Chapter 8

Inflation and Dark Energy

8.1 Cosmological inflation

8.1.1 Motivation

- In the preceding chapters, we have seen the remarkable success of the cosmological standard model, which is built upon the two symmetry assumptions underlying the class of Friedmann-Lemaître-Robertson-Walker models which experienced a Big Bang a finite time ago. We shall now discuss a fundamental problem of these models, and a possible way out.

- Historically, the problem was raised in a different way, but it becomes obvious with the very straightforward realisation that it is by no means obvious why the CMB should appear as isotropic as it is, and why there should be large coherent structures in it.

- Let us begin with the so-called *comoving particle horizon*, which is the distance that light can travel between the Big Bang and time t. Since light travels on null geodesics, $\mathrm{d}s = 0$, a radial light ray propagates according to $c\mathrm{d}t = a\mathrm{d}w$ [cf. (2.2)]. Therefore,

$$w(t) = \int_0^t \mathrm{d}w = c \int_0^t \frac{\mathrm{d}t'}{a} = c \int_0^t \frac{\mathrm{d}a}{a\dot{a}} \,. \tag{8.1}$$

- Between the Big Bang and the recombination time t_{rec}, the integrand in (8.1) can be approximated by the expansion rate for a matter-dominated universe, or

$$\frac{\dot{a}^2}{a^2} = H_0^2 \Omega_{\mathrm{m}0} a^{-3} \tag{8.2}$$

according to (2.7). Thus,

$$a\dot{a} = H_0 \sqrt{\Omega_{\mathrm{m}0} a} \,, \tag{8.3}$$

the comoving particle horizon becomes

$$w(t_{\rm rec}) = \frac{c}{H_0\sqrt{\Omega_{\rm m0}}}\int_0^{a_{\rm rec}}\frac{{\rm d}a}{\sqrt{a}} = \frac{2c\sqrt{a_{\rm rec}}}{H_0\sqrt{\Omega_{\rm m0}}}\,, \tag{8.4}$$

and the *physical* particle horizon at the time of recombination is $r_{\rm rec} = a_{\rm rec}w_{\rm rec}$.

- On the other hand, we have seen in (7.29) that the angular-diameter distance to the CMB is

$$D_{\rm ang}(a_{\rm rec}) \approx \frac{2ca_{\rm rec}}{H_0\sqrt{\Omega_{\rm m0}}}\,, \tag{8.5}$$

which implies that the *angular* size of the particle horizon is

$$\theta_{\rm rec} = \frac{r_{\rm rec}}{D_{\rm ang}(a_{\rm rec})} \approx \sqrt{a_{\rm rec}} \approx 2^\circ\,. \tag{8.6}$$

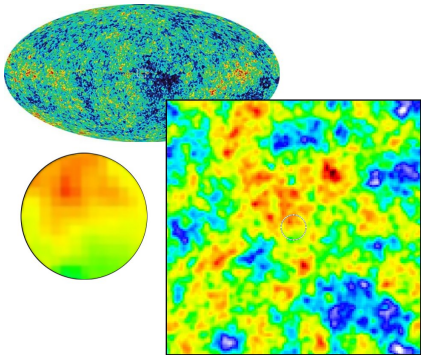


Illustration of the causality problem: The particle horizon at CMB decoupling corresponds to a circle of $\sim 2^\circ$ radius.

- The physical meaning of the particle horizon is that no event between the Big Bang and recombination can exert any influence on a given particle if it is more than the horizon length away. Our simple calculation thus shows that we cannot understand how causal processes could establish identical physical conditions in patches of the sky with radius a few degrees. Points on the CMB separated by larger angles were never causally connected before the CMB was released. It is therefore not at trivial that the CMB could have attained almost the same temperature across the entire sky! The simple fact that the CMB is almost entirely isotropic across the sky thus poses a problem which the standard cosmological model is apparently unable to solve. Moreover, the formation of coherent structures larger than the particle horizon remains mysterious. This is one way to state the *horizon problem*.

- It is sometimes called the *causality problem*: How can coherent structures in the CMB be larger than the particle horizon was at recombination?

