe contains a constant vacuum density and normal matter with $\rho \propto a^{-3}$, there is a unique epoch at which these two contributions cross over, and we seem to be living near to that time. This coincidence calls for some explanation. One might think of appealing to anthropic ideas, and these can limit $\Lambda$ to some extent: if the universe became vacuum-dominated at $z > 1000$, gravitational instability as discussed in the next section would have been impossible—so that galaxies, stars and observers would not have been possible. However, Weinberg (1989) argues

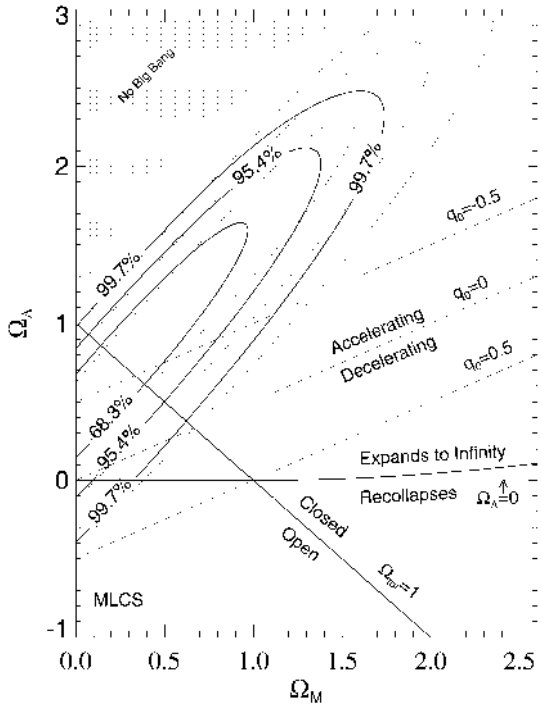


**Figure 2.6.** Confidence contours on the $\Omega_v$–$\Omega_m$ plane, according to Riess *et al* (1998). Open models of all but the lowest densities are apparently ruled out, and non-zero $\Lambda$ is strongly preferred. If we restrict ourselves to $k = 0$, then $\Omega_m \simeq 0.3$ is required. The constraints perpendicular to the $k = 0$ line are not very tight, but CMB data can help here in limiting the allowed degree of curvature.

that $\Lambda$ could have been much larger than its actual value without making observers impossible. Efstathiou (1995) attempted to construct a probability distribution for $\Lambda$ by taking this to be proportional to the number density of galaxies that result in a given model. However, there is no general agreement on how to set a probability measure for this problem.

It would be more satisfactory if we had some physical mechanism that guaranteed the coincidence, and one possibility has been suggested. We already have one coincidence, in that we live relatively close in time to the era of matter–radiation equality ($z \sim 10^3$, as opposed to $z \sim 10^{80}$ for the GUT era). What is required is a cosmological 'constant' that switches on around the equality era. Zlatev *et al* (1999) have suggested how this might happen. The idea is to use the vacuum properties of a homogeneous scalar field as the physical origin of the negative-pressure term detected via SNe. This idea of a 'rolling' $\Lambda$ was first explored by Ratra and Peebles (1988), and there has recently been a tendency

towards use of the fanciful term 'quintessence'. In any case, it is important to appreciate that the idea uses exactly the same physical elements that we discussed in the context of inflation: there is some $V(\phi)$, causing the expectation value of $\phi$ to obey the damped oscillator equation of motion, so the energy density and pressure are

$$\rho_\phi = \dot{\phi}^2/2 + V$$
$$p_\phi = \dot{\phi}^2/2 - V.$$

This gives us two extreme equations of state:

(i) vacuum-dominated, with $V \gg \dot{\phi}^2/2$, so that $p = -\rho$;
(ii) kinetic-dominated, with $V \ll \dot{\phi}^2/2$, so that $p = \rho$.

In the first case, we know that $\rho$ does not alter as the universe expands, so the vacuum rapidly tends to dominate over normal matter. In the second case, the equation of state is the unusual $\Gamma = 2$, so we get the rapid behaviour $\rho \propto a^{-6}$. If a quintessence-dominated universe starts off with a large kinetic term relative to the potential, it may seem that things should always evolve in the direction of being potential-dominated. However, this ignores the detailed dynamics of the situation: for a suitable choice of potential, it is possible to have a *tracker field*, in which the kinetic and potential terms remain in a constant proportion, so that we can have $\rho \propto a^{-\alpha}$, where $\alpha$ can be anything we choose.

Putting this condition in the equation of motion shows that the potential is required to be exponential in form. More importantly, we can generalize to the case where the universe contains scalar field and ordinary matter. Suppose the latter dominates, and obeys $\rho_\mathrm{m} \propto a^{-\alpha}$. It is then possible to have the scalar-field density obeying the same $\rho \propto a^{-\alpha}$ law, provided

$$V(\phi) = \frac{2}{\lambda^2}(6/\alpha - 1)\exp[-\lambda\phi].$$

The scalar-field density is $\rho_\phi = (\alpha/\lambda^2)\rho_\mathrm{total}$ (see, e.g., Liddle and Scherrer 1999). The impressive thing about this solution is that the quintessence density stays a fixed fraction of the total, whatever the overall equation of state: it automatically scales as $a^{-4}$ at early times, switching to $a^{-3}$ after the matter–radiation equality.

This is not quite what we need, but it shows how the effect of the overall equation of state can affect the rolling field. Because of the $3H\dot{\phi}$ term in the equation of motion, $\phi$ 'knows' whether or not the universe is matter dominated. This suggests that a more complicated potential than the exponential may allow the arrival of matter domination to trigger the desired $\Lambda$-like behaviour. Zlatev *et al* suggest two potentials which might achieve this:

$$V(\phi) = M^{4+\beta}\phi^{-\beta} \qquad \text{or} \qquad V(\phi) = M^4[\exp(m_\mathrm{P}/\phi) - 1].$$

The evolution in these potentials may be described by $w(t)$, where $w = p/\rho$. We need $w \simeq 1/3$ in the radiation era, changing to $w \simeq -1$ today. The evolution

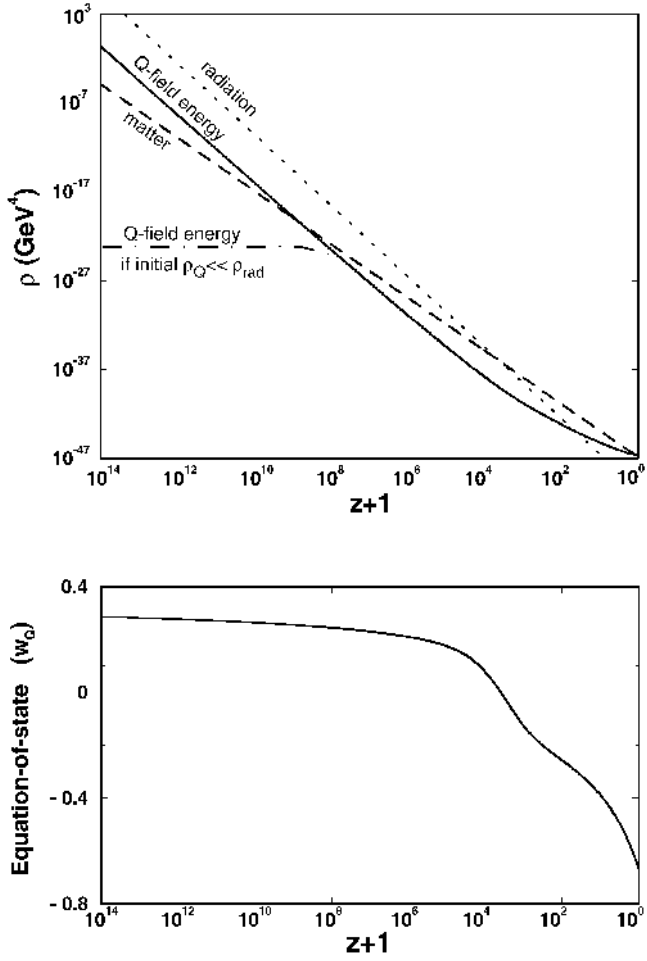


**Figure 2.7.** This figure, taken from Zlatev *et al* (1999), shows the evolution of the density in the 'quintessence' field (top panel), together with the effective equation of state of the quintessence vacuum (bottom panel), for the case of the inverse exponential potential. This allows vacuum energy to lurk at a few per cent of the total throughout the radiation era, but switching on a cosmological constant after the universe becomes matter dominated.

in the inverse exponential potential is shown in figure 2.7, demonstrating that the required behaviour can be found. However, a slight fine-tuning is still required, in that the trick only works for $M \sim 1$ meV, so there has to be an energy coincidence with the energy scale of matter–radiation equality.

So, the idea of tracker fields does not remove completely the puzzle concerning the level of the present-day vacuum energy. In a sense, relegating the

solution to a potential of unexplained form may seem a retrograde step. However, it is at least a testable step: the prediction of figure 2.7 is that $w \simeq -0.8$ today, so that the quintessence density scales as $\rho \propto a^{-0.6}$. This is a significant difference from the classical $w = -1$ vacuum energy, and it should be detectable as the SNe data improve. The existing data already require approximately $w < -0.5$, so there is the entrancing prospect that the equation of state for the vacuum will soon become the subject of experimental study.

## 2.6 Dynamics of structure formation

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. This section summarizes some of the key results concerning how such structure can arise via gravitational instability.

### 2.6.1 Linear perturbations

The study of cosmological perturbations can be presented as a complicated exercise in linearized general relativity; fortunately, much of the essential physics can be extracted from a Newtonian approach. We start by writing down the fundamental equations governing fluid motion (non-relativistic for now):

$$\begin{aligned} &\text{Euler:} && \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = -\frac{\boldsymbol{\nabla} p}{\rho} - \boldsymbol{\nabla}\Phi \\ &\text{energy:} && \frac{\mathrm{D}\rho}{\mathrm{D}t} = -\rho \boldsymbol{\nabla}\cdot\boldsymbol{v} \\ &\text{Poisson:} && \nabla^2\Phi = 4\pi G\rho, \end{aligned}$$

where $\mathrm{D}/\mathrm{D}t = \partial/\partial t + \boldsymbol{v}\cdot\boldsymbol{\nabla}$ is the usual convective derivative. We now produce the *linearized equations of motion* by collecting terms of first order in perturbations about a homogeneous background: $\rho = \rho_0 + \delta\rho$ etc. As an example, consider the energy equation:

$$[\partial/\partial t + (\boldsymbol{v}_0 + \delta\boldsymbol{v})\cdot\boldsymbol{\nabla}](\rho_0 + \delta\rho) = -(\rho_0 + \delta\rho)\boldsymbol{\nabla}\cdot(\boldsymbol{v}_0 + \delta\boldsymbol{v}).$$

For no perturbation, the zero-order equation is

$$(\partial/\partial t + \boldsymbol{v}_0\cdot\boldsymbol{\nabla})\rho_0 = -\rho_0\boldsymbol{\nabla}\cdot\boldsymbol{v}_0;$$

since $\rho_0$ is homogeneous and $\boldsymbol{v}_0 = H\boldsymbol{x}$ is the Hubble expansion, this just says $\dot{\rho}_0 = -3H\rho_0$. Expanding the full equation and subtracting the zeroth-order equation gives the equation for the perturbation:

$$(\partial/\partial t + \boldsymbol{v}_0\cdot\boldsymbol{\nabla})\delta\rho + \delta\boldsymbol{v}\cdot\boldsymbol{\nabla}(\rho_0 + \delta_\rho) = -(\rho_0 + \delta\rho)\boldsymbol{\nabla}\cdot\delta\boldsymbol{v} - \delta\rho\boldsymbol{\nabla}\cdot\boldsymbol{v}_0.$$

Now, for sufficiently small perturbations, terms containing a product of perturbations such as $\delta\boldsymbol{v}\cdot\nabla\delta_\rho$ must be negligible in comparison with the first-order terms. Remembering that $\rho_0$ is homogeneous leaves the linearized equation

$$[\partial/\partial t + \boldsymbol{v}_0\cdot\nabla]\delta\rho = -\rho_0\nabla\cdot\delta\boldsymbol{v} - \delta\rho\nabla\cdot\boldsymbol{v}_0.$$

It is straightforward to perform the same steps with the other equations; the results look simpler if we define the fractional density perturbation

$$\delta \equiv \frac{\delta\rho}{\rho_0}.$$

As before, when dealing with time derivatives of perturbed quantities, the full convective time derivative $\mathrm{D}/\mathrm{D}t$ can always be replaced by $\mathrm{d}/\mathrm{d}t \equiv \partial/\partial t + \boldsymbol{v}_0\cdot\nabla$, which is the time derivative for an observer comoving with the unperturbed expansion of the universe. We then can write

$$\frac{\mathrm{d}}{\mathrm{d}t}\delta\boldsymbol{v} = -\frac{\nabla\delta p}{\rho_0} - \nabla\delta\Phi - (\delta\boldsymbol{v}\cdot\nabla)\boldsymbol{v}_0$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\delta = -\nabla\cdot\delta\boldsymbol{v}$$
$$\nabla^2\delta\Phi = 4\pi G\rho_0\delta.$$

There is now only one complicated term to be dealt with: $(\delta\boldsymbol{v}\cdot\nabla)\boldsymbol{v}_0$ on the right-hand side of the perturbed Euler equation. This is best attacked by writing it in components:

$$[(\delta\boldsymbol{v}\cdot\nabla)\boldsymbol{v}_0]_j = [\delta v]_i\nabla_i[v_0]_j = H[\delta v]_j,$$

where the last step follows because $\boldsymbol{v}_0 = H\boldsymbol{x}_0 \Rightarrow \nabla_i[v_0]_j = H\delta_{ij}$. This leaves a set of equations of motion that have no explicit dependence on the global expansion speed $v_0$; this is only present implicitly through the use of convective time derivatives $\mathrm{d}/\mathrm{d}t$.

These equations of motion are written in Eulerian coordinates: proper length units are used, and the Hubble expansion is explicitly present through the velocity $\boldsymbol{v}_0$. The alternative approach is to use the comoving coordinates formed by dividing the Eulerian coordinates by the scale factor $a(t)$:

$$\boldsymbol{x}(t) = a(t)\boldsymbol{r}(t)$$
$$\delta\boldsymbol{v}(t) = a(t)\boldsymbol{u}(t).$$

The next step is to translate spatial derivatives into comoving coordinates:

$$\nabla_x = \frac{1}{a}\nabla_{\mathrm{r}}.$$

To keep the notation simple, subscripts on $\nabla$ will normally be omitted hereafter, and spatial derivatives will be with respect to comoving coordinates. The

linearized equations for conservation of momentum and matter as experienced by fundamental observers moving with the Hubble flow then take the following simple forms in comoving units:

$$\dot{\boldsymbol{u}} + 2\frac{\dot{a}}{a}\boldsymbol{u} = \frac{\boldsymbol{g}}{a} - \frac{\boldsymbol{\nabla}\delta p}{\rho_0}$$
$$\dot{\delta} = -\boldsymbol{\nabla}\cdot\boldsymbol{u},$$

where dots stand for $\mathrm{d}/\mathrm{d}t$. The peculiar gravitational acceleration $\boldsymbol{\nabla}\delta\Phi/a$ is denoted by $\boldsymbol{g}$.

Before going on, it is useful to give an alternative derivation of these equations, this time working in comoving length units right from the start. First note that the comoving peculiar velocity $\boldsymbol{u}$ is just the time derivative of the comoving coordinate $\boldsymbol{r}$:

$$\dot{\boldsymbol{x}} = \dot{a}\boldsymbol{r} + a\dot{\boldsymbol{r}} = H\boldsymbol{x} + a\dot{\boldsymbol{r}},$$

where the right-hand side must be equal to the Hubble flow $H\boldsymbol{x}$, plus the peculiar velocity $\delta\boldsymbol{v} = a\boldsymbol{u}$. In this equation, dots stand for exact convective time derivatives—i.e. time derivatives measured by an observer who follows a particle's trajectory—rather than partial time derivatives $\partial/\partial t$. This allows us to apply the continuity equation immediately in comoving coordinates, since this equation is simply a statement that particles are conserved, independent of the coordinates used. The exact equation is

$$\frac{\mathrm{D}}{\mathrm{D}t}\rho_0(1+\delta) = -\rho_0(1+\delta)\boldsymbol{\nabla}\cdot\boldsymbol{u},$$

and this is easy to linearize because the background density $\rho_0$ is independent of time when comoving length units are used. This gives the first-order equation $\dot{\delta} = -\boldsymbol{\nabla}\cdot\boldsymbol{u}$ immediately. The equation of motion follows from writing the Eulerian equation of motion as $\ddot{\boldsymbol{x}} = \boldsymbol{g}_0 + \boldsymbol{g}$, where $\boldsymbol{g} = \boldsymbol{\nabla}\delta\Phi/a$ is the peculiar acceleration defined earlier, and $\boldsymbol{g}_0$ is the acceleration that acts on a particle in a homogeneous universe (neglecting pressure forces, for simplicity). Differentiating $\boldsymbol{x} = a\boldsymbol{r}$ twice gives

$$\ddot{\boldsymbol{x}} = a\dot{\boldsymbol{u}} + 2\dot{a}\boldsymbol{u} + \frac{\ddot{a}}{a}\boldsymbol{x} = \boldsymbol{g}_0 + \boldsymbol{g}.$$

The unperturbed equation corresponds to zero peculiar velocity and zero peculiar acceleration: $(\ddot{a}/a)\boldsymbol{x} = \boldsymbol{g}_0$; subtracting this gives the perturbed equation of motion $\boldsymbol{u} + 2(\dot{a}/a)\boldsymbol{u} = \boldsymbol{g}$, as before. This derivation is rather more direct than the previous route of working in Eulerian space. Also, it emphasizes that the equation of motion is exact, even though it happens to be linear in the perturbed quantities.

After doing all this, we still have three equations in the four variables $\delta$, $\boldsymbol{u}$, $\delta\Phi$ and $\delta p$. The system needs an equation of state in order to be closed; this may

be specified in terms of the sound speed

$$c_s^2 \equiv \frac{\partial p}{\partial \rho}.$$

Now think of a plane-wave disturbance $\delta \propto \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}$, where $\boldsymbol{k}$ is a comoving wavevector; in other words, suppose that the wavelength of a single Fourier mode stretches with the universe. All time dependence is carried by the amplitude of the wave, and so the spatial dependence can be factored out of time derivatives in these equations (which would not be true with a constant comoving wavenumber $k/a$). An equation for the amplitude of $\delta$ can then be obtained by eliminating $\boldsymbol{u}$:

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} = \delta(4\pi G\rho_0 - c_s^2 k^2/a^2).$$

This equation is the one that governs the gravitational amplification of density perturbations.

There is a critical proper wavelength, known as the *Jeans length*, at which we switch from the possibility of exponential growth for long-wavelength modes to standing sound waves at short wavelengths. This critical length is

$$\lambda_{\mathrm{J}} = c_s\sqrt{\frac{\pi}{G\rho}},$$

and clearly delineates the scale at which sound waves can cross an object in about the time needed for gravitational free-fall collapse. When considering perturbations in an expanding background, things are more complex. Qualitatively, we expect to have no growth when the 'driving term' on the right-hand side is negative. However, owing to the expansion, $\lambda_{\mathrm{J}}$ will change with time, and so a given perturbation may switch between periods of growth and stasis.

### 2.6.2 Dynamical effects of radiation

At early enough times, the universe was radiation dominated ($c_s = c/\sqrt{3}$) and the analysis so far does not apply. It is common to resort to general relativity perturbation theory at this point. However, the fields are still weak, and so it is possible to generate the results we need by using special relativity fluid mechanics and Newtonian gravity with a relativistic source term. For simplicity, assume that accelerations due to pressure gradients are negligible in comparison with gravitational accelerations (i.e. restrict the analysis to $\lambda \gg \lambda_{\mathrm{J}}$ from the start). The basic equations are then a simplified Euler equation and the full energy and gravitational equations:

$$\begin{aligned}
&\text{Euler:} && \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = -\nabla\Phi \\
&\text{energy:} && \frac{\mathrm{D}}{\mathrm{D}t}(\rho + p/c^2) = \frac{\partial}{\partial t}(p/c^2) - (\rho + p/c^2)\nabla\cdot\boldsymbol{v} \\
&\text{Poisson:} && \nabla^2\Phi = 4\pi G(\rho + 3p/c^2).
\end{aligned}$$

For total radiation domination, $p = \rho c^2/3$, and it is easy to linearize these equations as before. The main differences come from factors of 2 and 4/3 due to the non-negligible contribution of the pressure. The result is a continuity equation $\nabla \cdot \boldsymbol{u} = -(3/4)\dot{\delta}$, and the evolution equation for $\delta$:

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} = \frac{32\pi}{3} G\rho_0 \delta,$$

so the net result of all the relativistic corrections is a driving term on the right-hand side that is a factor 8/3 higher than in the matter-dominated case.

In both matter- and radiation-dominated universes with $\Omega = 1$, we have $\rho_0 \propto 1/t^2$:

$$\text{matter domination } (a \propto t^{2/3}): \qquad 4\pi G\rho_0 = \frac{2}{3t^2}$$

$$\text{radiation domination } (a \propto t^{1/2}): \qquad 32\pi G\rho_0/3 = \frac{1}{t^2}.$$

Every term in the equation for $\delta$ is thus the product of derivatives of $\delta$ and powers of $t$, and a power-law solution is obviously possible. If we try $\delta \propto t^n$, then the result is $n = 2/3$ or $-1$ for matter domination; for radiation domination, this becomes $n = \pm 1$. For the growing mode, these can be combined rather conveniently using the *conformal time* $\eta \equiv \int \mathrm{d}t/a$:

$$\delta \propto \eta^2.$$

Recall that $\eta$ is proportional to the comoving size of the horizon.

It is also interesting to think about the growth of matter perturbations in universes with non-zero vacuum energy, or even possibly some other exotic background with a peculiar equation of state. The differential equation for $\delta$ is as before, but $a(t)$ is altered. The way to deal with this is to treat a spherical perturbation as a small universe. Consider the Friedmann equation in the form

$$(\dot{a})^2 = \Omega_0^{\mathrm{tot}} H_0^2 a^2 + K,$$

where $K = -kc^2/R_0^2$; this emphasizes that $K$ is a constant of integration. A second constant of integration arises in the expression for time:

$$t = \int_0^a \dot{a}^{-1}\,\mathrm{d}a + C.$$

This lets us argue as before in the case of decaying modes: if a solution to the Friedmann equation is $a(t, K, C)$, then valid density perturbations are

$$\delta \propto \left(\frac{\partial \ln a}{\partial K}\right)_t \qquad \text{or} \qquad \left(\frac{\partial \ln a}{\partial C}\right)_t.$$

Since $\partial(\dot{a}^2)/\partial K = 1$, this gives the growing and decaying modes as

$$\delta \propto \begin{cases} (\dot{a}/a)\int_0^a (\dot{a})^{-3}\,\mathrm{d}a & \text{(growing mode)} \\ (\dot{a}/a) & \text{(decaying mode).} \end{cases}$$

(Heath 1977, see also section 10 of Peebles 1980).

The equation for the growing mode requires numerical integration in general, with $\dot{a}(a)$ given by the Friedmann equation. A very good approximation to the answer is given by Carroll *et al* (1992):

$$\frac{\delta(z=0,\Omega)}{\delta(z=0,\Omega=1)} \simeq \frac{5}{2}\Omega_\mathrm{m}\left[\Omega_\mathrm{m}^{4/7} - \Omega_\mathrm{v} + \left(1+\frac{1}{2}\Omega_\mathrm{m}\right)\left(1+\frac{1}{70}\Omega_\mathrm{v}\right)\right]^{-1}.$$

This fitting formula for the growth suppression in low-density universes is an invaluable practical tool. For flat models with $\Omega_\mathrm{m}+\Omega_\mathrm{v} = 1$, it says that the growth suppression is less marked than for an open universe—approximately $\Omega^{0.23}$ as against $\Omega^{0.65}$ if $\Lambda = 0$. This reflects the more rapid variation of $\Omega_\mathrm{v}$ with redshift; if the cosmological constant is important dynamically, this only became so very recently, and the universe spent more of its history in a nearly Einstein–de Sitter state by comparison with an open universe of the same $\Omega_\mathrm{m}$.

What about the case of collisionless matter in a radiation background? The fluid treatment is not appropriate here, since the two species of particles can interpenetrate. A particularly interesting limit is for perturbations well inside the horizon: the radiation can then be treated as a smooth, unclustered background that affects only the overall expansion rate. This is analogous to the effect of $\Lambda$, but an analytical solution does exist in this case. The perturbation equation is as before

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} = 4\pi G\rho_\mathrm{m}\delta,$$

but now $H^2 = 8\pi G(\rho_\mathrm{m}+\rho_\mathrm{r})/3$. If we change variable to $y \equiv \rho_\mathrm{m}/\rho_\mathrm{r} = a/a_\mathrm{eq}$, and use the Friedmann equation, then the growth equation becomes

$$\delta'' + \frac{2+3y}{2y(1+y)}\delta' - \frac{3}{2y(1+y)}\delta = 0$$

(for $k = 0$, as appropriate for early times). It may be seen by inspection that a growing solution exists with $\delta'' = 0$:

$$\delta \propto y + 2/3.$$

It is also possible to derive the decaying mode. This is simple in the radiation-dominated case ($y \ll 1$): $\delta \propto -\ln y$ is easily seen to be an approximate solution in this limit.

What this says is that, at early times, the dominant energy of radiation drives the universe to expand so fast that the matter has no time to respond, and $\delta$ is frozen at a constant value. At late times, the radiation becomes negligible, and the

growth increases smoothly to the Einstein–de Sitter $\delta \propto a$ behaviour (Mészáros 1974). The overall behaviour is therefore similar to the effects of pressure on a coupled fluid: for scales greater than the horizon, perturbations in matter and radiation can grow together, but this growth ceases once the perturbations enter the horizon. However, the explanations of these two phenomena are completely different.

### 2.6.3 The peculiar velocity field

The foregoing analysis shows that gravitational collapse inevitably generates deviations from the Hubble expansion, which are interesting to study in detail.

Consider first a galaxy that moves with some peculiar velocity in an otherwise uniform universe. Even though there is no peculiar gravitational acceleration acting, its velocity will decrease with time as the galaxy attempts to catch up with successively more distant (and therefore more rapidly receding) neighbours. If the proper peculiar velocity is $v$, then after time $\mathrm{d}t$ the galaxy will have moved a proper distance $x = v\,\mathrm{d}t$ from its original location. Its near neighbours will now be galaxies with recessional velocities $Hx = Hv\,\mathrm{d}t$, relative to which the peculiar velocity will have fallen to $v - Hx$. The equation of motion is therefore just

$$\dot{v} = -Hv = -\frac{\dot{a}}{a}v,$$

with the solution $v \propto a^{-1}$: peculiar velocities of non-relativistic objects suffer redshifting by exactly the same factor as photon momenta. It is often convenient to express the peculiar velocity in terms of its comoving equivalent, $\boldsymbol{v} \equiv a\boldsymbol{u}$, for which the equation of motion becomes $\dot{u} = -2Hu$. Thus, in the absence of peculiar accelerations and pressure forces, comoving peculiar velocities redshift away through the Hubble drag term $2H\boldsymbol{u}$.

If we now include the effects of peculiar acceleration, this simply adds the acceleration $g$ on the right-hand side. This gives the equation of motion

$$\dot{\boldsymbol{u}} + \frac{2\dot{a}}{a}\boldsymbol{u} = -\frac{\boldsymbol{g}}{a},$$

where $\boldsymbol{g} = \nabla\delta\Phi/a$ is the peculiar gravitational acceleration. Pressure terms have been neglected, so $\lambda \gg \lambda_{\mathrm{J}}$. Remember that throughout we are using comoving length units, so that $\nabla_{\mathrm{proper}} = \nabla/a$. This equation is the exact equation of motion for a single galaxy, so that the time derivative is $\mathrm{d}/\mathrm{d}t = \partial/\partial t + \boldsymbol{u}\cdot\nabla$. In linear theory, the second part of the time derivative can be neglected, and the equation then turns into one that describes the evolution of the linear peculiar velocity field at a fixed point in comoving coordinates.

The solutions for the peculiar velocity field can be decomposed into modes either parallel to $\boldsymbol{g}$ or independent of $\boldsymbol{g}$ (these are the homogeneous and inhomogeneous solutions to the equation of motion). The interpretation of these solutions is aided by knowing that the velocity field satisfies the *continuity*

*equation*: $\dot{\rho} = -\nabla \cdot (\rho \boldsymbol{v})$ in proper units, which obviously takes the same form $\dot{\rho} = -\nabla \cdot (\rho \boldsymbol{u})$ if lengths and densities are in comoving units. If we express the density as $\rho = \rho_0(1+\delta)$ (where in comoving units $\rho_0$ is just a number independent of time), the continuity equation takes the form

$$\dot{\delta} = -\nabla \cdot [(1+\delta)\boldsymbol{u}],$$

which becomes just

$$\nabla \cdot \boldsymbol{u} = -\dot{\delta}$$

in linear theory when both $\delta$ and $\boldsymbol{u}$ are small. This states that it is possible to have vorticity modes with $\nabla \cdot \boldsymbol{u} = 0$, for which $\dot{\delta}$ vanishes. We have already seen that $\delta$ either grows or decays as a power of time, so these modes require zero density perturbation, in which case the associated peculiar gravity also vanishes. These vorticity modes are thus the required homogeneous solutions, and they decay as $v = au \propto a^{-1}$, as with the kinematic analysis for a single particle. For any gravitational-instability theory, in which structure forms via the collapse of small perturbations laid down at very early times, it should therefore be a very good approximation to say that the linear velocity field must be curl-free.

For the growing modes, we want to try looking for a solution $\boldsymbol{u} = F(t)\boldsymbol{g}$. Then using continuity plus Gauss's theorem, $\nabla \cdot \boldsymbol{g} = 4\pi G a \rho \delta$, gives us

$$\delta \boldsymbol{v} = \frac{2f(\Omega)}{3H\Omega}\boldsymbol{g},$$

where the function $f(\Omega) \equiv (a/\delta)\,\mathrm{d}\delta/\mathrm{d}a$. A very good approximation to this (Peebles 1980) is $g \simeq \Omega^{0.6}$ (a result that is almost independent of $\Lambda$; Lahav *et al* 1991). Alternatively, we can work in Fourier terms. This is easy, as $\boldsymbol{g}$ and $\boldsymbol{k}$ are parallel, so that $\nabla \cdot \boldsymbol{u} = -\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{u} = -\mathrm{i}ku$. Thus, directly from the continuity equation,

$$\delta \boldsymbol{v}_k = -\frac{\mathrm{i}Hf(\Omega)a}{k}\delta_k \hat{\boldsymbol{k}}.$$

The $1/k$ factor shows clearly that peculiar velocities are much more sensitive probes of large-scale inhomogeneities than are density fluctuations. The existence of large-scale homogeneity in density requires $n > -3$, whereas peculiar velocities will diverge unless $n > -1$ on large scales.

### 2.6.4 Transfer functions

We have seen that power spectra at late times result from modifications of any primordial power by a variety of processes: growth under self-gravitation; the effects of pressure; dissipative processes. We now summarize the two main ways in which the power spectrum that exists at early times may differ from that which emerges at the present, both of which correspond to a reduction of small-scale fluctuations (at least, for adiabatic fluctuations; we shall not consider isocurvature modes here):

(1) Radiation effects. Prior to matter–radiation equality, we have already seen that perturbations inside the horizon are prevented from growing by radiation pressure. Once $z_{\rm eq}$ is reached, if collisionless dark matter dominates, perturbations on all scales can grow. We therefore expect a feature in the transfer function around $k \sim 1/r_{\rm H}(z_{\rm eq})$. In the matter-dominated approximation, we get

$$d_{\rm H} = \frac{2c}{H_0}(\Omega z)^{-1/2} \Rightarrow d_{\rm eq} = 39(\Omega h^2)^{-1}\ {\rm Mpc}.$$

The exact distance–redshift relation is

$$R_0\,{\rm d}r = \frac{c}{H_0}\frac{{\rm d}z}{(1+z)\sqrt{1+\Omega_{\rm m}z+(1+z)^2\Omega_{\rm r}}},$$

from which it follows that the correct answer for the horizon size including radiation is a factor $\sqrt{2}-1$ smaller: $d_{\rm eq} = 16.0(\Omega h^2)^{-1}$ Mpc.

(2) Damping. In addition to having their growth retarded, very small-scale perturbations will be erased entirely, which can happen in one of two ways. For collisionless dark matter, perturbations are erased simply by *free-streaming*: random particle velocities cause blobs to disperse. At early times ($kT > mc^2$), the particles will travel at $c$, and so any perturbation that has entered the horizon will be damped. This process switches off when the particles become non-relativistic; for massive particles, this happens long before $z_{\rm eq}$ (resulting in cold dark matter (CDM)). For massive neutrinos, however, it happens *at* $z_{\rm eq}$: only perturbations on very large scales survive in the case of hot dark matter (HDM). In a purely baryonic universe, the corresponding process is called *Silk damping*: the mean free path of photons due to scattering by the plasma is non-zero, and so radiation can diffuse out of a perturbation, convecting the plasma with it.

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:

$$T_k \equiv \frac{\delta_k(z=0)}{\delta_k(z)D(z)},$$

where $D(z)$ is the linear growth factor between redshift $z$ and the present. The normalization redshift is arbitrary, so long as it refers to a time before any scale of interest has entered the horizon.

It is invaluable in practice to have some accurate analytic formulae that fit the numerical results for transfer functions. We give below results for some common models of particular interest (illustrated in figure 2.8, along with other cases where a fitting formula is impractical). For the models with collisionless dark matter, $\Omega_{\rm B} \ll \Omega$ is assumed, so that all lengths scale with the horizon size at matter–radiation equality, leading to the definition

$$q \equiv \frac{k}{\Omega h^2\ {\rm Mpc}^{-1}}.$$

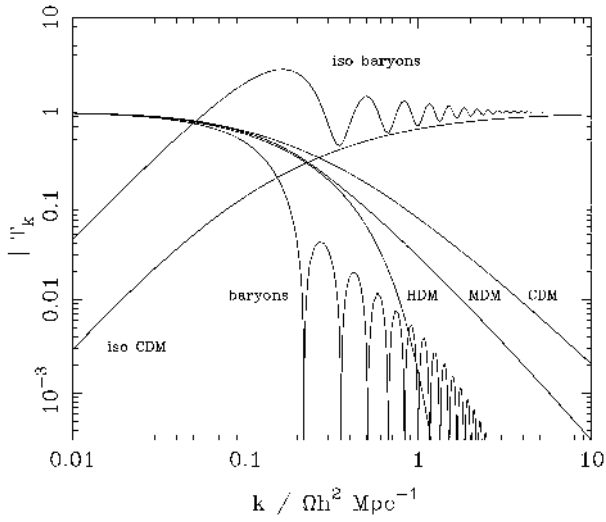


**Figure 2.8.** A plot of transfer functions for various models. For adiabatic models, $T_k \to 1$ at small $k$, whereas the opposite is true for isocurvature models. A number of possible matter contents are illustrated: pure baryons; pure CDM; pure HDM; MDM (30% HDM, 70% CDM). For dark-matter models, the characteristic wavenumber scales proportional to $\Omega h^2$. The scaling for baryonic models does not obey this exactly; the plotted cases correspond to $\Omega = 1$, $h = 0.5$.

We consider the following cases:

(1) adiabatic CDM;
(2) adiabatic massive neutrinos (one massive, two massless); and
(3) isocurvature CDM; these expressions come from Bardeen *et al* (1986; BBKS).

Since the characteristic length-scale in the transfer function depends on the horizon size at matter–radiation equality, the temperature of the CMB enters. In these formulae, it is assumed to be exactly 2.7 K; for other values, the characteristic wavenumbers scale $\propto T^{-2}$. For these purposes massless neutrinos count as radiation, and three species of these contribute a total density that is 0.68 that of the photons.

(1) $T_k = \dfrac{\ln(1+2.34q)}{2.34q}[1+3.89q+(16.1q)^2+(5.46q)^3+(6.71q)^4]^{-1/4}$
(2) $T_k = \exp(-3.9q-2.1q^2)$
(3) $T_k = (5.6q)^2(1+[15.0q+(0.9q)^{3/2}+(5.6q)^2]^{1.24})^{-1/1.24}$.

The case of mixed dark matter (MDM: a mixture of massive neutrinos and CDM) is more complex. See Pogosyan and Starobinksy (1995) for a fit in this case.

These expressions assume pure dark matter, which is unrealistic. At least for CDM models, a non-zero baryonic density lowers the apparent dark-matter density parameter. We can define an apparent shape parameter $\Gamma$ for the transfer function:

$$q \equiv (k/h\ \mathrm{Mpc}^{-1})/\,\Gamma,$$

and $\Gamma = \Omega h$ in a model with zero baryon content. This parameter was originally defined by Efstathiou *et al* (1992), in terms of a CDM model with $\Omega_{\mathrm{B}} = 0.03$. Peacock and Dodds (1994) showed that the effect of increasing $\Omega_{\mathrm{B}}$ was to preserve the CDM-style spectrum shape, but to shift to lower values of $\Gamma$. This shift was generalized to models with $\Omega \neq 1$ by Sugiyama (1995):

$$\Gamma = \Omega h \exp[-\Omega_{\mathrm{B}}(1 + \sqrt{2h}/\Omega)].$$

Note the oscillations in $T(k)$ for high baryon content; these can be significant even in CDM-dominated models when working with high-precision data. Eisenstein and Hu (1998) are to be congratulated for their impressive persistence in finding an accurate fitting formula that describes these wiggles. This is invaluable for carrying out a search of a large parameter space. An interesting question is whether these 'wiggles' survive evolution into the nonlinear regime: Meiksin *et al* (1999) showed that most do not, but that observable signatures of baryons remain on large scales.

### 2.6.5 The spherical model

An overdense sphere is a very useful nonlinear model, as it behaves in exactly the same way as a closed sub-universe. The density perturbation needs not be a uniform sphere: any spherically symmetric perturbation will clearly evolve at a given radius in the same way as a uniform sphere containing the same amount of mass. In what follows, therefore, density refers to the *mean* density inside a given sphere. The equations of motion are the same as for the scale factor, and we can therefore write down the cycloid solution immediately. For a matter-dominated universe, the relation between the proper radius of the sphere and time is

$$r = A(1 - \cos\theta)$$
$$t = B(\theta - \sin\theta),$$

and $A^3 = GMB^2$, just from $\ddot{r} = -GM/r^2$. Expanding these relations up to order $\theta^5$ gives $r(t)$ for small $t$:

$$r \simeq \frac{A}{2}\left(\frac{6t}{B}\right)^{2/3}\left[1 - \frac{1}{20}\left(\frac{6t}{B}\right)^{2/3}\right],$$

and we can identify the density perturbation within the sphere:

$$\delta \simeq \frac{3}{20}\left(\frac{6t}{B}\right)^{2/3}.$$

This all agrees with what we knew already: at early times the sphere expands with the $a \propto t^{2/3}$ Hubble flow and density perturbations grow proportional to $a$.

We can now see how linear theory breaks down as the perturbation evolves. There are three interesting epochs in the final stages of its development, which we can read directly from the above solutions. Here, to keep things simple, we compare only with linear theory for an $\Omega = 1$ background.

(1) *Turnround*. The sphere breaks away from the general expansion and reaches a maximum radius at $\theta = \pi$, $t = \pi B$. At this point, the true density enhancement with respect to the background is just $[A(6t/B)^{2/3}/2]^3/r^3 = 9\pi^2/16 \simeq 5.55$.

(2) *Collapse*. If only gravity operates, then the sphere will collapse to a singularity at $\theta = 2\pi$. This occurs when $\delta_{\rm lin} = (3/20)(12\pi)^{2/3} \simeq 1.69$.

(3) *Virialization*. Consider the time at which the sphere has collapsed by a factor 2 from maximum expansion. At this point, it has kinetic energy $K$ related to potential energy $V$ by $V = -2K$. This is the condition for equilibrium, according to the *virial theorem*. For this reason, many workers take this epoch as indicating the sort of density contrast to be expected as the endpoint of gravitational collapse. This occurs at $\theta = 3\pi/2$, and the corresponding density enhancement is $(9\pi + 6)^2/8 \simeq 147$, with $\delta_{\rm lin} \simeq 1.58$. Some authors prefer to assume that this virialized size is eventually achieved only at collapse, in which case the contrast becomes $(6\pi)^2/2 \simeq 178$.

These calculations are the basis for a common 'rule of thumb', whereby one assumes that linear theory applies until $\delta_{\rm lin}$ is equal to some $\delta_{\rm c}$ a little greater than unity, at which point virialization is deemed to have occurred. Although this only applies for $\Omega = 1$, analogous results can be worked out from the full $\delta_{\rm lin}(z, \Omega)$ and $t(z, \Omega)$ relations; $\delta_{\rm lin} \simeq 1$ is a good criterion for collapse for any value of $\Omega$ likely to be of practical relevance. The full density contrast at virialization may be approximated by

$$1 + \delta_{\rm vir} \simeq 178\Omega^{-0.7}$$

(although flat $\Lambda$-dominated models show less dependence on $\Omega$; Eke *et al* 1996).

## 2.7 Quantifying large-scale structure

The next step is to see how these theoretical ideas can be confronted with statistical measures of the observed matter distribution, and to summarize what is known about the dimensionless density perturbation field

$$\delta(\boldsymbol{x}) \equiv \frac{\rho(\boldsymbol{x}) - \langle\rho\rangle}{\langle\rho\rangle}.$$

A critical feature of the $\delta$ field is that it inhabits a universe that is isotropic and homogeneous in its large-scale properties. This suggests that the statistical

properties of $\delta$ should also be homogeneous, even though it is a field that describes inhomogeneities.

We will often need to use the $\langle \cdots \rangle$ symbol, that denotes averaging over an ensemble of realizations of the statistical $\delta$ process. In practice, this will usually be equated to the spatial average over a sufficiently large volume. Fields that satisfy this property, whereby

$$\text{volume average} \leftrightarrow \text{ensemble average}$$

are termed *ergodic*.

### 2.7.1 Fourier analysis of density fluctuations

It is often convenient to consider building up a general field by the superposition of many modes. For a flat comoving geometry, the natural tool for achieving this is via Fourier analysis. How do we make a Fourier expansion of the density field in an infinite universe? If the field were periodic within some box of side $L$, then we would just have a sum over wave modes:

$$F(\boldsymbol{x}) = \sum F_{\boldsymbol{k}} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}.$$

The requirement of periodicity restricts the allowed wavenumbers to harmonic!boundary conditions

$$k_x = n\frac{2\pi}{L}, \qquad n = 1, 2\ldots,$$

with similar expressions for $k_y$ and $k_z$. Now, if we let the box become arbitrarily large, then the sum will go over to an integral that incorporates the density of states in $k$-space, exactly as in statistical mechanics. The Fourier relations in $n$ dimensions are thus

$$F(x) = \left(\frac{L}{2\pi}\right)^n \int F_k(k) \exp(-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x})\, \mathrm{d}^n k$$
$$F_k(k) = \left(\frac{1}{L}\right)^n \int F(x) \exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x})\, \mathrm{d}^n x.$$

As an immediate example of the Fourier machinery in action, consider the important quantity

$$\xi(\boldsymbol{r}) \equiv \langle \delta(\boldsymbol{x})\delta(\boldsymbol{x}+\boldsymbol{r})\rangle,$$

which is the autocorrelation function of the density field—usually referred to simply as the *correlation function*. The angle brackets indicate an averaging over the normalization volume $V$. Now express $\delta$ as a sum and note that $\delta$ is real, so that we can replace one of the two $\delta$'s by its complex conjugate, obtaining

$$\xi = \left\langle \sum_{\boldsymbol{k}} \sum_{\boldsymbol{k}'} \delta_{\boldsymbol{k}} \delta^*_{\boldsymbol{k}'} \mathrm{e}^{\mathrm{i}(\boldsymbol{k}'-\boldsymbol{k})\cdot\boldsymbol{x}} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}} \right\rangle.$$

Alternatively, this sum can be obtained without replacing $\langle\delta\delta\rangle$ by $\langle\delta\delta^*\rangle$, from the relation between modes with opposite wavevectors that holds for any real field: $\delta_{\boldsymbol{k}}(-\boldsymbol{k}) = \delta^*_{\boldsymbol{k}}(\boldsymbol{k})$. Now, by the periodic boundary conditions, all the cross terms with $\boldsymbol{k}' \neq \boldsymbol{k}$ average to zero. Expressing the remaining sum as an integral, we have

$$\xi(\boldsymbol{r}) = \frac{V}{(2\pi)^3}\int |\delta_{\boldsymbol{k}}|^2 \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}\,\mathrm{d}^3k.$$

In short, the correlation function is the Fourier transform of the *power spectrum*. This relation has been obtained by volume averaging, so it applies to the specific mode amplitudes and correlation function measured in any given realization of the density field. Taking ensemble averages of each side, the relation clearly also holds for the ensemble average power and correlations—which are really the quantities that cosmological studies aim to measure. We shall hereafter often use the alternative notation

$$P(k) \equiv \langle|\delta_k|^2\rangle$$

for the ensemble-average power.