- Another uncomfortable problem of the standard cosmological model is the flatness, or at least the near-flatness, of spatial hypersurfaces of our Universe. To see this, we write Friedmann's equation in the form

$$\begin{aligned} H^2(a) &= \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{Kc^2}{a^2} \\ &= H^2(a)\left[\Omega_{\rm total}(a) - \frac{Kc^2}{a^2H^2}\right]\,, \end{aligned} \tag{8.7}$$

from which we conclude

$$|\Omega_{\rm total}(a) - 1| = \frac{Kc^2}{a^2H^2}\,. \tag{8.8}$$

- According to (8.3), we have, in the matter-dominated era,

$$\mathrm{d}t = \frac{\sqrt{a}\mathrm{d}a}{H_0\sqrt{\Omega_{\mathrm{m0}}}} \quad \Rightarrow \quad a \propto t^{2/3} \,, \tag{8.9}$$

hence

$$a^2H^2 = \dot{a}^2 \propto t^{-2/3} \tag{8.10}$$

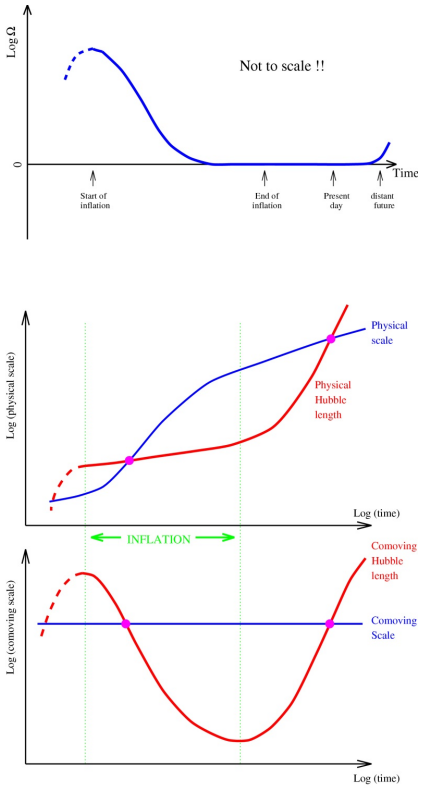
or, from (8.8),

$$|\Omega_{\mathrm{total}}(a) - 1| \propto t^{2/3} \,. \tag{8.11}$$

- This shows that any deviation of the total density parameter Ω_{total} from unity tends to grow with time. Thus, (spatial) flatness is an unstable property. If it is not very precisely flat in the beginning, the Universe will develop away from flatness. Since we know that spatial hypersurfaces are now almost flat, $|\Omega_{\mathrm{total}}(a)-1| \lesssim 1\%$, say, the deviation from flatness must have been at most

$$|\Omega_{\mathrm{total}}(a_{\mathrm{rec}}) - 1| \lesssim 1\% \left(\frac{4\times10^5}{1.4\times10^{10}}\right)^{2/3} \approx 10^{-5} \,, \tag{8.12}$$

or ten parts per million at the time of recombination. Clearly, this requires enormous fine-tuning. This is called the *flatness problem*: How can we understand flatness in the late universe without assuming an extreme degree of fine-tuning at early times?



8.1.2 The idea of inflation

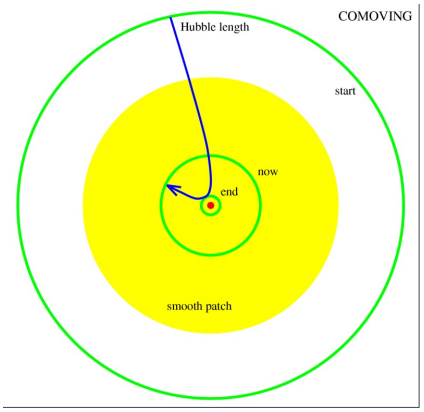
- Since the c/H is the Hubble radius, the quantity $r_{\mathrm{H}} \equiv c/(aH)$ is the *comoving* Hubble radius. During the matter-dominated era, $H \propto a^{-3/2}$ and thus $r_{\mathrm{H}} \propto \sqrt{a}$, while $H \propto a^{-2}$ and $r_{\mathrm{H}} \propto a$ during the radiation-dominated era. Therefore, the comoving Hubble radius typically grows with time. Since we can write (8.8) as

$$|\Omega_{\mathrm{total}}(a) - 1| = Kr_{\mathrm{H}}^2 \,, \tag{8.13}$$

this is equivalent to the flatness problem.

- This motivates the idea that at least the flatness problem would be solved if the comoving Hubble radius could, at least for some sufficiently long period, *shrink* with time. If that could be arranged, any deviation of $\Omega_{\mathrm{total}}(a)$ from unity would be driven towards zero.