In an isotropic universe, the density perturbation spectrum cannot contain a preferred direction, and so we must have an isotropic power spectrum: $\langle|\delta_{\boldsymbol{k}}|^2(\boldsymbol{k})\rangle = |\delta_k|^2(k)$. The angular part of the $k$-space integral can therefore be performed immediately: introduce spherical polars with the polar axis along $\boldsymbol{k}$, and use the reality of $\xi$ so that $\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \rightarrow \cos(kr\cos\theta)$. In three dimensions, this yields

$$\xi(r) = \frac{V}{(2\pi)^3}\int P(k)\frac{\sin kr}{kr}4\pi k^2\,\mathrm{d}k.$$

We shall usually express the power spectrum in dimensionless form, as the variance per $\ln k$ ($\Delta^2(k) = \mathrm{d}\langle\delta^2\rangle/\mathrm{d}\ln k \propto k^3P[k]$):

$$\Delta^2(k) \equiv \frac{V}{(2\pi)^3}4\pi k^3P(k) = \frac{2}{\pi}k^3\int_0^\infty \xi(r)\frac{\sin kr}{kr}r^2\,\mathrm{d}r.$$

This gives a more easily visualizable meaning to the power spectrum than does the quantity $VP(k)$, which has dimensions of volume: $\Delta^2(k) = 1$ means that there are order-unity density fluctuations from modes in the logarithmic bin around wavenumber $k$. $\Delta^2(k)$ is therefore the natural choice for a Fourier-space counterpart to the dimensionless quantity $\xi(r)$.

This shows that the power spectrum is a central quantity in cosmology, but how can we predict its functional form? For decades, this was thought to be impossible, and so a minimal set of assumptions was investigated. In the absence of a physical theory, we should not assume that the spectrum contains any preferred length scale, otherwise we should then be compelled to explain this feature. Consequently, the spectrum must be a featureless power law:

$$\langle|\delta_k|^2\rangle \propto k^n.$$

The index $n$ governs the balance between large-and small-scale power.

A power-law spectrum implies a power-law correlation function. If $\xi(r) = (r/r_0)^{-\gamma}$, with $\gamma = n + 3$, the corresponding 3D power spectrum is

$$\Delta^2(k) = \frac{2}{\pi}(kr_0)^{\gamma}\Gamma(2-\gamma)\sin\frac{(2-\gamma)\pi}{2} \equiv \beta(kr_0)^{\gamma}$$

($= 0.903(kr_0)^{1.8}$ if $\gamma = 1.8$). This expression is only valid for $n < 0$ ($\gamma < 3$); for larger values of $n$, $\xi$ must become negative at large $r$ (because $P(0)$ must vanish, implying $\int_0^{\infty} \xi(r) r^2 \, \mathrm{d}r = 0$). A cut-off in the spectrum at large $k$ is needed to obtain physically sensible results.

Most important of all is the scale-invariant spectrum, which corresponds to the value $n = 1$, i.e. $\Delta^2 \propto k^4$. To see how the name arises, consider a perturbation $\delta\Phi$ in the gravitational potential:

$$\nabla^2\delta\Phi = 4\pi G\rho_0\delta \Rightarrow \delta\Phi_k = -4\pi G\rho_0\delta_k/k^2.$$

The two powers of $k$ pulled down by $\nabla^2$ mean that, if $\Delta^2 \propto k^4$ for the power spectrum of density fluctuations, then $\Delta^2_{\Phi}$ is a constant. Since potential perturbations govern the flatness of spacetime, this says that the scale-invariant spectrum corresponds to a metric that is a *fractal*: spacetime has the same degree of 'wrinkliness' on each resolution scale.

### 2.7.2 The CDM model

The CDM model is the simplest model for structure formation, and it is worth examining in some detail. The CDM linear-theory spectrum modifications are illustrated in figure 2.9. The primordial power-law spectrum is reduced at large $k$, by an amount that depends on both the quantity of dark matter and its nature. Generally the bend in the spectrum occurs near $1/k$ of the order of the horizon size at matter–radiation equality, proportional to $(\Omega h^2)^{-1}$. For a pure CDM universe, with scale-invariant initial fluctuations ($n = 1$), the observed spectrum depends only on two parameters. One is the shape $\Gamma = \Omega h$, and the other is a normalization. On the shape front, a government health warning is needed, as follows. It has been quite common to take $\Gamma$-based fits to observations as indicating a *measurement* of $\Omega h$, but there are three reasons why this may give incorrect answers:

(1) The dark matter may not be CDM. An admixture of HDM will damp the spectrum more, mimicking a lower CDM density.
(2) Even in a CDM-dominated universe, baryons can have a significant effect, making $\Gamma$ lower than $\Omega h$.
(3) The strongest (and most-ignored) effect is tilt: if $n \neq 1$, then even in a pure CDM universe a $\Gamma$-model fit to the spectrum will give a badly incorrect estimate of the density (the change in $\Omega h$ is roughly $0.3(n-1)$; Peacock and Dodds 1994).

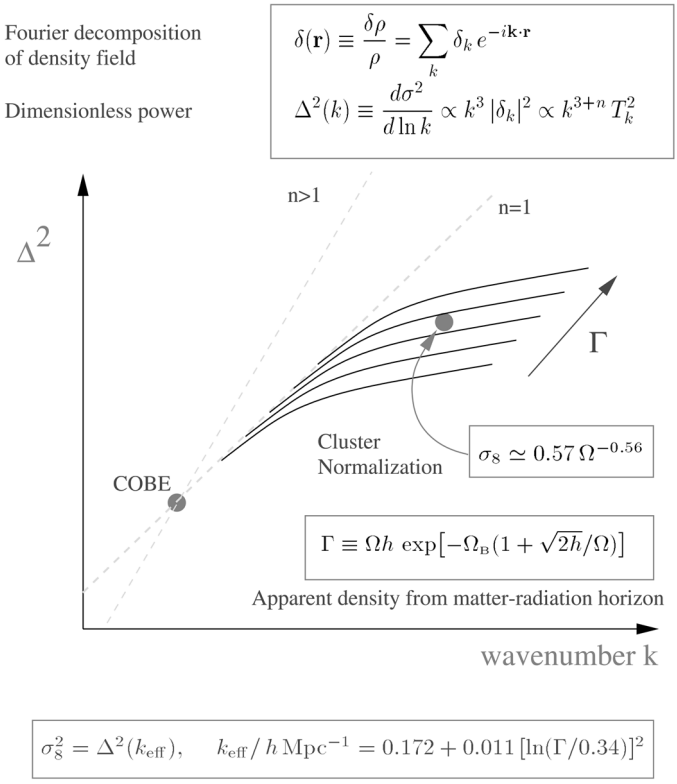


**Figure 2.9.** This figure illustrates how the primordial power spectrum is modified as a function of density in a CDM model. For a given tilt, it is always possible to choose a density that satisfies both the COBE and cluster normalizations.

The other parameter is the normalization. This can be set at a number of points. The COBE normalization comes from large-angle CMB anisotropies, and is sensitive to the power spectrum at $k \simeq 10^{-3}h$ Mpc$^{-1}$. The alternative is to set the normalization near the quasilinear scale, using the abundance of rich clusters. Many authors have tried this calculation, and there is good agreement on the answer:

$$\sigma_8 \simeq (0.5 - 0.6)\Omega_{\rm m}^{-0.6},$$

where $\sigma_8$ is the fractional rms variation in the linear density field, when convolved with a sphere of radium $8h^{-1}$ Mpc (White *et al* 1993, Eke *et al* 1996, Viana and Liddle 1996). In many ways, this is the most sensible normalization to use for LSS studies, since it does not rely on an extrapolation from larger scales.

### 2.7.3 Karhunen–Loève and all that

A key question for these statistical measures is how accurate they are—i.e. how much does the result for a given finite sample depart from the ideal statistic averaged over an infinite universe? Terminology here can be confusing, in that a distinction is sometimes made between *sampling variance* and *cosmic variance*. The former is to be understood as arising from probing a given volume only with a finite number of galaxies (e.g. just the bright ones), so that $\sqrt{N}$ statistics limit our knowledge of the mass distribution within that region. The second term concerns whether we have reached a fair sample of the universe, and depends on whether there is significant power in density perturbation modes with wavelengths larger than the sample depth. Clearly, these two aspects are closely related.

The quantitative analysis of these errors is most simply performed in Fourier space, and was given by Feldman *et al* (1994). The results can be understood most simply by comparison with an idealized complete and uniform survey of a volume $L^3$, with periodicity scale $L$. For an infinite survey, the arbitrariness of the spatial origin means that different modes are uncorrelated:

$$\langle \delta_k(\boldsymbol{k}_i)\delta_k^*(\boldsymbol{k}_j)\rangle = P(k)\delta_{ij}.$$

Each mode has an exponential distribution in power (because the complex coefficients $\delta_k$ are 2D Gaussian-distributed variables on the Argand plane), for which the mean and rms are identical. The fractional uncertainty in the mean power measured over some $k$-space volume is then just determined by the number of uncorrelated modes averaged over

$$\frac{\delta\bar{P}}{\bar{P}} = \frac{1}{N_{\rm modes}^{1/2}}; \qquad N_{\rm modes} = \left(\frac{L}{2\pi}\right)^3 \int \mathrm{d}^3k.$$

The only subtlety is that, because the density field is real, modes at $k$ and $-k$ are perfectly correlated. Thus, if the $k$-space volume is a shell, the effective number of uncorrelated modes is only half this expression.

Analogous results apply for an arbitrary survey selection function. In the continuum limit, the Kroneker delta in the expression for mode correlation would be replaced a term proportional to a delta-function, $\delta[\boldsymbol{k}_i - \boldsymbol{k}_j]$. Now, multiplying the infinite ideal survey by a survey window, $\rho(\boldsymbol{r})$, is equivalent to convolution in the Fourier domain, with the result that the power per mode is correlated over $k$-space separations of order $1/D$, where $D$ is the survey depth.

Given this expression for the fractional power, it is clear that the precision of the estimate can be manipulated by appropriate weighting of the data: giving increased weight to the most distant galaxies increases the effective survey volume, boosting the number of modes. This sounds too good to be true, and of course it is: the previous expression for the fractional power error applies to the sum of true clustering power and shot noise. The latter arises because we transform a point process. Given a set of $N$ galaxies, we would estimate Fourier

coefficients via $\delta_k = (1/N)\sum_i \exp(-\mathrm{i}\boldsymbol{k}\cdot x_i)$. From this, the expectation power is

$$\langle|\delta_k|^2\rangle = P(k) + 1/N.$$

The existence of an additive discreteness correction is no problem, but the *fluctuations* on the shot noise hide the signal of interest. Introducing weights boosts the shot noise, so there is an optimum choice of weight that minimizes the uncertainty in the power after shot-noise subtraction. Feldman *et al* (1994) showed that this weight is

$$w = (1 + \bar{n}P)^{-1},$$

where $\bar{n}$ is the expected galaxy number density as a function of position in the survey.

Since the correlation of modes arises from the survey selection function, it is clear that weighting the data changes the degree of correlation in $k$ space. Increasing the weight in low-density regions increases the effective survey volume, and so shrinks the $k$-space coherence scale. However, the coherence scale continues to shrink as distant regions of the survey are given greater weight, whereas the noise goes through a minimum. There is thus a trade-off between the competing desirable criteria of high $k$-space resolution and low noise. Tegmark (1996) shows how weights may be chosen to implement any given prejudice concerning the relative importance of these two criteria. See also Hamilton (1997a, b) for similar arguments.

Finally, we note that this discussion strictly applies only to the case of Gaussian density fluctuations—which cannot be an accurate model on nonlinear scales. In fact, the errors in the power spectrum are increased on nonlinear scales, and modes at all $k$ have their amplitudes coupled to some degree by nonlinear evolution. These effects are not easy to predict analytically, and are best dealt with by running numerical simulations (see Meiksin and White 1999, Scoccimarro *et al* 1999).

Given these difficulties with correlated results, it is attractive to seek a method where the data can be decomposed into a set of statistics that are completely uncorrelated with each other. Such a method is provided by the Karhunen–Loève formalism. Vogeley and Szalay (1996) argued as follows. Define a column vector of data $\underline{d}$; this can be quite abstract in nature, and could be e.g. the numbers of galaxies in a set of cells, or a set of Fourier components of the transformed galaxy number counts. Similarly, for CMB studies, $\underline{d}$ could be $\delta T/T$ in a set of pixels, or spherical-harmonic coefficients $a_{\ell m}$. We assume that the mean can be identified and subtracted off, so that $\langle\underline{d}\rangle = 0$ in ensemble average. The statistical properties of the data are then described by the covariance matrix

$$C_{ij} \equiv \langle d_i d_j^*\rangle$$

(normally the data will be real, but it is convenient to keep things general and include the complex conjugate).

Suppose we seek to expand the datavector in terms of a set of new orthonormal vectors:

$$\underline{d} = \sum_i a_i \underline{\psi}_i; \quad \underline{\psi}_i^* \cdot \underline{\psi}_j = \delta_{ij}.$$

The expansion coefficients are extracted in the usual way: $a_j = \underline{d} \cdot \underline{\psi}_j^*$. Now require that these coefficients be statistically uncorrelated, $\langle a_i a_j^* \rangle = \lambda_i \delta_{ij}$ (no sum on $i$). This gives

$$\underline{\psi}_i^* \cdot \langle \underline{d}\,\underline{d}^* \rangle \cdot \underline{\psi}_j = \lambda_i \delta_{ij},$$

where the dyadic $\langle \underline{d}\,\underline{d}^* \rangle$ is $\underline{\underline{C}}$, the correlation matrix of the data vector: $(\underline{d}\,\underline{d}^*)_{ij} \equiv d_i d_j^*$. Now, the effect of operating this matrix on one of the $\underline{\psi}_i$ must be expandable in terms of the complete set, which shows that the $\underline{\psi}_j$ must be the eigenvectors of the correlation matrix:

$$\langle \underline{d}\,\underline{d}^* \rangle \cdot \underline{\psi}_j = \lambda_j \underline{\psi}_j.$$

Vogeley and Szalay (1996) further show that these uncorrelated modes are optimal for representing the data: if the modes are arranged in order of decreasing $\lambda$, and the series expansion truncated after $n$ terms, the rms truncation error is minimized for this choice of eigenmodes. To prove this, consider the truncation error

$$\underline{\epsilon} = \underline{d} - \sum_{i=1}^{n} a_i \underline{\psi}_i = \sum_{i=n+1}^{\infty} a_i \underline{\psi}_i.$$

The square of this is

$$\langle \epsilon^2 \rangle = \sum_{i=n+1}^{\infty} \langle |a_i|^2 \rangle,$$

where $\langle |a_i|^2 \rangle = \underline{\psi}_i^* \cdot \underline{\underline{C}} \cdot \underline{\psi}_i$, as before. We want to minimize $\langle \epsilon^2 \rangle$ by varying the $\underline{\psi}_i$, but we need to do this in a way that preserves normalization. This is achieved by introducing a Lagrange multiplier, and minimizing

$$\sum \underline{\psi}_i^* \cdot \underline{\underline{C}} \cdot \underline{\psi}_i + \lambda(1 - \underline{\psi}_i^* \cdot \underline{\psi}_i).$$

This is easily solved if we consider the more general problem where $\underline{\psi}_i^*$ and $\underline{\psi}_i$ are independent vectors:

$$\underline{\underline{C}} \cdot \underline{\psi}_i = \lambda \psi_i.$$

In short, the eigenvectors of $\underline{\underline{C}}$ are optimal in a least-squares sense for expanding the data. The process of truncating the expansion is a form of lossy *data compression*, since the size of the data vector can be greatly reduced without significantly affecting the fidelity of the resulting representation of the universe.

The process of diagonalizing the covariance matrix of a set of data also goes by the more familiar name of *principal components analysis* (PCA), so what is the difference between the KL approach and PCA? In the previous discussion, they

are identical, but the idea of choosing an optimal eigenbasis is more general than PCA. Consider the case where the covariance matrix can be decomposed into a 'signal' and a 'noise' term:

$$\underline{\underline{C}} = \underline{\underline{S}} + \underline{\underline{N}},$$

where $\underline{\underline{S}}$ depends on cosmological parameters that we might wish to estimate, whereas $\underline{\underline{N}}$ is some fixed property of the experiment under consideration. In the simplest imaginable case, $\underline{\underline{N}}$ might be a diagonal matrix, so PCA diagonalizes both $\underline{\underline{S}}$ and $\underline{\underline{N}}$. In this case, ranking the PCA modes by eigenvalue would correspond to ordering the modes according to signal-to-noise ratio. Data compression by truncating the mode expansion then does the sensible thing: it rejects all modes of low signal-to-noise ratio.

However, in general these matrices will not commute, and there will not be a single set of eigenfunctions that are common to the $\underline{\underline{S}}$ and $\underline{\underline{N}}$ matrices. Normally, this would be taken to mean that it is impossible to find a set of coordinates in which both are diagonal. This conclusion can however be evaded, as follows. When considering the effect of coordinate transformations on vectors and matrices, we are normally forced to consider only rotation-like transformations that preserve the norm of a vector (e.g. in quantum mechanics, so that states stay normalized). Thus, we write $\underline{d}' = \underline{\underline{R}} \cdot \underline{d}$, where $\underline{\underline{R}}$ is unitary, so that $\underline{\underline{R}} \cdot \underline{\underline{R}}^\dagger = \underline{\underline{I}}$. If $\underline{\underline{R}}$ is chosen so that its columns are the eigenvalues of $\underline{\underline{N}}$, then the transformed noise matrix, $\underline{\underline{R}} \cdot \underline{\underline{N}} \cdot \underline{\underline{R}}^\dagger$, is diagonal. Nevertheless, if the transformed $\underline{\underline{S}}$ is not diagonal, the two will not commute. This apparently insuperable problem can be solved by using the fact that the data vectors are entirely abstract at this stage. There is therefore no reason not to consider the further transformation of scaling the data, so that $\underline{\underline{N}}$ becomes proportional to the identity matrix. This means that the transformation is no longer unitary – but there is no physical reason to object to a change in the normalization of the data vectors.

Suppose we therefore make a further transformation

$$\underline{d}'' = \underline{\underline{W}} \cdot \underline{d}'.$$

The matrix $\underline{\underline{W}}$ is related to the rotated noise matrix:

$$\underline{\underline{N}}' = \mathrm{diag}(n_1, n_2, \ldots) \Rightarrow \underline{\underline{W}} = \mathrm{diag}(1/\sqrt{n_1}, 1/\sqrt{n_2}, \ldots).$$

This transformation is termed *prewhitening* by Vogeley and Szalay (1996), since it converts the noise matrix to white noise, in which each pixel has a unit noise that is uncorrelated with other pixels. The effect of this transformation on the full covariance matrix is

$$C''_{ij} \equiv \langle d''_i d''^*_j \rangle \Rightarrow \underline{\underline{C}}'' = (\underline{\underline{W}} \cdot \underline{\underline{R}}) \cdot \underline{\underline{C}} \cdot (\underline{\underline{W}} \cdot \underline{\underline{R}})^\dagger.$$

After this transformation, the noise and signal matrices certainly do commute, and the optimal modes for expanding the new data are once again the PCA

eigenmodes in the new coordinates:

$$\underline{\underline{C}}'' \cdot \underline{\psi}_i'' = \lambda \underline{\psi}_i''.$$

These eigenmodes must be expressible in terms of some modes in the original coordinates, $\underline{e}_i$:

$$\underline{\psi}_i'' = (\underline{\underline{W}} \cdot \underline{\underline{R}}) \cdot \underline{e}_i.$$

In these terms, the eigenproblem is

$$(\underline{\underline{W}} \cdot \underline{\underline{R}}) \cdot \underline{\underline{C}} \cdot (\underline{\underline{W}} \cdot \underline{\underline{R}})^\dagger \cdot (\underline{\underline{W}} \cdot \underline{\underline{R}}) \cdot \underline{e}_i = \lambda (\underline{\underline{W}} \cdot \underline{\underline{R}}) \cdot \underline{e}_i.$$

This can be simplified using $\underline{\underline{W}}^\dagger \cdot \underline{\underline{W}} = \underline{\underline{N}}'^{-1}$ and $\underline{\underline{N}}'^{-1} = \underline{\underline{R}} \cdot \underline{\underline{N}}^{-1} \underline{\underline{R}}^\dagger$, to give

$$\underline{\underline{C}} \cdot \underline{\underline{N}}^{-1} \cdot \underline{e}_i = \lambda \underline{e}_i,$$

so the required modes are eigenmodes of $\underline{\underline{C}} \cdot \underline{\underline{N}}^{-1}$. However, care is required when considering the orthonormality of the $\underline{e}_i$: $\underline{\psi}_i^\dagger \cdot \underline{\psi}_j = \underline{e}_i^\dagger \cdot \underline{\underline{N}}^{-1} \cdot \underline{e}_j$, so the $\underline{e}_i$ are not orthonormal. If we write $\underline{d} = \sum_i a_i \underline{e}_i$, then

$$a_i = (\underline{\underline{N}}^{-1} \cdot \underline{e}_i)^\dagger \cdot \underline{d} \equiv \underline{\psi}_i^\dagger \cdot \underline{d}.$$

Thus, the modes used to extract the compressed data by dot product satisfy $\underline{\underline{C}} \cdot \underline{\psi} = \lambda \underline{\underline{N}} \cdot \underline{\psi}$, or finally

$$\underline{\underline{S}} \cdot \underline{\psi} = \lambda \underline{\underline{N}} \cdot \underline{\psi},$$

given a redefinition of $\lambda$. The optimal modes are thus eigenmodes of $\underline{\underline{N}}^{-1} \cdot \underline{\underline{S}}$, hence the name *signal-to-noise eigenmodes* (Bond 1995, Bunn 1995).

It is interesting to appreciate that the set of KL modes just discussed is also the 'best' set of modes to choose from a completely different point of view: they are the modes that are optimal for estimation of a parameter via maximum likelihood. Suppose we write the compressed data vector, $\underline{x}$, in terms of a non-square matrix $\underline{\underline{A}}$ (whose rows are the basis vectors $\underline{\psi}_i^*$):

$$\underline{x} = \underline{\underline{A}} \cdot \underline{d}.$$

The transformed covariance matrix is

$$\underline{\underline{D}} \equiv \langle \underline{x} \underline{x}^\dagger \rangle = \underline{\underline{A}} \cdot \underline{\underline{C}} \cdot \underline{\underline{A}}^\dagger.$$

For the case where the original data obeyed Gaussian statistics, this is true for the compressed data also, so the likelihood is

$$-2 \ln \mathcal{L} = \ln \det \underline{\underline{D}} + \underline{x}^* \cdot \underline{\underline{D}}^{-1} \cdot \underline{x} + \text{constant}.$$

The normal variance on some parameter $p$ (on which the covariance matrix depends) is

$$\frac{1}{\sigma_{\rm p}^2} = \frac{{\rm d}^2[-2 \ln \mathcal{L}]}{{\rm d}q^2}.$$

Without data, we do not know this, so it is common to use the expectation value of the right-hand side as an estimate (recently, there has been a tendency to dub this the 'Fisher matrix').

We desire to optimize $\sigma_p$ by an appropriate choice of data-compression vectors, $\underline{\psi}_j$. By writing $\sigma_p$ in terms of $\underline{\underline{A}}$, $\underline{\underline{C}}$ and $\underline{d}$, it may eventually be shown that the desired optimal modes satisfy

$$\left(\frac{\mathrm{d}}{\mathrm{d}p}\underline{\underline{C}}\right)\cdot\underline{\psi} = \lambda\,\underline{\underline{C}}\cdot\underline{\psi}.$$

For the case where the parameter of interest is the cosmological power, the matrix on the left-hand side is just proportional to $\underline{\underline{S}}$, so we have to solve the eigenproblem

$$\underline{\underline{S}}\cdot\underline{\psi} = \lambda\underline{\underline{C}}\cdot\underline{\psi}.$$

With a redefinition of $\lambda$, this becomes

$$\underline{\underline{S}}\cdot\underline{\psi} = \lambda\underline{\underline{N}}\cdot\underline{\psi}.$$

The optimal modes for parameter estimation in the linear case are thus identical to the PCA modes of the prewhitened data discussed earlier. The more general expression was given by Tegmark *et al* (1997), and it is only in this case, where the covariance matrix is not necessarily linear in the parameter of interest, that the KL method actually differs from PCA.

The reason for going to all this trouble is that the likelihood can now be evaluated much more rapidly, using the compressed data. This allows extensive model searches over large parameter spaces that would be infeasible with the original data (since inversion of an $N \times N$ covariance matrix takes a time proportional to $N^3$). Note, however, that the price paid for this efficiency is that a different set of modes need to be chosen depending on the model of interest, and that these modes will not in general be optimal for expanding the dataset itself. Nevertheless, it may be expected that application of these methods will inevitably grow as datasets increase in size. Present applications mainly prove that the techniques work: see Matsubara *et al* (2000) for application to the LCRS (Las Campanas Redshift Survey), or Padmanabhan *et al* (1999) for the UZC (Updated Zwicky Catalog) survey. The next generation of experiments will probably be forced to resort to data compression of this sort, rather than using it as an elegant alternative method of analysis.

### 2.7.4 Projection on the sky

A more common situation is where we lack any distance data; we then deal with a projection on the sky of a magnitude-limited set of galaxies at different depths. The statistic that is observable is the angular correlation function, $w(\theta)$, or its angular power spectrum $\Delta^2_\theta$. If the sky were flat, the relation between these would

be the usual *Hankel transform* pair:

$$w(\theta) = \int_0^\infty \Delta_\theta^2 J_0(K\theta)\,\mathrm{d}K/K,$$
$$\Delta_\theta^2 = K^2 \int_0^\infty w(\theta) J_0(K\theta)\theta\,\mathrm{d}\theta.$$

For power-law clustering, $w(\theta) = (\theta/\theta_0)^{-\epsilon}$, this gives

$$\Delta_\theta^2(K) = (K\theta_0)^\epsilon 2^{1-\epsilon}\frac{\Gamma(1-\epsilon/2)}{\Gamma(\epsilon/2)},$$

which is equal to $0.77(K\theta_0)^\epsilon$ for $\epsilon = 0.8$. At large angles, these relations are not quite correct. We should really expand the sky distribution in *spherical harmonics*:

$$\delta(\hat{\boldsymbol{q}}) = \sum a_\ell^m Y_{\ell m}(\hat{\boldsymbol{q}}),$$

where $\hat{\boldsymbol{q}}$ is a unit vector that specifies direction on the sky. The functions $Y_{\ell m}$ are the eigenfunctions of the angular part of the $\nabla^2$ operator: $Y_{\ell m}(\theta,\phi) \propto \exp(\mathrm{i}m\phi)P_\ell^m(\cos\theta)$, where $P_\ell^m$ are the *associated Legendre polynomials* (see e.g. section 6.8 of Press *et al* 1992). Since the spherical harmonics satisfy the orthonormality relation $\int Y_{\ell m}Y^*_{\ell' m'}\,\mathrm{d}^2q = \delta_{\ell\ell'}\delta_{mm'}$, the inverse relation is

$$a_\ell^m = \int \delta(\hat{\boldsymbol{q}})Y^*_{\ell m}\,\mathrm{d}^2q.$$

The analogues of the Fourier relations for the correlation function and power spectrum are

$$w(\theta) = \frac{1}{4\pi}\sum_\ell \sum_{m=-\ell}^{m=+\ell} |a_\ell^m|^2 P_\ell(\cos\theta)$$
$$|a_\ell^m|^2 = 2\pi\int_{-1}^{1} w(\theta)P_\ell(\cos\theta)\,\mathrm{d}\cos\theta.$$

For small $\theta$ and large $\ell$, these go over to a form that looks like a flat sky, as follows. Consider the asymptotic forms for the Legendre polynomials and the $J_0$ Bessel function:

$$P_\ell(\cos\theta) \simeq \sqrt{\frac{2}{\pi\ell\sin\theta}}\cos\left[\left(\ell+\frac{1}{2}\right)\theta - \frac{1}{4}\pi\right]$$
$$J_0(z) \simeq \sqrt{\frac{2}{\pi z}}\cos\left[z - \frac{1}{4}\pi\right],$$

for respectively $\ell \to \infty$, $z \to \infty$; see chapters 8 and 9 of Abramowitz and Stegun (1965). This shows that, for $\ell \gg 1$, we can approximate the small-angle

correlation function in the usual way in terms of an angular power spectrum $\Delta^2_\theta$ and angular wavenumber $K$:

$$w(\theta) = \int_0^\infty \Delta^2_\theta(K) J_0(K\theta) \frac{\mathrm{d}K}{K}$$
$$\Delta^2_\theta(K = \ell + \tfrac{1}{2}) = \frac{2\ell+1}{8\pi} \sum_{\mathrm{m}} |a^m_\ell|^2.$$

An important relation is that between the angular and spatial power spectra. In outline, this is derived as follows. The perturbation seen on the sky is

$$\delta(\hat{\boldsymbol{q}}) = \int_0^\infty \delta(\boldsymbol{y}) y^2 \phi(y)\, \mathrm{d}y,$$

where $\phi(y)$ is the *selection function*, normalized such that $\int y^2\phi(y)\,\mathrm{d}y = 1$, and $y$ is comoving distance. The function $\phi$ is the comoving density of objects in the survey, which is given by the integrated luminosity function down to the luminosity limit corresponding to the limiting flux of the survey seen at different redshifts; a flat universe ($\Omega = 1$) is assumed for now. Now write down the Fourier expansion of $\delta$. The plane waves may be related to spherical harmonics via the expansion of a plane wave in *spherical Bessel functions* $j_\ell(x) = (\pi/2x)^{1/2} J_{n+1/2}(x)$ (see chapter 10 of Abramowitz and Stegun (1965) or section 6.7 of Press *et al* (1992)):

$$\mathrm{e}^{\mathrm{i}kr\cos\theta} = \sum_0^\infty (2\ell+1) i^\ell P_\ell(\cos\theta) j_\ell(kr),$$

plus the spherical harmonic addition theorem

$$P_\ell(\cos\theta) = \frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{m=+\ell} Y^*_{\ell m}(\hat{\boldsymbol{q}}) Y_{\ell m}(\hat{\boldsymbol{q}}'); \qquad \hat{\boldsymbol{q}} \cdot \hat{\boldsymbol{q}}' = \cos\theta.$$

These relations allow us to take the angular correlation function $w(\theta) = \langle \delta(\hat{\boldsymbol{q}}) \delta(\hat{\boldsymbol{q}}') \rangle$ and transform it to give the angular power spectrum coefficients. The actual manipulations involved are not as intimidating as they may appear, but they are left as an exercise and we simply quote the final result (Peebles 1973):

$$\langle |a^m_\ell|^2 \rangle = 4\pi \int \Delta^2(k) \frac{\mathrm{d}k}{k} \left[ \int y^2 \phi(y) j_\ell(ky)\, \mathrm{d}y \right]^2.$$

What is the analogue of this formula for small angles? Rather than manipulating large-$\ell$ Bessel functions, it is easier to start again from the correlation function. By writing as before the overdensity observed at a particular

direction on the sky as a radial integral over the spatial overdensity, with a weighting of $y^2\phi(y)$, we see that the angular correlation function is

$$\langle\delta(\hat{\boldsymbol{q}}_1)\delta(\hat{\boldsymbol{q}}_2)\rangle = \int\!\!\int \langle\delta(\boldsymbol{y}_1)\delta(\boldsymbol{y}_2)\rangle y_1^2 y_2^2 \phi(y_1)\phi(y_2)\,\mathrm{d}y_1\,\mathrm{d}y_2.$$

We now change variables to the mean and difference of the radii, $y \equiv (y_1+y_2)/2$; $x \equiv (y_1 - y_2)$. If the depth of the survey is larger than any correlation length, we only get a signal when $y_1 \simeq y_2 \simeq y$. If the selection function is a slowly varying function, so that the thickness of the shell being observed is also of order of the depth, the integration range on $x$ may be taken as being infinite. For small angles, we then obtain *Limber's equation*:

$$w(\theta) = \int_0^\infty y^4\phi^2\,\mathrm{d}y \int_{-\infty}^\infty \xi\left(\sqrt{x^2+y^2\theta^2}\right)\mathrm{d}x$$

(see sections 51 and 56 of Peebles 1980). Theory usually supplies a prediction about the linear density field in the form of the power spectrum, and so it is convenient to recast Limber's equation:

$$w(\theta) = \int_0^\infty y^4\phi^2\,\mathrm{d}y \int_0^\infty \pi\,\Delta^2(k) J_0(ky\theta)\,\mathrm{d}k/k^2.$$

This power-spectrum version of Limber's equation is already in the form required for the relation to the angular power spectrum, and so we obtain the direct small-angle relation between spatial and angular power spectra:

$$\Delta_\theta^2 = \frac{\pi}{K}\int \Delta^2(K/y) y^5\phi^2(y)\,\mathrm{d}y.$$

This is just a convolution in log space, and is considerably simpler to evaluate and interpret than the $w$–$\xi$ version of Limber's equation.

Finally, note that it is not difficult to make allowance for spatial curvature in this discussion. Write the RW metric in the form

$$c^2\,\mathrm{d}\tau^2 = c^2\,\mathrm{d}t^2 - R^2\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\theta^2\right];$$

for $k = 0$, the notation $y = R_0 r$ was used for comoving distance, where $R_0 = (c/H_0)|1-\Omega|^{-1/2}$. The radial increment of comoving distance was $\mathrm{d}x = R_0\,\mathrm{d}r$, and the comoving distance between two objects was $(\mathrm{d}x^2+y^2\theta^2)^{1/2}$. To maintain this version of Pythagoras's theorem, we clearly need to keep the definition of $y$ and redefine radial distance: $\mathrm{d}x = R_0\,\mathrm{d}r\,C(y)$, where $C(y) = [1-k(y/R_0)^2]^{-1/2}$. The factor $C(y)$ appears in the non-Euclidean comoving volume element, $\mathrm{d}V \propto y^2 C(y)\,\mathrm{d}y$, so that we now require the normalization equation for $\phi$ to be

$$\int_0^\infty y^2\phi(y)C(y)\,\mathrm{d}y = 1.$$

The full version of Limber's equation therefore gains two powers of $C(y)$, but one of these is lost in converting between $R_0\,\mathrm{d}r$ and $\mathrm{d}x$:

$$w(\theta) = \int_0^\infty [C(y)]^2 y^4 \phi^2\,\mathrm{d}y \int_{-\infty}^\infty \xi\left(\sqrt{x^2 + y^2\theta^2}\right)\frac{\mathrm{d}x}{C(y)}.$$

The net effect is therefore to replace $\phi^2(y)$ by $C(y)\phi^2(y)$, so that the full power-spectrum equation is

$$\Delta_\theta^2 = \frac{\pi}{K}\int \Delta^2(K/y)C(y)y^5\phi^2(y)\,\mathrm{d}y.$$

It is also straightforward to allow for evolution. The power version of Limber's equation is really just telling us that the angular power from a number of different radial shells adds incoherently, so we just need to use the actual evolved power at that redshift. These integral equations can be inverted numerically to obtain the real-space 3D clustering results from observations of 2D clustering; see Baugh and Efstathiou (1993, 1994).

### 2.7.5 Nonlinear clustering: a problem for CDM?

Observations of galaxy clustering extend into the highly nonlinear regime, $\xi \lesssim 10^4$, so it is essential to understand how this nonlinear clustering relates to the linear-theory initial conditions. A useful trick for dealing with this problem is to think of the density field under full nonlinear evolution as consisting of a set of collapsed, virialized clusters. What is the density profile of one of these objects? At least at separations smaller than the clump separation, the density profile of the clusters is directly related to the correlation function, since this just measures the number density of neighbours of a given galaxy. For a very steep cluster profile, $\rho \propto r^{-\epsilon}$, most galaxies will lie near the centres of clusters, and the correlation function will be a power law, $\xi(r) \propto r^{-\gamma}$, with $\gamma = \epsilon$. In general, because the correlation function is the convolution of the density field with itself, the two slopes differ. In the limit that clusters do not overlap, the relation is $\gamma = 2\epsilon - 3$ (for $3/2 < \epsilon < 3$; see Peebles 1974 or McClelland and Silk 1977). In any case, the critical point is that the correlation function may be be thought of as arising directly from the density profiles of clumps in the density field.

In this picture, it is easy to see how $\xi$ will evolve with redshift, since clusters are virialized objects that do not expand. The hypothesis of *stable clustering* states that, although the separation of clusters will alter as the universe expands, their internal density structure will stay constant with time. This hypothesis clearly breaks down in the outer regions of clusters, where the density contrast is small and linear theory applies, but it should be applicable to small-scale clustering. Regarding $\xi$ as a density profile, its small-scale shape should therefore be fixed in *proper* coordinates, and its amplitude should scale as $(1+z)^{-3}$ owing to the changing mean density of unclustered galaxies, which dilute the clustering at high

redshift. Thus, with $\xi \propto r^{-\gamma}$, we obtain the comoving evolution

$$\xi(r, z) \propto (1+z)^{\gamma-3} \qquad \text{(nonlinear)}.$$

Since the observed $\gamma \simeq 1.8$, this implies slower evolution than is expected in the linear regime:

$$\xi(r, z) \propto (1+z)^{-2} g(\Omega) \qquad \text{(linear)}.$$

This argument does not so far give a relation between the nonlinear slope $\gamma$ and the index $n$ of the linear spectrum. However, the linear and nonlinear regimes match at the scale of quasilinearity, i.e. $\xi(r_0) = 1$; each regime must make the same prediction for how this break point evolves. The linear and nonlinear predictions for the evolution of $r_0$ are, respectively, $r_0 \propto (1+z)^{-2/(n+3)}$ and $r_0 \propto (1+z)^{-(3-\gamma)/\gamma}$, so that $\gamma = (3n+9)/(n+5)$. In terms of an effective index $\gamma = 3 + n_{\rm NL}$, this becomes

$$n_{\rm NL} = -\frac{6}{5+n}.$$

The power spectrum resulting from power-law initial conditions will evolve self-similarly with this index. Note the narrow range predicted: $-2 < n_{\rm NL} < -1$ for $-2 < n < +1$, with an $n = -2$ spectrum having the same shape in both linear and nonlinear regimes.

For many years it was thought that only these limiting cases of extreme linearity or nonlinearity could be dealt with analytically, but in a marvelous piece of alchemy, Hamilton *et al* (1991; HKLM) suggested a general way of understanding the linear $\leftrightarrow$ nonlinear mapping. This initial idea was extended into a workable practical scheme by Peacock and Dodds (1996), allowing the effects of nonlinear evolution to be calculated to a few per cent accuracy for a wide class of spectra.

Indications from the angular clustering of faint galaxies (Efstathiou *et al* 1991) and directly from redshift surveys (Le Fèvre *et al* 1996) are that the observed clustering of galaxies evolves at about the linear-theory rate for $z \lesssim 0.5$, rather more rapidly than the scaling solution would indicate. However, any interpretation of such data needs to assume that galaxies are unbiased tracers of the mass, whereas the observed high amplitude of clustering of quasars at $z \simeq 1$ ($r_0 \simeq 7h^{-1}$ Mpc; see Shanks *et al* 1987, Shanks and Boyle 1994) were an early warning that some high-redshift objects had clustering that is apparently not due to gravity alone. When it eventually became possible to measure correlations of normal galaxies at $z \gtrsim 1$ directly, a similar effect was found, with the comoving strength of clustering being comparable to its value at $z = 0$ (e.g. Adelberger *et al* 1998, Carlberg *et al* 2000). This presumably states that the increasing degree of bias due to high-redshift galaxies being rare objects swamps the gravitational evolution of density fluctuations.

A number of authors have pointed out that the detailed spectral shape inferred from galaxy data appears to be inconsistent with that of nonlinear

evolution from CDM initial conditions. (e.g. Efstathiou *et al* 1990, Klypin *et al* 1996, Peacock 1997). Perhaps the most detailed work was carried out by the Virgo consortium, who carried out $N = 256^3$ simulations of a number of CDM models (Jenkins *et al* 1998). Their results are shown in figure 2.10, which gives the nonlinear power spectrum at various times (cluster normalization is chosen for $z = 0$) and contrasts this with the APM data. The lower small panels are the scale-dependent bias that would be required if the model did, in fact, describe the real universe, defined as

$$b(k) \equiv \left( \frac{\Delta^2_{\rm gals}(k)}{\Delta^2_{\rm mass}} \right)^{1/2}.$$

In all cases, the required bias is non-monotonic; it rises at $k \gtrsim 5h^{-1}$ Mpc, but also displays a bump around $k \simeq 0.1h^{-1}$ Mpc. If real, this feature seems impossible to understand as a genuine feature of the mass power spectrum; certainly, it is not at a scale where the effects of even a large baryon fraction would be expected to act (Eisenstein *et al* 1998, Meiksin *et al* 1999).

### 2.7.6 Real-space and redshift-space clustering

Peculiar velocity fields are responsible for the distortion of the clustering pattern in redshift space, as first clearly articulated by Kaiser (1987). For a survey that subtends a small angle (i.e. in the *distant-observer approximation*), a good approximation to the anisotropic redshift-space Fourier spectrum is given by the Kaiser function together with a damping term from nonlinear effects:

$$\delta^s_k = \delta^r_k (1 + \beta\mu^2) D(k\sigma\mu),$$

where $\beta = \Omega_{\rm m}^{0.6}/b$, $b$ being the linear bias parameter of the galaxies under study, and $\mu = \hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{r}}$. For an exponential distribution of relative small-scale peculiar velocities (as seen empirically), the damping function is $D(y) \simeq (1 + y^2/2)^{-1/2}$, and $\sigma \simeq 400$ km s$^{-1}$ is a reasonable estimate for the pairwise velocity dispersion of galaxies (e.g. Ballinger *et al* 1996).

In principle, this distortion should be a robust way to determine $\Omega$ (or at least $\beta$). In practice, the effect has not been easy to see with past datasets. This is mainly a question of depth: a large survey is needed in order to beat down the shot noise, but this tends to favour bright spectroscopic limits. This limits the result both because relatively few modes in the linear regime are sampled, and also because local survey volumes will tend to violate the small-angle approximation. Strauss and Willick (1995) and Hamilton (1998) review the practical application of redshift-space distortions. In the next section, preliminary results are presented from the 2dF Galaxy Redshift Survey, which shows the distortion effect clearly for the first time.

Peculiar velocities may be dealt with by using the correlation function evaluated explicitly as a 2D function of transverse ($r_\perp$) and radial ($r_\parallel$) separation.

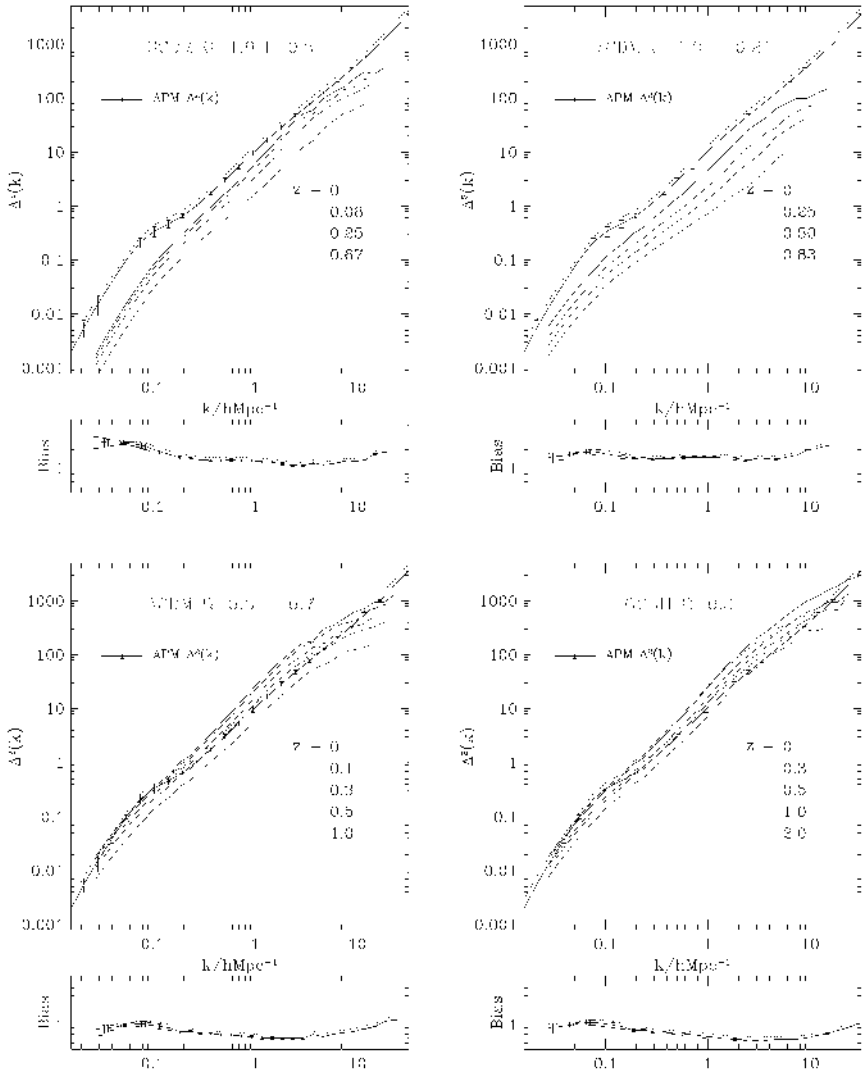


**Figure 2.10.** The nonlinear evolution of various CDM power spectra, as determined by the Virgo consortium (Jenkins *et al* 1998). The broken curves show the evolving spectra for the mass, which at no time match the shape of the APM data. This is expressed in the lower small panels as a scale-dependent bias at $z = 0$: $b^2(k) = P_{\rm APM}/P_{\rm mass}$.