- Conveniently, such an arrangement would also remove or at least alleviate the causality problem. Since the Hubble length characterises the radius of the observable universe, it could be driven inside the horizon and thus move the entire observable universe into a causally-connected region. When the hypothesised epoch



The universe can be driven into flatness (top) if the comoving Hubble radius can shrink for sufficiently long time (middle). This can also solve the causality problem (bottom).

of a shrinking comoving Hubble radius is over, it starts expanding again, but if the reduction was sufficiently large, it could remain within the causally-connected region at least until the present.

- How could such a *shrinking* comoving Hubble radius be arranged? Obviously, we require

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{c}{aH} = -\frac{c}{(aH)^2}(\dot{a}H + a\dot{H}) = -\frac{c}{(aH)^2}\left(\frac{\dot{a}^2}{a} + \ddot{a} - \frac{\dot{a}^2}{a}\right) < 0 \,, \tag{8.14}$$

which is possible if and only if $\ddot{a} > 0$, in other words, if the expansion of the Universe *accelerates*.

- This appears counter-intuitive because the cosmic expansion is dominated by gravity, which should be attractive and thus necessarily *decelerate* the expansion. The first law of thermodynamics implies the matter condition

$$\rho c^2 + 3p < 0 \quad \Rightarrow \quad p < -\frac{\rho c^2}{3} \,. \tag{8.15}$$

In other words, cosmic acceleration is possible if the dominant ingredient of the cosmic fluid has sufficiently *negative pressure*.

- When applied to a cosmic sub-volume $V = a^3$, the first law of thermodynamics

$$\mathrm{d}E + p\mathrm{d}V = 0 \quad \Rightarrow \quad \mathrm{d}(\rho c^2 a^3) + p\mathrm{d}a^3 = 0 \tag{8.16}$$

because any heat current would violate isotropy. We thus obtain the equation

$$(\dot{\rho}a^3 + 3\rho a^2\dot{a})c^2 + 3pa^2\dot{a} = 0 \,, \tag{8.17}$$

which implies the density evolution

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) \,. \tag{8.18}$$

- The cosmological constant must have $\dot{\rho} = 0$ and therefore $p = -\rho/c^2$. It has a suitable equation-of-state for cosmic acceleration. (We will see later that Type Ia supernovae, and other observations, motivate a cosmological constant and thus cosmic acceleration).

- If we bring Friedmann's equation (2.7) into the form

$$a^2H^2 = H_0^2\left[\Omega_{\mathrm{m}0}a^{-2} + \Omega_{\mathrm{m}0}a^{-1} - \Omega_{\mathrm{K}} + \Omega_{\Lambda 0}a^2\right] \,, \tag{8.19}$$

it is obvious that a cosmological constant dominates quickly once it becomes comparable to the other density components, because it has the highest power of the scale factor a attached. Once it dominates, (8.19) becomes

$$\dot{a} = H_0\sqrt{\Omega_{\Lambda 0}}a \quad \Rightarrow \quad a \propto \exp\left(H_0\sqrt{\Omega_{\Lambda 0}}t\right) \,, \tag{8.20}$$

and the universe enters into exponential expansion.

8.1.3 Slow roll, structure formation, and observational constraints

- We have seen that we need inflation to solve the flatness and causality problems, and inflation needs a form of matter with *negative pressure*. What could that be? Fortunately, conditions like that are not hard to arrange for particle physics.

- Consider a scalar field ϕ with a self-interaction potential $V(\phi)$. Then, field theory shows that pressure and density of the scalar field are related by the *equation of state*

$$p_\phi = w\rho_\phi c^2 \quad \text{with} \quad w \equiv \frac{\frac{1}{2}\dot{\phi}^2 - V}{\frac{1}{2}\dot{\phi}^2 + V} \,. \tag{8.21}$$

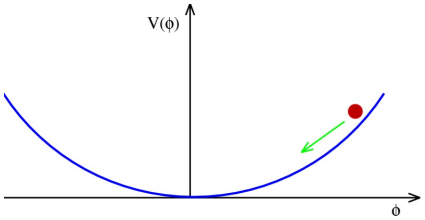
  Evidently, negative pressure is possible if the kinetic energy of the scalar field is sufficiently smaller than its potential energy. For the cosmological-constant case, $\dot{\phi} = 0$, we have $w = -1$ or $p = -\rho c^2$, in agreement with the conclusion from (8.18).

- In other words, a suitably strongly self-interacting scalar field has exactly the properties we need. Inflation, i.e. accelerated expansion, broadly requires $\dot{\phi}^2$ to be sufficiently smaller than V.