Integrating along the redshift axis then gives the *projected correlation function*, which is independent of the velocities

$$w_{\rm p}(r_\perp) \equiv \int_{-\infty}^{\infty} \xi(r_\perp, r_\parallel)\,{\rm d}r_\parallel = 2\int_{r_\perp}^{\infty} \xi(r)\frac{r\,{\rm d}r}{(r^2 - r_\perp^2)^{1/2}}.$$

In principle, this statistic can be used to recover the real-space correlation function by using the inverse relation for the *Abel integral equation*:

$$\xi(r) = -\frac{1}{\pi}\int_r^\infty w_p'(y)\frac{\mathrm{d}y}{(y^2-r^2)^{1/2}}.$$

An alternative notation for the projected correlation function is $\Xi(r_\perp)$ (Saunders *et al* 1992). Note that the projected correlation function is not dimensionless, but has dimensions of length. The quantity $\Xi(r_\perp)/r_\perp$ is more convenient to use in practice as the projected analogue of $\xi(r)$.

### 2.7.7 The state of the art in LSS

We now consider the confrontation of some of these tools with observations. In the past few years, much attention has been attracted by the estimate of the galaxy power spectrum from the automatic plate measuring (APM) survey (Baugh and Efstathiou 1993, 1994, Maddox *et al* 1996). The APM result was generated from a catalogue of $\sim 10^6$ galaxies derived from UK Schmidt Telescope photographic plates scanned with the Cambridge APM machine; because it is based on a *deprojection* of angular clustering, it is immune to the complicating effects of redshift-space distortions. The difficulty, of course, is in ensuring that any low-level systematics from e.g. spatial variations in magnitude zero point are sufficiently well controlled that they do not mask the cosmological signal, which is of order $w(\theta) \lesssim 0.01$ at separations of a few degrees.

The best evidence that the APM survey has the desired uniformity is the *scaling test*, where the correlations in fainter magnitude slices are expected to move to smaller scales and be reduced in amplitude. If we increase the depth of the survey by some factor $D$, the new angular correlation function will be

$$w'(\theta) = \frac{1}{D}w(D\theta).$$

The APM survey passes this test well; once the overall redshift distribution is known, it is possible to obtain the spatial power spectrum by inverting a convolution integral:

$$w(\theta) = \int_0^\infty y^4\phi^2\,\mathrm{d}y\int_0^\infty \pi\Delta^2(k)J_0(ky\theta)\,\mathrm{d}k/k^2$$

(where zero spatial curvature is assumed). Here, $\phi(y)$ is the comoving density at comoving distance $y$, normalized so that $\int y^2\phi(y)\,\mathrm{d}y = 1$.

This integral was inverted numerically by Baugh and Efstathiou (1993), and gives an impressively accurate determination of the power spectrum. The error estimates are derived empirically from the scatter between independent regions of the sky, and so should be realistic. If there are no undetected systematics, these error bars state that the power is very accurately determined. The APM result

has been investigated in detail by a number of authors (e.g. Gaztañaga and Baugh 1998, Eisenstein and Zaldarriaga 1999) and found to be robust; this has significant implications if true.

Because of the sheer number of galaxies, plus the large volume surveyed, the APM survey outperforms redshift surveys of the past, at least for the purpose of determining the power spectrum. The largest surveys of recent years (CfA: Huchra *et al* 1990, LCRS: Shectman *et al* 1996, PSCz: Saunders *et al* 2000) contain of the order of $10^4$ galaxy redshifts, and their statistical errors are considerably larger than those of the APM. On the other hand, it is of great importance to compare the results of deprojection with clustering measured directly in 3D.

This comparison was carried out by Peacock and Dodds (1994; PD94). The exercise is not straightforward, because the 3D results are affected by redshift-space distortions; also, different galaxy tracers can be biased to different extents. The approach taken was to use each dataset to reconstruct an estimate of the linear spectrum, allowing the relative bias factors to float in order to make these estimates agree as well as possible (figure 2.11). To within a scatter of perhaps a factor 1.5 in power, the results were consistent with a $\Gamma \simeq 0.25$ CDM model. Even though the subsequent sections will discuss some possible disagreements with the CDM models at a higher level of precision, the general existence of CDM-like curvature in the spectrum is likely to be an important clue to the nature of the dark matter.

An important general lesson can be drawn from the lack of large-amplitude features in the power spectrum. This is a strong indication that collisionless matter is deeply implicated in forming large-scale structure. Purely baryonic models contain large bumps in the power spectrum around the Jeans' length prior to recombination ($k \sim 0.03\Omega h^2$ Mpc$^{-1}$), whether the initial conditions are isocurvature or adiabatic. It is hard to see how such features can be reconciled with the data, beyond a 'visibility' in the region of 20%.

The proper resolution of many of the observational questions regarding the large-scale distribution of galaxies requires new generations of redshift survey that push beyond the $N = 10^5$ barrier. Two groups are pursuing this goal. The Sloan survey (e.g. Margon 1999) is using a dedicated 2.5-m telescope to measure redshifts for approximately 700 000 galaxies to $r = 18.2$ in the North Galactic Cap. The 2dF Galaxy Redshift Survey (e.g. Colless 1999) is using a fraction of the time on the 3.9-m Anglo-Australian Telescope plus Two-Degree Field spectrograph to measure 250 000 galaxies from the APM survey to $B_{\rm J} = 19.45$ in the South Galactic Cap. At the time of writing, the Sloan spectroscopic survey has yet to commence. However, the 2dFGRS project has measured in excess of 100 000 redshifts, and some preliminary clustering results are given here. For more details of the survey, particularly the team members whose hard work has made all this possible, see http://www.mso.anu.edu.au/2dFGRS/.

One of the advantages of 2dFGRS is that it is a fully sampled survey, so that the space density out to the depth imposed by the magnitude limit

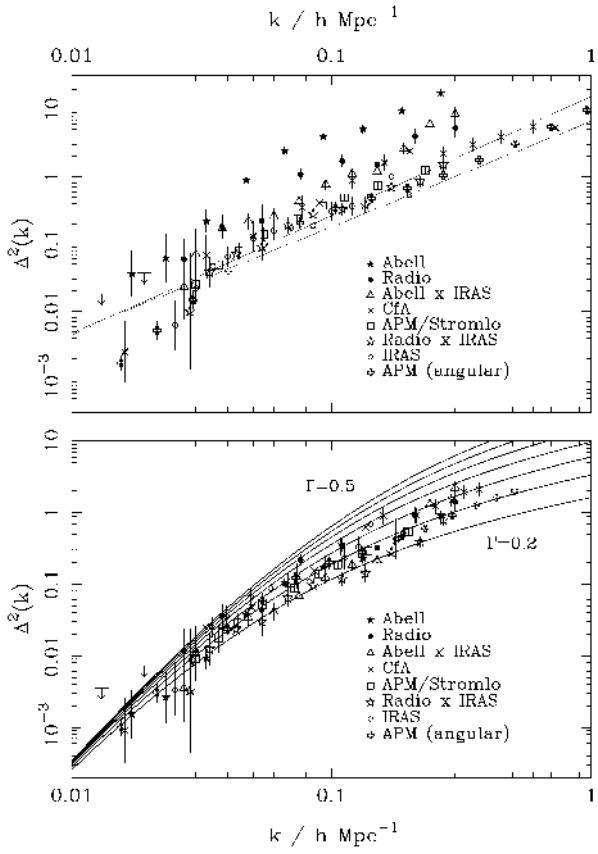


**Figure 2.11.** The PD94 compilation of power-spectrum measurements. The upper panel shows raw power measurements; the lower shows these data corrected for relative bias, nonlinear effects and redshift-space effects.

(median $z = 0.12$) is as high as nature allows: apart from a tail of low surface brightness galaxies (inevitably omitted from any spectroscopic survey), the 2dFGRS measure all the galaxies that exist over a cosmologically representative volume. It is the first to achieve this goal. The fidelity of the resulting map of the galaxy distribution can be seen in figure 2.12, which shows a small subset of the data: a slice of thickness 4 degrees, centred at declination $-27°$.

An issue with using the 2dFGRS data in their current form is that the sky has to be divided into circular 'tiles' each two degrees in diameter ('2dF' = 'two-degree field', within which the AAT is able to measure 400 spectra simultaneously; see http://www.aao.gov.au/2df/ for details of the instrument). The tiles are positioned adaptively, so that larger overlaps occur in regions of high galaxy density. In this way, it is possible to place a fibre on >95% of all galaxies.

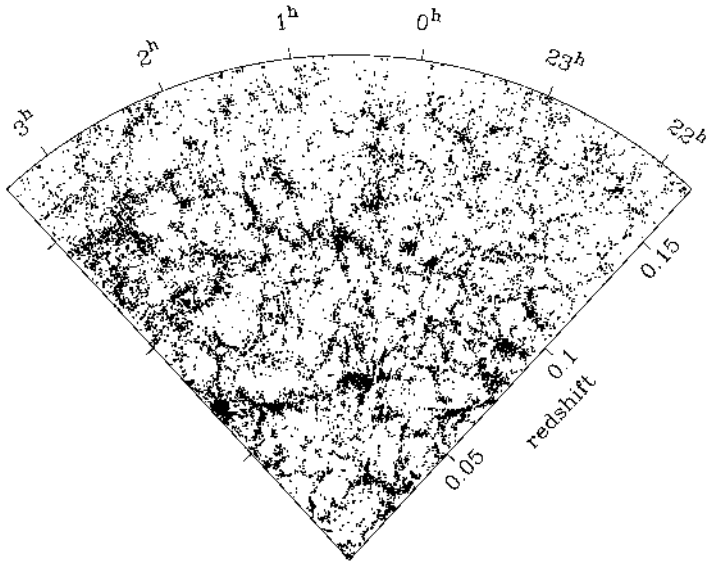


**Figure 2.12.** A four-degree thick slice of the southern strip of the 2dF Galaxy Redshift Survey. This restricted region alone contains 16 419 galaxies.

However, while the survey is in progress, there exist parts of the sky where the overlapping tiles have not yet been observed, and so the effective sampling fraction is only $\simeq$50%. These effects can be allowed for in two different ways. In clustering analyses, we compare the counts of pairs (or $n$-tuplets) of galaxies in the data to the corresponding counts involving an unclustered random catalogue. The effects of variable sampling can therefore be dealt with either by making the density of random points fluctuate according to the sampling, or by weighting observed galaxies by the reciprocal of the sampling factor for the zone in which they lie. The former approach is better from the point of view of shot noise, but the latter may be safer if there is any suspicion that the sampling fluctuations are correlated with real structure on the sky. In practice, both strategies give identical answers for the results below.

At the two-point level, the most direct quantity to compute is the *redshift–space correlation function*. This is an anisotropic function of the orientation of a galaxy pair, owing to peculiar velocities. We therefore evaluate $\xi$ as a function of 2D separation in terms of coordinates both parallel and perpendicular to the line of sight. If the comoving radii of two galaxies are $y_1$ and $y_2$ and their total separation is $r$, then we define coordinates

$$\pi \equiv |y_1 - y_2|; \qquad \sigma = \sqrt{r^2 - \pi^2}.$$

The correlation function measured in these coordinates is shown in figure 2.13. In

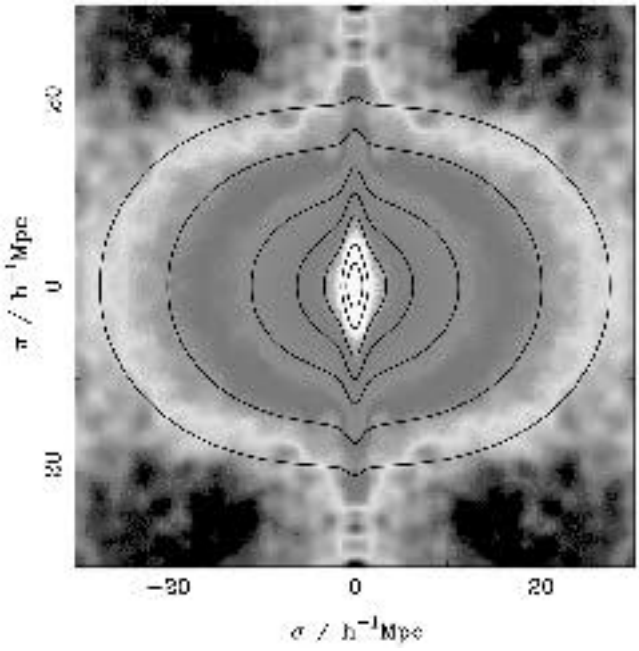


**Figure 2.13.** The redshift–space correlation function for the 2dFGRS, $\xi(\sigma, \pi)$, plotted as a function of transverse ($\sigma$) and radial ($\pi$) pair separation. The function was estimated by counting pairs in boxes of side $0.2h^{-1}$ Mpc, and then smoothing with a Gaussian of rms width $0.5h^{-1}$ Mpc. This plot clearly displays redshift distortions, with 'fingers of God' elongations at small scales and the coherent Kaiser flattening at large radii. The overplotted contours show model predictions with flattening parameter $\beta \equiv \Omega^{0.6}/b = 0.4$ and a pairwise dispersion of $\sigma_{\rm p} = 4h^{-1}$ Mpc. Contours are plotted at $\xi = 10, 5, 2, 1, 0.5, 0.2, 0.1$.

evaluating $\xi(\sigma, \pi)$, the optimal radial weight discussed earlier has been applied, so that the noise at large $r$ should be representative of true cosmic scatter.

The 2dFGRS results for the redshift-space correlation function results are shown in figure 2.13, and display very clearly the two signatures of redshift-space distortions discussed earlier. The *fingers of God* from small-scale random velocities are very clear, as indeed has been the case from the first redshift surveys (e.g. Davis and Peebles 1983). However, this is arguably the first time that the large-scale flattening from coherent infall has been really obvious in the data.

A good way to quantify the flattening is to analyse the clustering as a function of angle into Legendre polynomials:

$$\xi_\ell(r) = \frac{2\ell+1}{2}\int_{-1}^{1} \xi(\sigma = r\sin\theta, \pi = r\cos\theta) P_\ell(\cos\theta)\,{\rm d}\cos\theta.$$

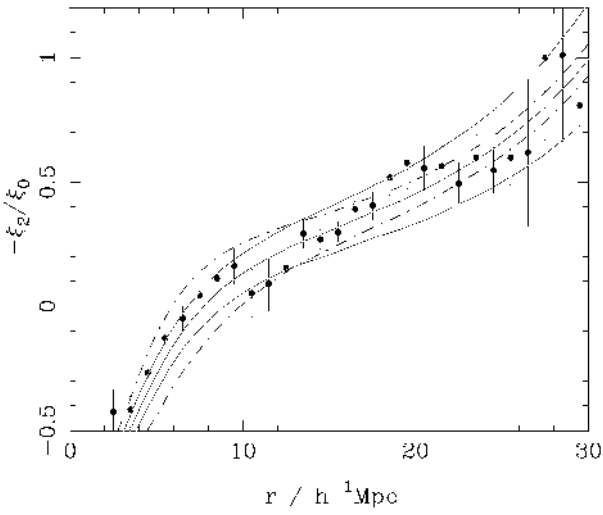


**Figure 2.14.** The flattening of the redshift–space correlation function is quantified by the quadrupole-to-monopole ratio, $\xi_2/\xi_0$. This quantity is positive where fingers-of-God distortion dominates, and is negative where coherent infall dominates. The full curves show model predictions for $\beta = 0.3$, 0.4 and 0.5, with $\sigma_p = 4h^{-1}$ Mpc (full), plus $\beta = 0.4$ with $\sigma_p = 3$, 4 and $5h^{-1}$ Mpc (chain). At large radii, the effects of fingers-of-God are becoming relatively small, and values of $\beta \simeq 0.4$ are clearly appropriate.

The quadrupole-to-monopole ratio should be a clear indicator of coherent infall. In linear theory, it is given by

$$\frac{\xi_2}{\xi_0} = f(n)\frac{4\beta/3 + 4\beta^2/7}{1 + 2\beta/3 + \beta^2/5},$$

where $f(n) = (3+n)/n$ (Hamilton 1992). On small and intermediate scales, the effective spectral index is negative, so the quadrupole-to-monopole ratio should be negative, as observed.

However, it is clear that the results on the largest scales are still significantly affected by finger-of-God smearing. The best way to interpret the observed effects is to calculate the same quantities for a model. To achieve this, we use the observed APM 3D power spectrum, plus the distortion model discussed earlier. This gives the plots shown in figure 2.14. The free parameter is $\beta$, and this has a best-fit value close to 0.4, approximately consistent with other arguments for a universe with $\Omega = 0.3$ and a small degree of large-scale galaxy bias.

### 2.7.8 Galaxy formation and biased clustering

We now come to the difficult question of the relation between the galaxy distribution and the large-scale density field. The formation of galaxies must be

a non-local process to some extent, and the modern paradigm was introduced by White and Rees (1978): galaxies form through the cooling of baryonic material in virialized halos of dark matter. The virial radii of these systems are in excess of 0.1 Mpc, so there is the potential for large differences in the correlation properties of galaxies and dark matter on these scales.

A number of studies have indicated that the observed galaxy correlations may indeed be reproduced by CDM models. The most direct approach is a numerical simulation that includes gas, and relevant dissipative processes. This is challenging, but just starting to be feasible with current computing power Pearce *et al* 1999). The alternative is 'semi-analytic' modelling, in which the merging history of dark-matter halos is treated via the extended Press–Schechter theory (Bond *et al* 1991), and the location of galaxies within halos is estimated using dynamical-friction arguments (e.g. Kauffmann *et al* 1993, 1999, Cole *et al* 1994, Somerville and Primack 1999, van Kampen *et al* 1999, Benson *et al* 2000a, b). Both these approaches have yielded similar conclusions, and shown how CDM models can match the galaxy data: specifically, the low-density flat $\Lambda$CDM model that is favoured on other grounds can yield a correlation function that is close to a single power law over $1000 \gtrsim \xi \gtrsim 1$, even though the mass correlations show a marked curvature over this range (Pearce *et al* 1999, Benson *et al* 2000a; see figure 2.15). These results are impressive, yet it is frustrating to have a result of such fundamental importance emerge from a complicated calculational apparatus. There is thus some motivation for constructing a simpler heuristic model that captures the main processes at work in the full semi-analytic models. The following section describes an approach of this sort (Peacock and Smith 2000; see also Seljak 2000).

An early model for galaxy clustering was suggested by Neyman *et al* (1953), in which the nonlinear density field was taken to be a superposition of randomly placed clumps. With our present knowledge about the evolution of CDM universes, we can make this idealized model considerably more realistic: hierarchical models are expected to contain a distribution of masses of clumps, which have density profiles that are more complicated than isothermal spheres. These issues are well studied in $N$-body simulations, and highly accurate fitting formulae exist, both for the mass function and for the density profiles. Briefly, we use the mass function of Sheth and Tormen (1999; ST) and the halo profiles of Moore *et al* (1999; M99).

$$f(\nu) = 0.216\,17[1 + (\sqrt{2}/\nu^2)^{0.3}]\exp[-\nu^2/(2\sqrt{2})]$$
$$\Rightarrow F(>\nu) = 0.322\,18[1 - \mathrm{erf}(\nu/2^{3/4})] + 0.147\,65\Gamma[0.2, \nu^2/(2\sqrt{2})],$$

where $\Gamma$ is the incomplete gamma function.

Recently, it has been claimed by Moore *et al* (1999; M99) that the commonly adopted density profile of Navarro *et al* (1996; NFW) is in error at small $r$. M99

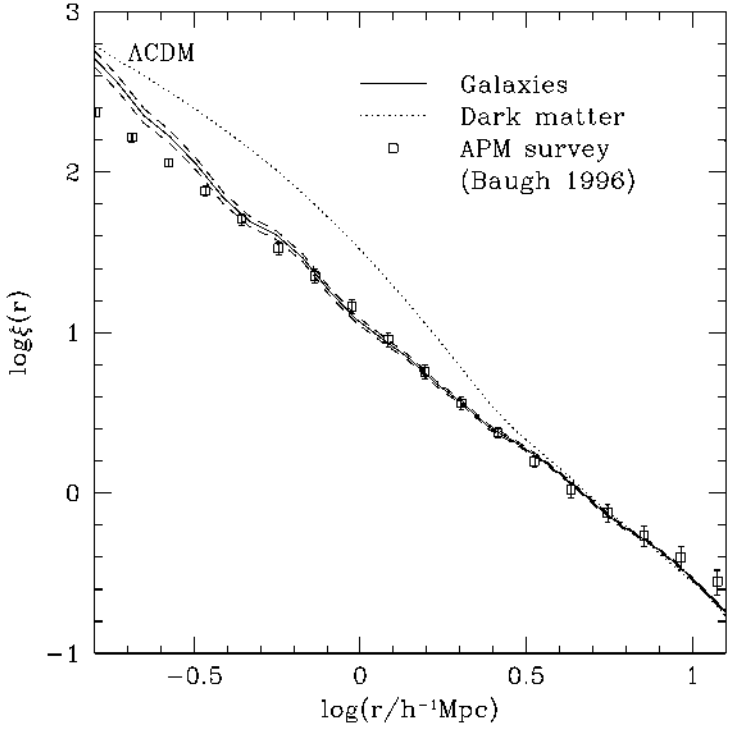


**Figure 2.15.** The correlation function of galaxies in the semi-analytical simulation of an LCDM universe by Benson *et al* (2000a).

proposed the alternative form

$$\rho/\rho_{\rm b} = \frac{\Delta_{\rm c}}{y^{3/2}(1+y^{3/2})} \qquad (r < r_{\rm vir}); \qquad y \equiv r/r_{\rm c}.$$

Using this model, it is then possible to calculate the correlations of the nonlinear density field, neglecting only the large-scale correlations in halo positions. The power spectrum determined in this way is shown in figure 2.16, and turns out to agree very well with the exact nonlinear result on small and intermediate scales. The lesson here is that a good deal of the nonlinear correlations of the dark matter field can be understood as a distribution of random clumps, provided these are given the correct distribution of masses and mass-dependent density profiles.

How can we extend this model to understand how the clustering of galaxies can differ from that of the mass? There are two distinct ways in which a degree of bias is inevitable:

(1) Halo occupation numbers. For low-mass halos, the probability of obtaining an $L^*$ galaxy must fall to zero. For halos with mass above this lower limit,

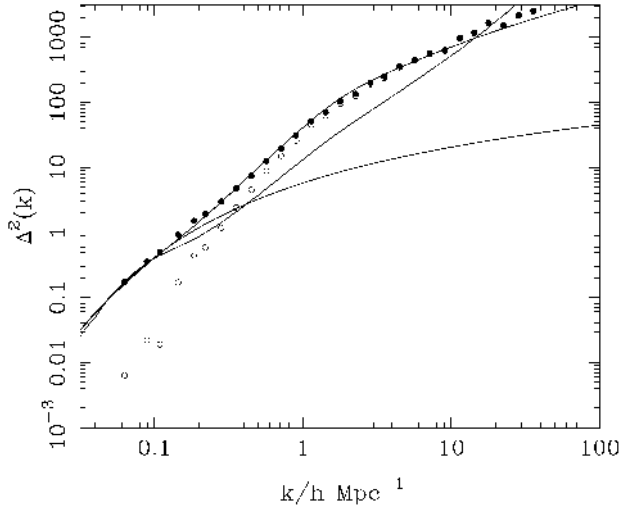


**Figure 2.16.** The power spectrum for the ΛCDM model. The full lines contrast the linear spectrum with the nonlinear spectrum, calculated according to the approximation of PD96. The spectrum according to randomly placed halos is denoted by open circles; if the linear power spectrum is added, the main features of the nonlinear spectrum are well reproduced.

the number of galaxies will in general not scale with halo mass.

(2) Non-locality. Galaxies can orbit within their host halos, so the probability of forming a galaxy depends on the overall halo properties, not just the density at a point. Also, the galaxies will end up at special places within the halos: for a halo containing only one galaxy, the galaxy will clearly mark the halo centre. In general, we expect one central galaxy and a number of satellites.

The numbers of galaxies that form in a halo of a given mass is the prime quantity that numerical models of galaxy formation aim to calculate. However, for a given assumed background cosmology, the answer may be determined empirically. Galaxy redshift surveys have been analysed via grouping algorithms similar to the 'friends-of-friends' method widely employed to find virialized clumps in $N$-body simulations. With an appropriate correction for the survey limiting magnitude, the observed number of galaxies in a group can be converted to an estimate of the total stellar luminosity in a group. This allows a determination of the All Galaxy System (AGS) luminosity function: the distribution of virialized clumps of galaxies as a function of their total luminosity, from small systems like the Local Group to rich Abell clusters.

The AGS function for the CfA survey was investigated by Moore *et al* (1993), who found that the result in blue light was well described by

$$\mathrm{d}\phi = \phi^*[(L/L^*)^\beta + (L/L^*)^\gamma]^{-1}\,\mathrm{d}L/L^*,$$

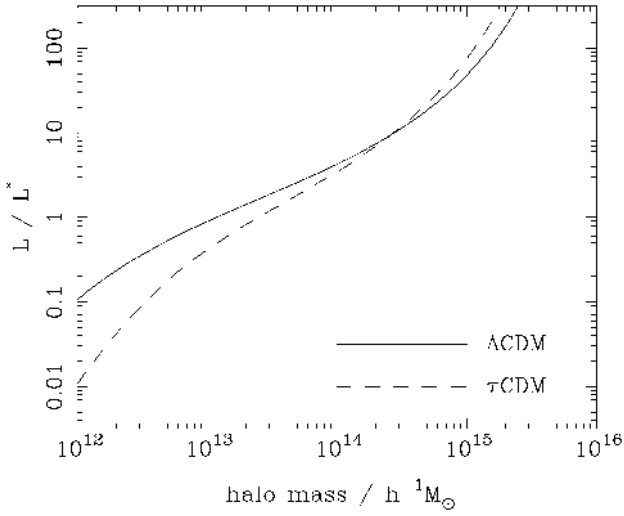


**Figure 2.17.** The empirical luminosity–mass relation required to reconcile the observed AGS luminosity function with two variants of CDM. $L^*$ is the characteristic luminosity in the AGS luminosity function ($L^* = 7.6 \times 10^{10} h^{-2} L_\odot$). Note the rather flat slope around $M = 10^{13}$–$10^{14} h^{-1} M_\odot$, especially for ΛCDM.

where $\phi^* = 0.001\,26h^3$ Mpc$^{-3}$, $\beta = 1.34$, $\gamma = 2.89$; the characteristic luminosity is $M^* = -21.42 + 5\log_{10} h$ in Zwicky magnitudes, corresponding to $M_{\rm B}^* = -21.71 + 5\log_{10} h$, or $L^* = 7.6 \times 10^{10} h^{-2} L_\odot$, assuming $M_{\rm B}^\odot = 5.48$. One notable feature of this function is that it is rather flat at low luminosities, in contrast to the mass function of dark-matter halos (see Sheth and Tormen 1999). It is therefore clear that any fictitious galaxy catalogue generated by randomly sampling the mass is unlikely to be a good match to observation. The simplest cure for this deficiency is to assume that the stellar luminosity per virialized halo is a monotonic, but nonlinear, function of halo mass. The required luminosity–mass relation is then easily deduced by finding the luminosity at which the integrated AGS density $\Phi(> L)$ matches the integrated number density of halos with mass $> M$. The result is shown in figure 2.17.

We can now return to the halo-based galaxy power spectrum and use the correct occupation number, $N$, as a function of mass. This needs a little care at small numbers, however, since the number of halos with occupation number unity affects the correlation properties strongly. These halos contribute no correlated pairs, so they simply dilute the signal from the halos with $N \geq 2$. The existence of antibias on intermediate scales can probably be traced to the fact that a large fraction of galaxy groups contain only one $> L_*$ galaxy. Finally, we need to put the galaxies in the correct location, as discussed before. If one galaxy always occupies the halo centre, with others acting as satellites, the small-scale

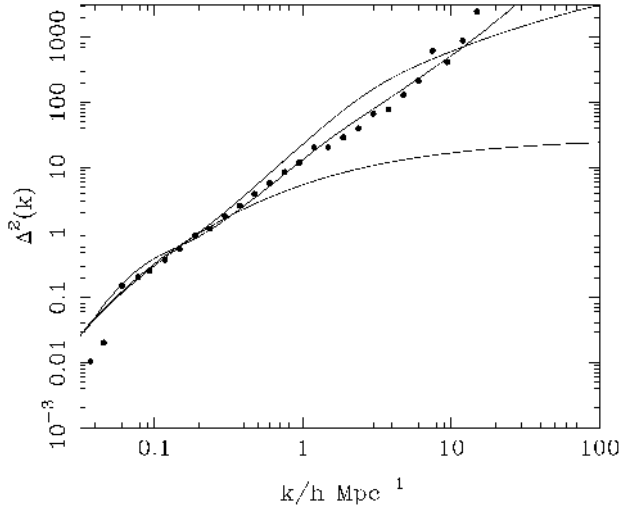


**Figure 2.18.** The power spectrum for a galaxy catalogue constructed from the ΛCDM model. A reasonable agreement with the APM data (full line) is achieved by simple empirical adjustment of the occupation number of galaxies as a function of halo mass, plus a scheme for placing the halos non-randomly within the halos. In contrast, the galaxy power spectrum differs significantly in shape from that of the dark matter (linear and nonlinear theory shown as in figure 2.16).

correlations automatically follow the slope of the halo density profile, which keeps them steep. The results of this exercise are shown in figure 2.18.

The results of this simple model are encouragingly similar to the scale-dependent bias found in the detailed calculations of Benson *et al* (2000a), shown in figure 2.15. There are thus grounds for optimism that we may be starting to attain a physical understanding of the origin of galaxy bias.

## 2.8 Cosmic background fluctuations

### 2.8.1 The hot big bang and the microwave background

What was the state of matter in the early phases of the big bang? Since the present-day expansion will cause the density to decline in the future, conditions in the past must have corresponded to high density—and thus to high temperature. We can deal with this quantitatively by looking at the thermodynamics of the fluids that make up a uniform cosmological model.

The expansion is clearly *adiathermal*, since the symmetry means that there can be no net heat flow through any surface. If the expansion is also reversible, then we can go one step further, because entropy change is defined in terms of

the heat that flows during a reversible change. If no heat flows during a reversible change, then entropy must be conserved, and the expansion will be *adiabatic*. This can only be an approximation, since there will exist irreversible microscopic processes. In practice, however, it will be shown later that the effects of these processes are overwhelmed by the entropy of thermal background radiation in the universe. It will therefore be an excellent approximation to treat the universe as if the matter content were a simple dissipationless fluid undergoing a reversible expansion. This means that, for a ratio of specific heats $\Gamma$, we get the usual adiabatic behaviour

$$T \propto R^{-3(\Gamma-1)}.$$

For radiation, $\Gamma = 4/3$ and we get just $T \propto 1/R$. A simple model for the energy content of the universe is to distinguish pressureless 'dust-like' matter (in the sense that $p \ll \rho c^2$) from relativistic 'radiation-like' matter (photons plus neutrinos). If these are assumed not to interact, then the energy densities scale as

$$\rho_{\rm m} \propto R^{-3} \qquad \rho_{\rm r} \propto R^{-4}$$

The universe must therefore have been *radiation-dominated* at some time in the past, where the densities of matter and radiation cross over. To anticipate, we know that the current radiation density corresponds to thermal radiation with $T \simeq 2.73$ K. In addition to this CMB, we also expect a background in neutrinos. This arises in the same way as the CMB: both photons and neutrinos are in thermal equilibrium at high redshift, but eventually fall out of equilibrium as the universe expands and reaction timescales lengthen. Subsequently, the number density of frozen-out background particles scales as $n \propto a^{-3}$, exactly as expected for a thermal background with $T \propto 1/a$. The background appears to stay in thermal equilibrium even though it has frozen out. If the neutrinos are massless and therefore relativistic, they contribute an energy density comparable to that of the photons (to be exact, a factor 0.68 times the photon density—see p 280 of Peacock (1999)). If there are no other contributions to the energy density from relativistic particles, then the total effective radiation density is $\Omega_{\rm r}h^2 \simeq 4.2 \times 10^{-5}$ and the redshift of *matter–radiation equality* is

$$1 + z_{\rm eq} = 23\,900\Omega h^2 (T/2.73\ {\rm K})^{-4}.$$

The time of this change in the global equation of state is one of the key epochs in determining the appearance of the present-day universe. By a coincidence, this epoch is close to another important event in cosmological history: *recombination*. Once the temperature falls below $\simeq 10^4$ K, ionized material can form neutral hydrogen. Observational astronomy is only possible from this point on, since Thomson scattering from electrons in ionized material prevents photon propagation. In practice, this limits the maximum redshift of observational interest to about 1000; unless $\Omega$ is very low or vacuum energy is important, a matter-dominated model is therefore a good approximation to reality.

In a famous piece of serendipity, the redshifted radiation from the last-scattering photosphere was detected as a 2.73 K microwave background by Penzias and Wilson (1965). Since the initial detection of the microwave background at $\lambda = 7.3$ cm, measurements of the spectrum have been made over an enormous range of wavelengths, from the depths of the Rayleigh–Jeans regime at 74 cm to well into the Wien tail at 0.5 mm. The most accurate measurements come from *COBE*—the NASA cosmic background explorer satellite. Early data showed the spectrum to be very close to a pure Planck function (Mather *et al* 1990), and the final result verifies the lack of any distortion with breathtaking precision. The COBE temperature measurement and 95% confidence range of

$$T = 2.728 \pm 0.004 \text{ K}$$

improves significantly on the ground-based experiments. The lack of distortion in the shape of the spectrum is astonishing, and limits the chemical potential to $|\mu| < 9 \times 10^{-5}$ (Fixsen *et al* 1996). These results also allow the limit $y \lesssim 1.5 \times 10^{-5}$ to be set on the Compton-scattering distortion parameter. These limits are so stringent that many competing cosmological models can be eliminated.

### 2.8.2 Mechanisms for primary fluctuations

At the last-scattering redshift ($z \simeq 1000$), gravitational instability theory says that fractional density perturbations $\delta \gtrsim 10^{-3}$ must have existed in order for galaxies and clusters to have formed by the present. A long-standing challenge in cosmology has been to detect the corresponding fluctuations in brightness temperature of the CMB radiation, and it took over 25 years of ever more stringent upper limits before the first detections were obtained, in 1992. The study of CMB fluctuations has subsequently blossomed into a critical tool for pinning down cosmological models.

This can be a difficult subject; the treatment given here is intended to be the simplest possible. For technical details see, e.g., Bond (1997), Efstathiou (1990), Hu and Sugiyama (1995), Seljak and Zaldarriaga (1996); for a more general overview, see White *et al* (1994) or Partridge (1995). The exact calculation of CMB anisotropies is complicated because of the increasing photon mean free path at recombination: a fluid treatment is no longer fully adequate. For full accuracy, the Boltzmann equation must be solved to follow the evolution of the photon distribution function. A convenient means for achieving this is provided by the public domain *CMBFAST* code (Seljak and Zaldarriaga 1996). Fortunately, these exact results can usually be understood via a more intuitive treatment, which is quantitatively correct on large and intermediate scales. This is effectively what would be called local thermodynamic equilibrium in stellar structure: imagine that the photons we see each originated in a region of space in which the radiation field was a Planck function of a given characteristic temperature. The observed

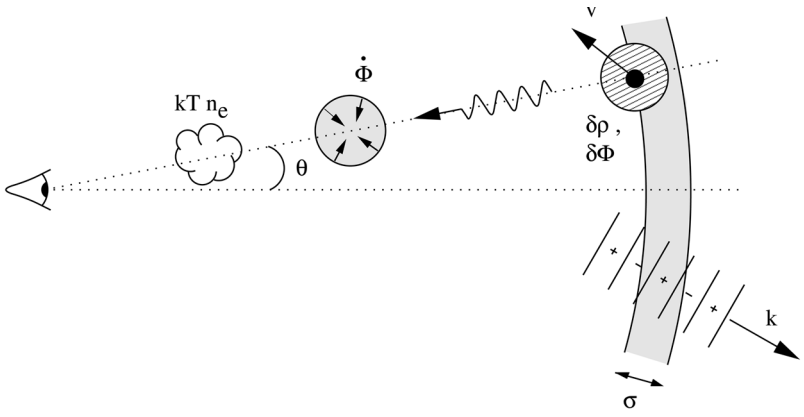


**Figure 2.19.** Illustrating the physical mechanisms that cause CMB anisotropies. The shaded arc on the right represents the last-scattering shell; an inhomogeneity on this shell affects the CMB through its potential, adiabatic and Doppler perturbations. Further perturbations are added along the line of sight by time-varying potentials (the Rees–Sciama effect) and by electron scattering from hot gas (the Sunyaev–Zeldovich effect). The density field at last scattering can be Fourier analysed into modes of wavevector $\boldsymbol{k}$. These spatial perturbation modes have a contribution that is, in general, damped by averaging over the shell of last scattering. Short-wavelength modes are more heavily affected (1) because more of them fit inside the scattering shell and (2) because their wavevectors point more nearly radially for a given projected wavelength.

brightness temperature field can then be thought of as arising from a superposition of these fluctuations in thermodynamic temperature.

We distinguish *primary anisotropies* (those that arise due to effects at the time of recombination) from *secondary anisotropies*, which are generated by scattering along the line of sight. There are three basic primary effects, illustrated in figure 2.19, which are important on respectively large, intermediate and small angular scales:

(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 brightness temperature.

To make quantitative progress, the next step is to see how to predict the size of these effects in terms of the spectrum of mass fluctuations.

### 2.8.3 The temperature power spectrum

The statistical treatment of CMB fluctuations is very similar to that of spatial density fluctuations. We have a 2D field of random fluctuations in brightness temperature, and this can be analysed by the same tools that are used in the case of 2D galaxy clustering.

Suppose that the fractional temperature perturbations on a patch of sky of side $L$ are Fourier expanded:

$$\frac{\delta T}{T}(\boldsymbol{X}) = \frac{L^2}{(2\pi)^2}\int T_K \exp(-\mathrm{i}\boldsymbol{K}\cdot\boldsymbol{X})\,\mathrm{d}^2K$$
$$T_K(\boldsymbol{K}) = \frac{1}{L^2}\int \frac{\delta T}{T}(\boldsymbol{X})\exp(\mathrm{i}\boldsymbol{K}\cdot\boldsymbol{X})\,\mathrm{d}^2X,$$

where $\boldsymbol{X}$ is a 2D position vector on the sky, and $\boldsymbol{K}$ is a 2D wavevector. This is only a valid procedure if the patch of sky under consideration is small enough to be considered flat; we give the full machinery later. We will normally take the units of length to be angle on the sky, although they could also in principle be $h^{-1}$ Mpc at a given redshift. 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, $R_\mathrm{H}$:

$$R_\mathrm{H} = \frac{2c}{\Omega_\mathrm{m}H_0} \qquad \text{(open)}$$
$$R_\mathrm{H} \simeq \frac{2c}{\Omega_\mathrm{m}^{0.4}H_0} \qquad \text{(flat);}$$

the latter approximation for models with $\Omega_\mathrm{m}+\Omega_\mathrm{v}=1$ is due to Vittorio and Silk (1991).

As with the density field, it is convenient to define a dimensionless power spectrum of fractional temperature fluctuations,

$$\mathcal{T}^2 \equiv \frac{L^2}{(2\pi)^2}2\pi K^2|T_K|^2,$$

so that $\mathcal{T}^2$ is the fractional variance in temperature from modes in a unit range of $\ln K$. The corresponding dimensionless spatial statistic is the two-point correlation function

$$C(\theta) = \left\langle \frac{\delta T}{T}(\psi)\frac{\delta T}{T}(\psi+\theta)\right\rangle,$$

which is the Fourier transform of the power spectrum, as usual:

$$C(\theta) = \int \mathcal{T}^2(K)J_0(K\theta)\frac{\mathrm{d}K}{K}.$$

Here, the Bessel function comes from the angular part of the Fourier transform:

$$\int \exp(\mathrm{i}x\cos\phi)\,\mathrm{d}\phi = 2\pi J_0(x).$$

Now, in order to predict the observed anisotropy of the microwave background, the problem we must solve is to integrate the temperature perturbation field through the *last-scattering shell*. In order to do this, we assume that the sky is flat; we also neglect curvature of the 3-space, although this is only strictly valid for flat models with $k = 0$. Both these restrictions mean that the results are not valid for very large angles. Now, introducing the Fourier expansion of the 3D temperature perturbation field (with coefficients $T_k^{3\mathrm{D}}$) we can construct the observed 2D temperature perturbation field by integrating over $k$ space and optical depth:

$$\frac{\delta T}{T} = \frac{V}{(2\pi)^3}\iint T_k^{3\mathrm{D}}\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}\,\mathrm{d}^3k\,\mathrm{e}^{-\tau}\,\mathrm{d}\tau.$$

A further simplification is possible if we approximate $\mathrm{e}^{-\tau}\,\mathrm{d}\tau$ by a Gaussian in comoving radius:

$$\exp(-\tau)\,\mathrm{d}\tau \propto \exp[-(r-r_{\mathrm{LS}})^2/2\sigma_{\mathrm{r}}^2]\,\mathrm{d}r.$$

This says that we observe radiation from a last-scattering shell centred at comoving distance $r_{\mathrm{LS}}$ (which is very nearly identical to $r_{\mathrm{H}}$, since the redshift is so high). The thickness of this shell is of the order of the mean free path to Compton scattering at recombination, which is approximately

$$\sigma_{\mathrm{r}} = 7(\Omega h^2)^{-1/2}\ \mathrm{Mpc}$$

(see p 287 of Peacock 1999).

The 2D power spectrum is thus a smeared version of the 3D one: any feature that appears at a particular wavenumber in 3D will cause a corresponding feature at the same wavenumber in 2D. A particularly simple converse to this rule arises when there are *no* features: the 3D power spectrum is scale-invariant ($\mathcal{T}_{3\mathrm{D}}^2 = $ constant). In this case, for scales large enough that we can neglect the radial smearing from the last-scattering shell,

$$\mathcal{T}_{2\mathrm{D}}^2 = \mathcal{T}_{3\mathrm{D}}^2$$

so that the pattern on the CMB sky is also scale invariant. To apply this machinery for a general spectrum, we now need quantitative expressions for the spatial temperature anisotropies.

**Sachs–Wolfe effect.** To relate to density perturbations, use Poisson's equation $\nabla^2\delta\Phi_k = 4\pi G\rho\delta_k$. The effect of $\nabla^2$ is to pull down a factor of $-k^2/a^2$ ($a^2$ because $k$ is a comoving wavenumber). Eliminating $\rho$ in terms of $\Omega$ and $z_{\mathrm{LS}}$ gives

$$T_k = -\frac{\Omega(1+z_{\mathrm{LS}})}{2}\left(\frac{H_0}{c}\right)^2\frac{\delta_k(z_{\mathrm{LS}})}{k^2}.$$

**Doppler source term.** The effect here is just the Doppler effect from the scattering of photons by moving plasma:

$$\frac{\delta T}{T} = \frac{\delta \boldsymbol{v} \cdot \hat{\boldsymbol{r}}}{c}.$$

Using the standard expression for the linear peculiar velocity, the corresponding $k$-space result is

$$T_k = -\mathrm{i}\sqrt{\Omega(1+z_{\mathrm{LS}})}\left(\frac{H_0}{c}\right)\frac{\delta_k(z_{\mathrm{LS}})}{k}\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{r}}.$$

**Adiabatic source term.** This is the simplest of the three effects mentioned earlier:

$$T_k = \frac{\delta_k(z_{\mathrm{LS}})}{3},$$

because $\delta n_\gamma / n_\gamma = \delta\rho/\rho$ and $n_\gamma \propto T^3$. However, this simplicity conceals a paradox. Last scattering occurs only when the universe recombines, which occurs at roughly a fixed temperature: $kT \sim \chi$, the ionization potential of hydrogen. Surely, then, we should just be looking back to a surface of constant temperature? Hot and cold spots should normalize themselves away, so that the last-scattering sphere appears uniform. The solution is that a denser spot recombines *later*: it is therefore less redshifted and appears hotter. In algebraic terms, the observed temperature perturbation is

$$\left(\frac{\delta T}{T}\right)_{\mathrm{obs}} = -\frac{\delta z}{1+z} = \frac{\delta\rho}{\rho},$$

where the last expression assumes linear growth, $\delta \propto (1+z)^{-1}$. Thus, even though a more correct picture for the temperature anisotropies seen on the sky is of a crinkled surface at constant temperature, thinking of hot and cold spots gives the right answer. Any observable cross-talk between density perturbations and delayed recombination is confined to effects of order higher than linear.