- Moreover, we need inflation to operate long enough to drive the total matter density parameter sufficiently close to unity for it to remain there to the present day. These two conditions are conventionally cast into the form

$$\epsilon \equiv \frac{1}{24\pi G}\left(\frac{V'}{V}\right) \ll 1 \quad \text{and} \quad \eta \equiv \frac{1}{8\pi G}\frac{V''}{V} \ll 1 \,. \tag{8.22}$$

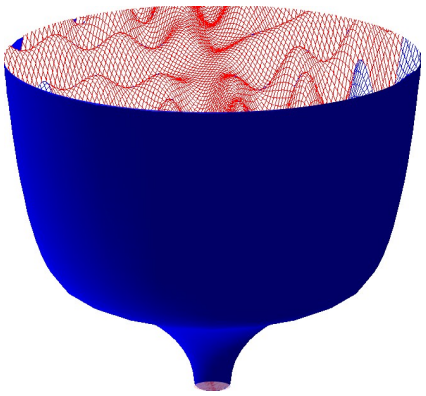
  They are called the *slow-roll conditions*. The first assures that inflation can set in, because if it is satisfied, the potential has a small gradient and cannot drive rapid rapid changes in the scalar field. The second restricts the curvature of the potential and thus assures that the inflationary condition is satisfied long enough.



The slow-roll conditions mean that the potential must be sufficiently flat for inflation to set in, and gently curved for it to last long enough.

- Estimates show that inflation needs to expand the Universe by $\sim 50 \ldots 60$ e-foldings (i.e. by a factor of $\mathrm{e}^{50\ldots60}$) for solving the causality and flatness problems.

- Inflation ends once the slow-roll conditions are violated. By then, the Universe will have become extremely cold. While the density of the inflaton field will be approximately the same as at the onset of inflation (as for the cosmological constant, this is a consequence of the negative pressure), all other matter and radiation fields will have their energy densities lowered by factors of $a^{-3\ldots4}$.



Inflation may expand quantum fluctuations to cosmological scales. It is possible that the large-scale structure in the universe originated from inevitable quantum fluctuations in the very early universe.

- Once ϵ approaches unity, the kinetic term $\dot{\phi}^2$ will dominate the potential, and the scalar field will start oscillating rapidly. It is assumed that the scalar field then decays into ordinary matter which fills or *reheats* the Universe after inflation is over.

- It is an extremely interesting aspect of inflation that it also provides a mechanism for seeding structure formation. As any other quantum field, the *inflaton field* ϕ must have undergone vacumm oscillations because the zero-point energy of a quantum harmonic oscillator cannot vanish due to Heisenberg's uncertainty principle.

- These vacuum oscillations cause the spontaneous creation and annihilation of particle-antiparticle pairs. Once inflation sets in, vacuum fluctuation modes are quickly driven out of the horizon and loose causal connection. Then, they cannot decay any more and "freeze in". Thus, inflation introduces the breath-taking notion that density fluctuations in our Universe today may have been seeded by vacuum fluctuations of the inflaton field before inflation set in and enlarged them to cosmological scales.

- This idea has precisely quantifiable consequences. First, by the central limit theorem, it demands that linear density fluctuations in the Universe should be a Gaussian random field. This is because they arise from incoherent superposition of extremely many independent fluctuation modes whose amplitude and wave number are all drawn from the same probability distribution. Under these circumstances, the central limit theorem shows that the result, i.e. the superposition of all these modes, must be a Gaussian random field.

- Second, it implies that the statistics of density fluctuations in the Universe today must be explicable by the statistics of vacuum fluctuations in a scalar quantum field. This is indeed the case. The power spectrum resulting from this consideration is very close to the scale-free Harrison-Zel'dovich-Peebles shape introduced in Sect. 1.2.2,

$$P_\delta(k) \propto k^n \ , \tag{8.23}$$

  where $n \approx 1$.

- The *spectral index* n would be precisely unity if inflation lasted forever. Since this was obviously not so, n must deviate slightly from unity, and detailed calculations show that it must be slightly smaller,

$$n = 1 + 2\eta - 6\epsilon \ . \tag{8.24}$$

  The latest WMAP measurements do in fact show that

$$n = 0.963^{+0.014}_{-0.015} \ . \tag{8.25}$$

The completely scale-invariant spectrum, $n = 1$, is excluded at 2.5σ.