We now draw these results together to form the spatial power spectrum of CMB fluctuations in terms of the power spectrum of mass fluctuations at last scattering:

$$\mathcal{T}^2_{3\mathrm{D}} = [(f_{\mathrm{A}} + f_{\mathrm{SW}})^2(k) + f_{\mathrm{V}}^2(k)\mu^2]\Delta^2_k(z_{\mathrm{LS}}),$$

where $\mu \equiv \hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{r}}$. The dimensionless factors can be written most simply as

$$\begin{aligned} f_{\mathrm{SW}} &= -\frac{2}{(kD_{\mathrm{LS}})^2} \\ f_{\mathrm{V}} &= \frac{2}{kD_{\mathrm{LS}}} \\ f_{\mathrm{A}} &= 1/3, \end{aligned}$$

where

$$D_{\rm LS} = \frac{2c}{\Omega_{\rm m}^{1/2} H_0}(1+z_{\rm LS})^{-1/2} = 184(\Omega h^2)^{-1/2}\ {\rm Mpc}$$

is the comoving horizon size at last scattering (a result that is independent of whether there is a cosmological constant).

We can see immediately from these expressions the relative importance of the various effects on different scales. The Sachs–Wolfe effect dominates for wavelengths $\gtrsim 1h^{-1}$ Gpc; Doppler effects then take over but are almost immediately dominated by adiabatic effects on the smallest scales.

These expressions apply to perturbations for which only gravity has been important up until last scattering, i.e. those larger than the horizon at $z_{\rm eq}$. For smaller wavelengths, a variety of additional physical processes act on the radiation perturbations, generally reducing the predicted anisotropies. An accurate treatment of these effects is not really possible without a more complicated analysis, as is easily seen by considering the thickness of the last-scattering shell, $\sigma_{\rm r} = 7(\Omega h^2)^{-1/2}$ Mpc. This clearly has to be of the same order of magnitude as the photon mean free path at this time; on any smaller scales, a fluid approximation for the radiation is inadequate and a proper solution of the Boltzmann equation is needed. Nevertheless, some qualitative insight into the small-scale processes is possible. The radiation fluctuations will be damped relative to the baryon fluid by photon diffusion, characterized by the Silk-damping scale, $\lambda_{\rm S} = 2.7(\Omega\Omega_{\rm B}h^6)^{-1/4}$ Mpc. Below the horizon scale at $z_{\rm eq}$, $16(\Omega h^2)^{-1}$ Mpc, there is also the possibility that dark-matter perturbations can grow while the baryon fluid is still held back by radiation pressure, which results in adiabatic radiation fluctuations that are less than would be predicted from the dark-matter spectrum alone. In principle, this suggests a suppression factor of $(1+z_{\rm eq})/(1+z_{\rm LS})$ or roughly a factor 10. In detail, the effect is an oscillating function of scale, since we have seen that baryonic perturbations oscillate as sound waves when they come inside the horizon:

$$\delta_{\rm b} \propto (3c_{\rm S})^{1/4} \exp\left(\pm {\rm i}\int k c_{\rm S}\,{\rm d}\tau\right);$$

here, $\tau$ stands for conformal time. There is thus an oscillating signal in the CMB, depending on the exact phase of these waves at the time of last scattering. These oscillations in the fluid of baryons plus radiation cause a set of acoustic peaks in the small-scale power spectrum of the CMB fluctuations (see later).

### 2.8.4 Large-scale fluctuations and CMB power spectrum

The flat-space formalism becomes inadequate for very large angles; the proper basis functions to use are the spherical harmonics:

$$\frac{\delta T}{T}(\hat{\boldsymbol{q}}) = \sum a_\ell^m Y_{\ell m}(\hat{\boldsymbol{q}}),$$

where $\hat{\boldsymbol{q}}$ is a unit vector that specifies direction on the sky. Since the spherical harmonics satisfy the orthonormality relation $\int Y_{\ell m} Y^*_{\ell' m'} \, \mathrm{d}^2 q = \delta_{\ell\ell'}\delta_{mm'}$, the inverse relation is

$$a_\ell^m = \int \frac{\delta T}{T} Y^*_{\ell m} \, \mathrm{d}^2 q.$$

The analogues of the Fourier relations for the correlation function and power spectrum are

$$C(\theta) = \frac{1}{4\pi} \sum_\ell \sum_{m=-\ell}^{m=+\ell} |a_\ell^m|^2 P_\ell(\cos\theta)$$

$$|a_\ell^m|^2 = 2\pi \int_{-1}^{1} C(\theta) P_\ell(\cos\theta) \, \mathrm{d}\cos\theta.$$

These are exact relations, governing the actual correlation structure of the observed sky. However, the sky we see is only one of infinitely many possible realizations of the statistical process that yields the temperature perturbations; as with the density field, we are more interested in the *ensemble average power*. A common notation is to define $C_\ell$ as the expectation value of $|a_\ell^m|^2$:

$$C(\theta) = \frac{1}{4\pi} \sum_\ell (2\ell+1) C_\ell P_\ell(\cos\theta), \qquad C_\ell \equiv \langle |a_\ell^m|^2 \rangle,$$

where now $C(\theta)$ is the ensemble-averaged correlation. For small $\theta$ and large $\ell$, the exact form reduces to a Fourier expansion:

$$C(\theta) = \int_0^\infty \mathcal{T}^2(K) J_0(K\theta) \frac{\mathrm{d}K}{K},$$

$$\mathcal{T}^2(K = \ell + \tfrac{1}{2}) = \frac{(\ell + \frac{1}{2})(2\ell+1)}{4\pi} C_\ell.$$

The effect of filtering the microwave sky with the beam of a telescope may be expressed as a multiplication of the $C_\ell$, as with convolution in Fourier space:

$$C_{\mathrm{S}}(\theta) = \frac{1}{4\pi} \sum_\ell (2\ell+1) W_\ell^2 C_\ell P_\ell(\cos\theta).$$

When the telescope beam is narrow in angular terms, the Fourier limit can be used to deduce the appropriate $\ell$-dependent filter function. For example, for a Gaussian beam of *FWHM* (full-width to half maximum) $2.35\sigma$, the filter function is $W_\ell = \exp(-\ell^2\sigma^2/2)$.

For the large-scale temperature anisotropy, we have already seen that what matters is the Sachs–Wolfe effect, for which we have derived the spatial anisotropy power spectrum. The spherical harmonic coefficients for a spherical slice through such a field can be deduced using the results for large-angle galaxy

clustering, in the limit of a selection function that goes to a delta function in radius:

$$C_\ell^{\rm SW} = 16\pi \int (kD_{\rm LS})^{-4} \Delta_k^2(z_{\rm LS}) j_\ell^2(kR_{\rm H}) \frac{{\rm d}k}{k},$$

where the $j_\ell$ are *spherical Bessel functions* (see chapter 10 of Abramowitz and Stegun 1965). This formula, derived by Peebles (1982), strictly applies only to spatially flat models, since the Fourier expansion of the density field is invalid in an open model. Nevertheless, since the curvature radius $R_0$ subtends an angle of $\Omega/[2(1-\Omega)^{1/2}]$, even the lowest few multipoles are not seriously affected by this point, provided $\Omega \gtrsim 0.1$.

For simple mass spectra, the integral for the $C_\ell$ can be performed analytically. The case of most practical interest is a scale-invariant spectrum ($\Delta_k^2 \propto k^4$), for which the integral scales as

$$C_\ell = \frac{6}{\ell(\ell+1)} C_2$$

(see equation (6.574.2) of Gradshteyn and Ryzhik 1980). The direct relation between the mass fluctuation spectrum and the multipole coefficients of CMB fluctuations mean that either can be used as a measure of the normalization of the spectrum.

### 2.8.5 Predictions of CMB anisotropies

We are now in a position to understand the characteristic angular structure of CMB fluctuations. The change-over from scale-invariant Sachs–Wolfe fluctuations to fluctuations dominated by Doppler scattering has been shown to occur at $k \simeq D_{\rm LS}$. This is one critical angle (call it $\theta_1$); its definition is $\theta_1 = D_{\rm LS}/R_{\rm H}$, and for a matter-only model it takes the value

$$\theta_1 = 1.8\Omega^{1/2} \text{ degrees}.$$

For flat low-density models with significant vacuum density, $R_{\rm H}$ is smaller; $\theta_1$ and all subsequent angles would then be larger by about a factor $\Omega^{-0.6}$ (i.e. $\theta_1$ is roughly independent of $\Omega$ in flat $\Lambda$-dominated models).

The second dominant scale is the scale of last-scattering smearing set by $\sigma_{\rm r} = 7(\Omega h^2)^{-1/2}$ Mpc. This subtends an angle

$$\theta_2 = 4\Omega^{1/2} \text{ arcmin}.$$

Finally, a characteristic scale in many density power spectra is set by the horizon at $z_{\rm eq}$. This is $16(\Omega h^2)^{-1}$ Mpc and subtends

$$\theta_3 = 9h^{-1} \text{ arcmin},$$

independent of $\Omega$. This is quite close to $\theta_2$, so that alterations in the transfer function are an effect of secondary importance in most models.

We therefore expect that all scale-invariant models will have similar CMB power spectra: a flat Sachs–Wolfe portion down to $K \simeq 1$ deg$^{-1}$, followed by a bump where Doppler and adiabatic effects come in, which turns over on arcminute scales through damping and smearing. This is illustrated well in figure 2.22, which shows some detailed calculations of 2D power spectra, generated with the CMBFAST package. From these plots, the key feature of the anisotropy spectrum is clearly the peak at $\ell \sim 100$. This is often referred to as the *Doppler peak*, but it is not so clear that this name is accurate. Our simplified analysis suggests that Sachs–Wolfe anisotropy should dominate for $\theta > \theta_1$, with Doppler and adiabatic terms becoming of comparable importance at $\theta_1$, and adiabatic effects dominating at smaller scales. There are various effects that cause the simple estimate of adiabatic effects to be too large, but they clearly cannot be neglected for $\theta < \theta_1$. A better name, which is starting to gain currency, is the *acoustic peak*. In any case, it is clear that the peak is the key diagnostic feature of the CMB anisotropy spectrum: its height above the SW 'plateau' is sensitive to $\Omega_{\rm B}$ and its angular location depends on $\Omega$ and $\Lambda$. It is therefore no surprise that many experiments are currently attempting accurate measurements of this feature. Furthermore, it is apparent that sufficiently accurate experiments will be able to detect higher 'harmonics' of the peak, in the form of smaller oscillations of amplitude perhaps 20% in power, around $\ell \simeq 500$–$1000$. These features arise because the matter–radiation fluid undergoes small-scale oscillations, the phase of which at last scattering depends on wavelength, since the density oscillation varies roughly as $\delta \propto \exp(\mathrm{i}c_{\rm S}k\tau)$. Accurate measurement of these oscillations would pin down the sound speed at last scattering, and help give an independent measurement of the baryon density.

Since large-scale CMB fluctuations are expected to be dominated by gravitational potential fluctuations, it was possible to make relatively clear predictions of the likely level of CMB anisotropies, even in advance of the first detections. What was required was a measurement of the typical depth of large-scale potential wells in the universe, and many lines of argument pointed inevitably to numbers of order $10^{-5}$. This was already clear from the existence of massive clusters of galaxies with velocity dispersions of up to 1000 km s$^{-1}$:

$$v^2 \sim \frac{GM}{r} \Rightarrow \frac{\Phi}{c^2} \sim \frac{v^2}{c^2},$$

so the potential well of a cluster is of order $10^{-5}$ deep. More exactly, the abundance of rich clusters is determined by the amplitude $\sigma_8$, which measures $[\Delta^2(k)]^{1/2}$ at an effective wavenumber of very nearly $0.17h$ Mpc$^{-1}$. If we assume that this is a large enough scale so that what we are measuring is the amplitude of any scale-invariant spectrum, then the earlier expression for the temperature power spectrum gives

$$\sqrt{\mathcal{T}_{\rm SW}^2} \simeq 10^{-5.7}\Omega\sigma_8[g(\Omega)]^{-1}.$$

There were thus strong grounds to expect that large-scale fluctuations would be present at about the $10^{-5}$ level, and it was a significant boost to the credibility of the gravitational-instability model that such fluctuations were eventually seen.

In more detail, it is possible to relate the COBE anisotropy to the large-scale portion of the power spectrum. Górski *et al* (1995), Bunn *et al* (1995), and White and Bunn (1995) discuss the large-scale normalization from the two-year COBE data in the context of CDM-like models. The final four-year COBE data favour very slightly lower results, and we scale to these in what follows. For scale-invariant spectra and $\Omega = 1$, the best normalization is

$$\text{COBE} \Rightarrow \Delta^2(k) = \left( \frac{k}{0.0737h \text{ Mpc}^{-1}} \right)^4 .$$

Translated into other common notation for the normalization, this is equivalent to $Q_{\text{rms-ps}} = 18.0\,\mu\text{K}$, or $\delta_{\text{H}} = 2.05 \times 10^{-5}$ (see e.g. Peacock and Dodds 1994).

For low-density models, the earlier discussion suggests that the power spectrum should depend on $\Omega$ and the growth factor $g$ as $P \propto g^2/\Omega^2$. Because of the time dependence of the gravitational potential (integrated Sachs–Wolfe effect) and because of spatial curvature, this expression is not exact, although it captures the main effect. From the data of White and Bunn (1995), a better approximation is

$$\Delta^2(k) \propto \frac{g^2}{\Omega^2} g^{0.7}.$$

This applies for low-$\Omega$ models both with and without vacuum energy, with a maximum error of 2% in density fluctuation provided $\Omega > 0.2$. Since the rough power-law dependence of $g$ is $g(\Omega) \simeq \Omega^{0.65}$ and $\Omega^{0.23}$ for open and flat models respectively, we see that the implied density fluctuation amplitude scales approximately as $\Omega^{-0.12}$ and $\Omega^{-0.69}$ respectively for these two cases. The dependence is weak for open models, but vacuum energy implies much larger fluctuations for low $\Omega$.

Within the CDM model, it is always possible to satisfy both the large-scale COBE normalization and the small-scale $\sigma_8$ constraint, by appropriate choice of $\Gamma$ and $n$. This is illustrated in figure 2.20. Note that the vacuum energy affects the answer; for reasonable values of $h$ and reasonable baryon content, flat models require $\Omega_{\text{m}} \simeq 0.3$, whereas open models require $\Omega_{\text{m}} \simeq 0.5$ in order to be consistent with scale-invariant primordial fluctuations.

### 2.8.6 Geometrical degeneracy

The statistics of CMB fluctuations depend on a large number of parameters, and it can be difficult to understand what the effect of changing each one will be. Furthermore, the effects of some parameters tend to change things in opposite directions, so that there are degenerate directions in the parameter space, along which changes leave the CMB unaffected. These were analysed comprehensively by Efstathiou and Bond (1999), and we now summarize the main results.

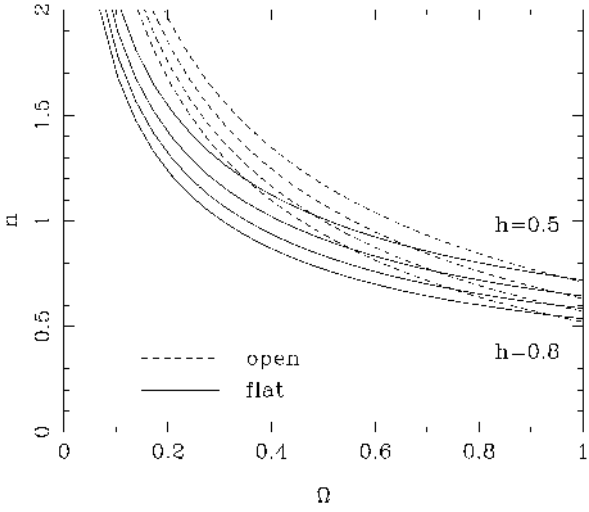


**Figure 2.20.** For 10% baryons, the value of $n$ needed to reconcile COBE and the cluster normalization in CDM models.

The usual expression for the comoving angular-diameter distance is

$$R_0 S_k(r) = \frac{c}{H_0}|1-\Omega|^{-1/2} S_k\left[\int_0^z \frac{|1-\Omega|^{1/2}\,\mathrm{d}z'}{\sqrt{(1-\Omega)(1+z')^2+\Omega_\mathrm{v}+\Omega_\mathrm{m}(1+z')^3}}\right],$$

where $\Omega = \Omega_\mathrm{m} + \Omega_\mathrm{v}$. Defining $\omega_i \equiv \Omega_i h^2$, this can be rewritten in a way that has no explicit $h$ dependence:

$$R_0 S_k(r) = \frac{3000\ \mathrm{Mpc}}{|\omega_k|^{1/2}} S_k\left[\int_0^z \frac{|\omega_k|^{1/2}\,\mathrm{d}z'}{\sqrt{\omega_k(1+z')^2+\omega_\mathrm{v}+\omega_\mathrm{m}(1+z')^3}}\right],$$

where $\omega_k \equiv (1-\Omega_\mathrm{m}-\Omega_\mathrm{v})h^2$. This parameter describes the curvature of the universe, treating it effectively as a physical density that scales as $\rho \propto a^{-2}$. This is convenient for the present formalism, but it is important to appreciate that curvature differs fundamentally from a hypothetical fluid with such an equation of state: the value of $\omega_k$ also sets the curvature index $k$.

The horizon distance at last scattering is $184\omega_\mathrm{m}^{-1/2}$ Mpc. Similarly, other critical length scales such as the sound horizon are governed by the relevant physical density, $\omega_\mathrm{b}$. Thus, if $\omega_\mathrm{m}$ and $\omega_\mathrm{b}$ are given, the shape of the spatial power spectrum is determined. The translation of this into an angular spectrum depends on the angular-diameter distance, which is a function of these parameters, plus $\omega_k$ and $\omega_\mathrm{v}$. Models in which $\omega_\mathrm{m}^{1/2} R_0 S_k(r)$ is a constant have the same angular horizon size. There is therefore a degeneracy between curvature ($\omega_k$) and vacuum ($\omega_\mathrm{v}$): these two parameters can be varied simultaneously to keep the same apparent distance, as illustrated in figure 2.21.

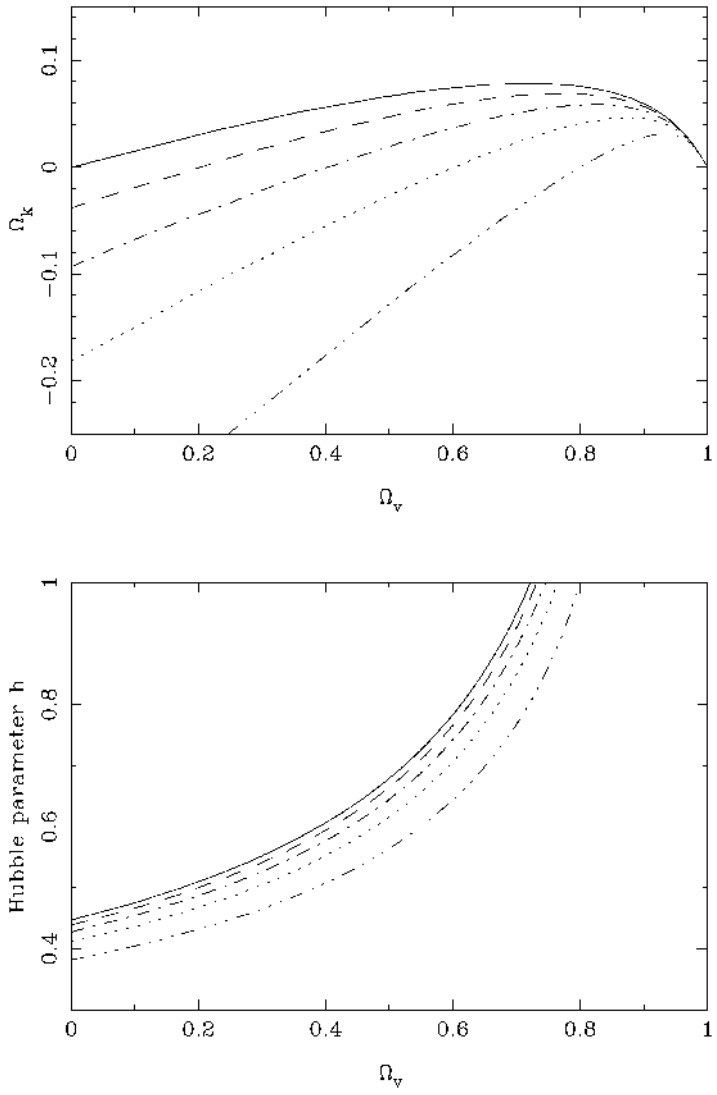


**Figure 2.21.** The geometrical degeneracy in the CMB means that models with fixed $\Omega_{\rm m}h^2$ and $\Omega_{\rm b}h^2$ can be made to look identical by varying the curvature against vacuum energy, while also varying the Hubble parameter. This degeneracy is illustrated here for the case $\omega_{\rm m} \equiv \Omega_{\rm m}h^2 = 0.2$. Models along a given line are equivalent from a CMB point of view; corresponding lines in the upper and lower panels have the same line style. The sensitivity to curvature is strong: if the universe appears to be flat, then it really must be so, unless it is very heavily vacuum dominated. Note that supplying external information about $h$ breaks the degeneracy. This figure assumes scalar fluctuations only; allowing tensor modes introduces additional degeneracies—mainly between the tensor fraction and tilt.

The physical degree of freedom here can be thought of as the Hubble constant. This is involved via the relation

$$h^2 = \omega_{\rm m} + \omega_{\rm v} + \omega_k,$$

so specifying $h$ in addition to the physical matter density fixes $\omega_{\rm v} + \omega_k$ and removes the degeneracy.

### 2.8.7 Small-scale data and outlook

The study of large-scale CMB anisotropies had a huge impact on cosmology in the 1990s, and the field seems likely to be of increasing importance over the next decade. This school was held at a particularly exciting time, as major new data on the CMB power spectrum arrived during the lectures (de Bernardis *et al* 2000, Hanany *et al* 2000). Although these developments are very recent, the situation already seems a good deal clearer than previously, and it is interesting to try to guess where the field is heading.

One immediate conclusion is that it increasingly seems that the relevant models are ones in which the primordial fluctuations were close to being adiabatic and Gaussian. Isocurvature models suffer from the high amplitude of the large-scale perturbations, and do not become any more attractive when modelled in detail (Hu *et al* 1995). Topological defects were for a long time hard to assess, since accurate predictions of their CMB properties were difficult to make. Recent progress does, however, indicate that these theories may have difficulty matching the main details of CMB anisotropies, even as they are presently known (Pen *et al* 1997).

We shall therefore concentrate on interpreting the data in terms of the simplest gravitational-instability models. Many of the features of these models are generic, although they are often spoken of as 'the inflationary predictions'. This statement needs to be examined carefully, since one of the possible prizes from a study of the CMB may be a test of inflation. CMB anisotropies in theories where structure forms via gravitational collapse were calculated in largely the modern way well before inflation was ever considered, by Peebles and Yu (1970). The difficulty in these calculations is the issue of super-horizon fluctuations. In a conventional hot big bang, these must be generated by some acausal process—indeed, an acausal origin is required even for large-scale homogeneity. Inflation is so far the only theory that generates such superhorizon modes at all naturally. Nevertheless, it is not acceptable to claim that detection of super-horizon modes amounts to a proof of inflation. Rather, we need some more characteristic signature of the specific process used by inflation: amplified quantum fluctuations.

We should thus review the predictions that simple models of inflation make for CMB anisotropies (see, e.g., chapter 11 of Peacock 1999 or Liddle and Lyth 2000 for more details). Inflation is driven by a scalar field $\phi$, with a potential

$V(\phi)$. As well as the characteristic energy density of inflation, $V$, this can be characterized by two dimensionless parameters

$$\epsilon \equiv \frac{m_{\rm P}^2}{16\pi}(V'/V)^2$$
$$\eta \equiv \frac{m_{\rm P}^2}{8\pi}(V''/V),$$

where $m_{\rm P}$ is the Planck mass, $V' = {\rm d}V/{\rm d}\phi$, and all quantities are evaluated towards the end of inflation, when the present large-scale structure modes were comparable in size to the inflationary horizon. Prior to transfer-function effects, the primordial fluctuation spectrum is specified by a horizon-scale amplitude (extrapolated to the present) $\delta_{\rm H}$ and a slope $n$:

$$\Delta^2(k) = \delta_{\rm H}^2 \left(\frac{ck}{H_0}\right)^{3+n}.$$

The inflationary predictions for these numbers are

$$\delta_{\rm H} \sim \frac{V^{1/2}}{m_{\rm P}^2 \epsilon^{1/2}}$$
$$n = 1 - 6\epsilon + 2\eta,$$

which leaves us in the unsatisfactory position of having two observables and three parameters.

The critical ingredient for testing inflation by making further predictions is the possibility that, in addition to scalar modes, the CMB could also be affected by gravitational waves (following the original insight of Starobinsky 1985). We therefore distinguish explicitly between scalar and tensor contributions to the CMB fluctuations by using appropriate subscripts. The former category are those described by the Sachs–Wolfe effect, and are gravitational potential fluctuations that relate directly to mass fluctuations. The relative amplitude of tensor and scalar contributions depended on the inflationary parameter $\epsilon$ alone:

$$\frac{C_\ell^{\rm T}}{C_\ell^{\rm S}} \simeq 12.4\epsilon \simeq 6(1-n).$$

The second relation to the *tilt* (which is defined to be $1 - n$) is less general, as it assumes a polynomial-like potential, so that $\eta$ is related to $\epsilon$. If we make this assumption, inflation can be tested by measuring the tilt and the tensor contribution. For simple models, this test should be feasible: $V = \lambda\phi^4$ implies $n \simeq 0.95$ and $C_\ell^{\rm T}/C_\ell^{\rm S} \simeq 0.3$. To be safe, we need one further observation, and this is potentially provided by the spectrum of $C_\ell^{\rm T}$. Suppose we write separate power-law index definitions for the scalar and tensor anisotropies:

$$C_\ell^{\rm S} \propto \ell^{n_{\rm S}-3}, \qquad C_\ell^{\rm T} \propto \ell^{n_{\rm T}-3}.$$

From the discussion of the Sachs–Wolfe effect, we know that, on large scales, the scalar index is the same as index in the matter power spectrum: $n_{\rm S} = n = 1 - 6\epsilon + 2\eta$. By the same method, it is easily shown that $n_{\rm T} = 1 - 2\epsilon$ (although different definitions of $n_{\rm T}$ are in use in the literature; the convention here is that $n = 1$ always corresponds to a constant $\mathcal{T}^2(\ell)$). Finally, then, we can write the *inflationary consistency equation*:

$$\frac{C_\ell^{\rm T}}{C_\ell^{S}} = 6.2(1 - n_{\rm T}).$$

The slope of the scalar perturbation spectrum is the only quantity that contains $\eta$, and so $n_{\rm S}$ is not involved in a consistency equation, since there is no independent measure of $\eta$ with which to compare it.

From the point of view of an inflationary purist, the scalar spectrum is therefore an annoying distraction from the important business of measuring the tensor contribution to the CMB anisotropies. A certain degree of degeneracy exists here (see Bond *et al* 1994), since the tensor contribution has no acoustic peak; $C_\ell^{\rm T}$ is roughly constant up to the horizon scale and then falls. A spectrum with a large tensor contribution therefore closely resembles a scalar-only spectrum with smaller $\Omega_{\rm b}$ (and hence a relatively lower peak). One way in which this degeneracy may be lifted is through polarization of the CMB fluctuations. A non-zero polarization is inevitable because the electrons at last scattering experience an anisotropic radiation field. Thomson scattering from an anisotropic source will yield polarization, and the practical size of the fractional polarization $P$ is of the order of the quadrupole radiation anisotropy at last scattering: $P \gtrsim 1\%$. Furthermore, the polarization signature of tensor perturbations differs from that of scalar perturbations (e.g. Seljak 1997, Hu and White 1997); the different contributions to the total unpolarized $C_\ell$ can in principle be disentangled, allowing the inflationary test to be carried out.

How do these theoretical expectations match with the recent data, shown in figure 2.22? In many ways, the match to prediction is startlingly good: there is a very clear acoustic peak at $\ell \simeq 220$, which has very much the height and width expected for the principal peak in adiabatic models. As we have seen, the location of this peak is sensitive to $\Omega$, since it measures directly the angular size of the horizon at last scattering, which scales as $\ell \propto \Omega^{-1/2}$ for open models. The cut-off at $\ell \simeq 1000$ caused by last-scattering smearing also moves to higher $\ell$ for low $\Omega$; if $\Omega$ were small enough, the smearing cut-off would be carried to large $\ell$, where it would be inconsistent with the upper limits to anisotropies on 10-arcminute scales. This tendency for open models to violate the upper limits to arcminute-scale anisotropies is in fact a long-standing problem, which allowed Bond and Efstathiou (1984) to deduce the following limit on CDM universes:

$$\Omega \gtrsim 0.3h^{-4/3}.$$

The known lack of a CMB peak at high $\ell$ was thus already a very strong argument for a flat universe (with the caveats expressed in the earlier section on geometrical

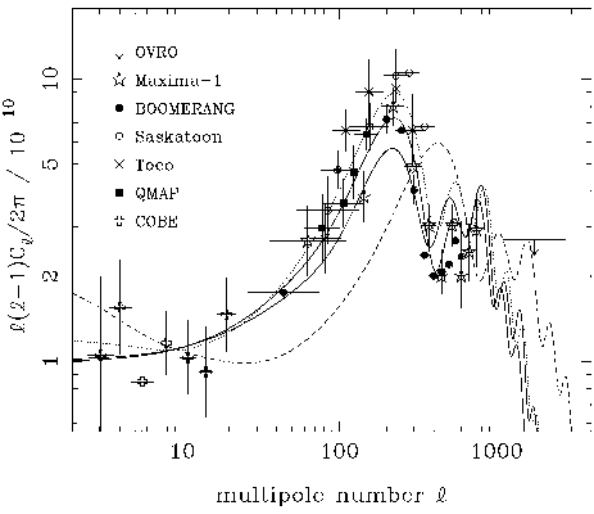


**Figure 2.22.** Angular power spectra $\mathcal{T}^2(\ell) = \ell(\ell+1)C_\ell/2\pi$ for the CMB, plotted against angular wavenumber $\ell$ in rad$^{-1}$. The experimental data are an updated version of the compilation described in White *et al* (1994), communicated by M White; see also Hancock *et al* (1997) and Jaffe *et al* (2000). Various model predictions for adiabatic scale-invariant CDM fluctuations are shown. The two full curves correspond to $(\Omega, \Omega_{\rm B}, h) = (1, 0.05, 0.5)$ and (1,0.1,0.5), with the higher $\Omega_{\rm B}$ increasing power by about 20% at the peak. The dotted line shows a flat $\Lambda$-dominated model with $(\Omega, \Omega_{\rm B}, h) = (0.3, 0.05, 0.65)$; the broken curve shows an open model with the same parameters. Note the very similar shapes of all the curves. The normalization has been set to the large-scale amplitude, and so any dependence on $\Omega$ is quite modest. The main effects are that open models shift the peak to the right, and that the height of the peak increases with $\Omega_{\rm B}$ and $h$.

degeneracy). Now that we have a direct detection of a peak at low $\ell$, this argument for a flat universe is even stronger.

If the basic adiabatic CDM paradigm is adopted, then we can move beyond generic statements about flatness to attempt to use the CMB to measure cosmological parameters. In a recent analysis (Jaffe *et al* 2000), the following best-fitting values for the densities in collisionless matter (c), baryons (b) and vacuum (v) were obtained, together with tight constraints on the power-spectrum index:

$$\Omega_{\rm c} + \Omega_{\rm b} + \Omega_{\rm v} = 1.11 \pm 0.07$$
$$\Omega_{\rm c} h^2 = 0.14 \pm 0.06$$
$$\Omega_{\rm b} h^2 = 0.032 \pm 0.005$$

$$n = 1.01 \pm 0.09.$$

The only parameter left undetermined by the CMB is the Hubble constant, $h$. Recent work (e.g. Mould *et al* 2000) suggests that this is now determined to an rms accuracy of 10%, and we adopt a central value of $h = 0.70$. This completes the cosmological model, requiring a total matter density parameter $\Omega_c + \Omega_b = 0.35 \pm 0.14$, very nicely consistent with what is required to be consistent with $\sigma_8$ for exactly scale-invariant fluctuations. The predicted fluctuation shape is also very sensible for this model: $\Gamma = 0.18$.

The fact that such a 'vanilla' model matches the main cosmological data so well is a striking achievement, but it raises a number of issues. One is that the baryon density inferred from the data exceeds that determined via primordial nucleosynthesis by about a factor 1.5. This may sound like good agreement, but both the CMB and nucleosynthesis are now impressively precise areas of science, and neither can easily accommodate the other's figure. The boring solution is that small systematics will eventually be identified that allow a compromise figure. Alternatively, this inconsistency could be the first sign that something is rotten in the basic framework. However, it is too early to make strong claims in this direction.

Of potentially greater significance is the fact that this successful fit has been achieved using scalar fluctuations alone; indeed, the tensor modes are not even mentioned by Jaffe *et al* (2000). To a certain extent, the presence of tensor modes can be hidden by adjustments in the other parameters. There can be no acoustic peak in the tensor contribution, so that the addition of tensors would require larger peak in the scalar component to compensate, pushing in the direction of universes that are of lower density, with larger baryon fractions. However, this would make it harder to keep higher harmonics of the acoustic oscillations low – and it is the lack of detection of any second and third peaks that forces the high baryon density in this solution. There would also be the danger of spoiling the very good agreement with other constraints, such as the $\sigma_8$ normalization. We therefore have to face the unpalatable fact that there is as yet no sign of the two generic signatures expected from inflationary models: tilt and a significant tensor contribution. It may be that the next generation of CMB experiments will detect such features at a low level. If they do not, and the initial conditions for structure formation remain as they presently seem to be (scale-invariant adiabatic scalar modes), then the heroic vision of using cosmology to probe physics near the Planck scale may not be achieved. The stakes are high.

## References

Abramowitz M and Stegun I A 1965 *Handbook of Mathematical Functions* (New York: Dover)

Adelberger K, Steidel C, Giavalisco M, Dickinson M, Pettini M and Kellogg M 1998 *Astrophys. J.* **505** 18

Ballinger W E, Peacock J A and Heavens A F 1996 *Mon. Not. R. Astron. Soc.* **282** 877
Bardeen J M, Bond J R, Kaiser N and Szalay A S 1986 *Astrophys. J.* **304** 15
Baugh C M and Efstathiou G 1993 *Mon. Not. R. Astron. Soc.* **265** 145
——1994 *Mon. Not. R. Astron. Soc.* **267** 323
Benson A J, Cole S, Frenk C S, Baugh C M and Lacey C G 2000a *Mon. Not. R. Astron. Soc.* **311** 793
Benson A J, Baugh C M, Cole S, Frenk C S, Lacey C G 2000b *Mon. Not. R. Astron. Soc.* **316** 107
Bond J R 1995 *Phys. Rev. Lett.* **74** 4369
——1997 Cosmology and large-scale structure *Proc. 60th Les Houches School* ed R Schaeffer *et al* (Amsterdam: Elsevier) p 469
Bond J R, Cole S, Efstathiou G and Kaiser N 1991 *Astrophys. J.* **379** 440
Bond J R, Crittenden R, Davis R L, Efstathiou G and Steinhardt P J 1994 *Phys. Rev. Lett.* **72** 13
Bond J R and Efstathiou G 1984 *Astrophys. J.* **285** L45
Bunn E F 1995 *PhD Thesis* University of California, Berkeley
Bunn E F, Scott D and White M 1995 *Astrophys. J.* **441** 9
Carlberg R G, Yee H K C, Morris S L, Lin H, Hall P B, Patton D, Sawicki M and Shepherd C W 2000 *Astrophys. J.* **542** 57
Carroll S M, Press W H and Turner E L 1992 *Annu. Rev. Astron. Astrophys.* **30** 499
Cole S, Aragón-Salamanca A, Frenk C S, Navarro J F and Zepf S E 1994 *Mon. Not. R. Astron. Soc.* **271** 781
Colless M 1999 *Phil. Trans. R. Soc.* A **357** 105
Davis M and Peebles P J E 1983 *Astrophys. J.* **267** 465
de Bernardis P *et al* 2000 *Nature* **404** 955
Efstathiou G 1990 Physics of the early Universe *Proc. 36th Scottish Universities Summer School in Physics* ed J A Peacock, A F Heavens and A T Davies (Bristol: Adam Hilger) p 361
——1995 *Mon. Not. R. Astron. Soc.* **274** L73
Efstathiou G, Bernstein G, Katz N, Tyson T and Guhathakurta P 1991 *Astrophys. J.* **380** 47
Efstathiou G and Bond J R 1999 *Mon. Not. R. Astron. Soc.* **304** 75
Efstathiou G, Bond J R and White S D M 1992 *Mon. Not. R. Astron. Soc.* **258** 1P
Efstathiou G, Sutherland W and Maddox S J 1990 *Nature* **348** 705
Eisenstein D J and Hu W 1998 *Astrophys. J.* **496** 605
Eisenstein D J and Zaldarriaga M 1999 *Astrophys. J.* **546** 2
Eke V R, Cole S and Frenk C S 1996 *Mon. Not. R. Astron. Soc.* **282** 263
Feldman H A, Kaiser N and Peacock J A 1994 *Astrophys. J.* **426** 23
Felten J E and Isaacman R 1986 *Rev. Mod. Phys.* **58** 689
Fixsen D J, Cheng E S, Gales J M, Mather J C, Shafer R A and Wright E L 1996 *Astrophys. J.* **473** 576
Gaztañaga E and Baugh C M 1998 *Mon. Not. R. Astron. Soc.* **294** 229
Górski K M, Ratra B, Sugiyama N and Banday A J 1995 *Astrophys. J.* **444** L65
Gradshteyn I S and Ryzhik I M 1980 *Table of Integrals, Series and Products* (New York: Academic Press)
Hamilton A J S 1992 *Astrophys. J.* **385** L5
——1997a *Mon. Not. R. Astron. Soc.* **289** 285
——1997b *Mon. Not. R. Astron. Soc.* **289** 295
——1998 *The Evolving Universe* ed D Hamilton (Dordrecht: Kluwer) pp 185–275

Hamilton A J S, Kumar P, Lu E and Matthews A 1991 *Astrophys. J.* **374** L1
Hanany S *et al* 2000 *Astrophys. J.* **545** L5
Hancock S *et al* 1997 *Mon. Not. R. Astron. Soc.* **289** 505
Heath D 1977 *Mon. Not. R. Astron. Soc.* **179** 351
Hu W, Bunn E F and Sugiyama N 1995 *Astrophys. J.* **447** L59
Hu W and Sugiyama N 1995 *Astrophys. J.* **444** 489
Hu W and White M 1997 *New Astronomy* **2** 323
Huchra J P, Geller M J, de Lapparant V and Corwin H G 1990 *Astrophys. J. Suppl.* **72** 433
Jaffe A *et al* 2000 *Preprint* astro-ph/0007333
Jenkins A, Frenk C S, Pearce F R, Thomas P A, Colberg J M, White S D M, Couchman H M P, Peacock J A, Efstathiou G and Nelson A H 1998 *Astrophys. J.* **499** 20
Kaiser N 1987 *Mon. Not. R. Astron. Soc.* **227** 1
Kauffmann G, Colberg J M, Diaferio A and White S D M 1999 *Mon. Not. R. Astron. Soc.* **303** 188
Kauffmann G, White S D M and Guiderdoni B 1993 *Mon. Not. R. Astron. Soc.* **264** 201
Klypin A, Primak J and Holtzman J 1996 *Astrophys. J.* **466** 13
Lahav O, Lilje P B, Primack J R and Rees M J 1991 *Mon. Not. R. Astron. Soc.* **251** 128
Le Fèvre O *et al* 1996 *Astrophys. J.* **461** 534
Liddle A R and Lyth D 1993 *Phys. Rep.* **231** 1
——2000 *Cosmological Inflation & Large-Scale Structure* (Cambridge: Cambridge University Press)
Liddle A R and Scherrer R J 1999 *Phys. Rev.* D **59** 023509 (astro-ph/9809272)
Mészáros P 1974 *Astron. Astrophys.* **37** 225
Maddox S J, Efstathiou G, Sutherland W J 1996 *Mon. Not. R. Astron. Soc.* **283** 1227
Margon B 1999 *Phil. Trans. R. Soc.* A **357** 93
Mather J C *et al* 1990 *Astrophys. J.* **354** L37
Matsubara T, Szalay A S and Landy S D 2000 *Astrophys. J.* **535** 1
McClelland J and Silk J 1977 *Astrophys. J.* **217** 331
Meiksin A A, White M 1999 *Mon. Not. R. Astron. Soc.* **308** 1179
Meiksin A A, White M and Peacock J A 1999 *Mon. Not. R. Astron. Soc.* **304** 851
Moore B, Frenk C S and White S D M 1993 *Mon. Not. R. Astron. Soc.* **261** 827
Moore B, Quinn T, Governato F, Stadel J and Lake G 1999 *Mon. Not. R. Astron. Soc.* **310** 1147
Mould J R *et al* 2000 *Astrophys. J.* **529** 786
Navarro J F, Frenk C S and White S D M 1996 *Astrophys. J.* **462** 563
Neyman J, Scott E L and Shane C D 1953 *Astrophys. J.* **117** 92
Padmanabhan N, Tegmark M and Hamilton A J S 1999 *Astrophys. J.* **550** 52
Partridge R B 1995 *3K: The Cosmic Microwave Background* (Cambridge: Cambridge University Press)
Peacock J A 1997 *Mon. Not. R. Astron. Soc.* **284** 885
——1999 *Cosmological Physics* (Cambridge: Cambridge University Press)
Peacock J A and Dodds S J 1994 *Mon. Not. R. Astron. Soc.* **267** 1020
——1996 *Mon. Not. R. Astron. Soc.* **280** L19
Peacock J A and Smith R E 2000 *Mon. Not. R. Astron. Soc.* **318** 1144
Pearce F R *et al* 1999 *Astrophys. J.* **521** L99
Peebles P J E 1973 *Astrophys. J.* **185** 413
——1974 *Astrophys. J.* **32** 197

——1980 *The Large-Scale Structure of the Universe* (Princeton, NJ: Princeton University Press)
——1982 *Astrophys. J.* **263** L1
Peebles P J E and Yu J T 1970 *Astrophys. J.* **162** 815
Pen U-L, Seljak U and Turok N 1997 *Phys. Rev. Lett.* **79** 1611
Penzias A A and Wilson R W 1965 *Astrophys. J.* **142** 419
Perlmutter S *et al* 1998 *Astrophys. J.* **517** 565
Pogosyan D and Starobinsky A A 1995 *Astrophys. J.* **447** 465
Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1992 *Numerical Recipes* 2nd edn (Cambridge: Cambridge University Press)
Ratra B and Peebles P J E 1988 *Phys. Rev.* D **37** 3406
Riess A G *et al* 1998 *Astron. J.* **116** 1009
Sachs R K and Wolfe A M 1967 *Astrophys. J.* **147** 73
Saunders W *et al* 2000 *Mon. Not. R. Astron. Soc.* **317** 55
Saunders W, Rowan-Robinson M and Lawrence A 1992 *Mon. Not. R. Astron. Soc.* **258** 134
Scoccimarro R, Zaldarriaga M and Hui L 1999 *Astrophys. J.* **527** 1
Seljak U 1997 *Astrophys. J.* **482** 6
——2000 *Mon. Not. R. Astron. Soc.* **318** 203
Seljak U and Zaldarriaga M 1996 *Astrophys. J.* **469** 437
Shanks T and Boyle B J 1994 *Mon. Not. R. Astron. Soc.* **271** 753
Shanks T, Fong R, Boyle B J and Peterson B A 1987 *Mon. Not. R. Astron. Soc.* **227** 739
Shectman S A, Landy S D, Oemler A, Tucker D L, Lin H, Kirshner R P and Schechter P L 1996 *Astrophys. J.* **470** 172
Sheth R K and Tormen G 1999 *Mon. Not. R. Astron. Soc.* **308** 11
Somerville R S and Primack J R 1999 *Mon. Not. R. Astron. Soc.* **310** 1087
Starobinsky A A 1985 *Sov. Astron. Lett.* **11** 133
Strauss M A and Willick J A 1995 *Phys. Rep.* **261** 271
Sugiyama N 1995 *Astrophys. J. Suppl.* **100** 281
Tegmark M 1996 *Mon. Not. R. Astron. Soc.* **280** 299
Tegmark M, Taylor A N and Heavens A F 1997 *Astrophys. J.* **480** 22
van Kampen E, Jimenez R and Peacock J A 1999 *Mon. Not. R. Astron. Soc.* **310** 43
Viana P T and Liddle A R 1996 *Mon. Not. R. Astron. Soc.* **281** 323
Vittorio N and Silk J 1991 *Astrophys. J.* **385** L9
Vogeley M S and Szalay A S 1996 *Astrophys. J.* **465** 34
Weinberg S 1972 *Gravitation & Cosmology* (New York: Wiley)
——1989 *Rev. Mod. Phys.* **61** 1
White M and Bunn E F 1995 *Astrophys. J.* **450** 477
White M, Scott D and Silk J 1994 *Annu. Rev. Astron. Astrophys.* **32** 319
White S D M, Efstathiou G and Frenk C S 1993 *Mon. Not. R. Astron. Soc.* **262** 1023
White S D M and Rees M 1978 *Mon. Not. R. Astron. Soc.* **183** 341
Zlatev I, Wang L and Steinhardt P J 1999 *Phys. Rev. Lett.* **82** 896