- The measured deviation of n from unity also restricts the number N of e-foldings completed by inflation; a value of $n = 0.963$ is consistent with 60 e-foldings.

- Another prediction of inflation is that it may excite not only scalar, but also tensor perturbations. Scalar perturbations lead to the density fluctuations, tensor perturbations correspond to gravitational waves. Vector perturbations do not play any role because they decay quickly as the universe expands. Inflation predicts that the ratio r between the amplitudes of tensor and scalar perturbations, taken in the limit of small wave numbers, is

$$r = 16\epsilon \, . \tag{8.26}$$

- An inflationary background of gravitational waves is in principle detectable through the polarisation of the CMB. Limits of order $r \lesssim 0.05$ are expected from the upcoming *Planck* satellite. Together with the result $n \neq 1$ from WMAP, it will then be possible to constrain viable inflation models, i.e. to constrain the shape of the inflaton potential.

8.2 Dark energy

8.2.1 Motivation

- The CMB shows us that the Universe is at least nearly spatially flat. Constraints from the CMB, and from kinematics and cluster evolution (we will discuss this later) show that the matter density alone cannot be responsible for flattening space, and primordial nucleosynthesis and the CMB show that baryons contribute at a very low level only. Something is missing, and it even dominates today's cosmic fluid.

- From structure formation, we know that this remaining constituent cannot clump on the scales covered by the galaxy surveys and below. It is thus different from dark matter. We call it *dark energy*. The type-Ia supernovae (later) tell us that it behaves at least very similar to a cosmological constant.

- Maybe the dark energy *is* a cosmological constant? Nothing currently indicates any deviation from this "simplest" assumption. So far, the cosmological constant is a perfectly viable description for all observational evidence we have.

- However, this is deeply unsatisfactory from the point of view of theoretical physics. The problem is the value of $\Omega_{\Lambda 0}$. As we have seen above, a self-interacting scalar field with negligible kinetic energy behaves like a cosmological constant. Then, its density should simply be given by its potential V. Simple arguments suggest that V should be the fourth power of the Planck mass, which turns out to be 120 *orders of magnitude* larger than the cosmological constant derived from observations. Since this fails, it seems natural to expect that the cosmological constant should vanish, but it does not. The main problem with the cosmological constant is therefore, why is it not zero if it is so small?

- The explanation of inflation by means of an inflaton field suggests one way out. As we have seen there, accelerated expansion can be driven by a self-interacting scalar field while its potential energy dominates. Moreover, it can be shown that if the potential V has an appropriate shape, the dark energy has attractor properties in the sense that a vast range of initial density values can evolve towards the same value today. Such models for a *dynamical dark energy* are theoretically very attractive.

8.2.2 Observational constraints?

- If the dark energy is indeed dynamical and provided by a self-interacting scalar field, how can we find out more about it? Reviewing the cosmological measurements we have discussed so far, it becomes evident that they are all derived from constraints on

  - cosmic time, as in the age of the Galaxy or of globular clusters, or in primordial nucleosynthesis;
  - distances, as in the spatial flatness derived from the CMB, the type-Ia supernovae or the geometry of cosmological weak lensing; or
  - the growth of cosmic structures, as in the acoustic oscillations in the CMB, the evolution of the cluster population, the structures in the galaxy distribution or the source of cosmological weak-lensing effects.

- We must therefore seek to constrain the dark energy by measurements of distances, times, and structure growth. Since they can all be traced back to the expansion behaviour of the universe as described by the Friedmann equation, we must see how the dark energy enters there, and what effects it can seed through it.

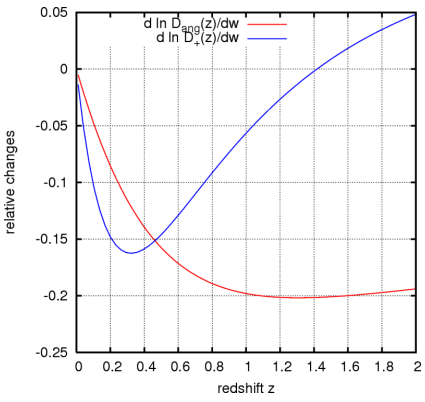
- Let us therefore assume that the dark energy is a suitably self-interacting, homogeneous scalar field. Then, its pressure can be described by
$$p = w(a)\rho c^2 \,, \tag{8.27}$$
where the equation-of-state parameter w is some function of a. According to (8.15), accelerated expansion needs $w < -1/3$, and the cosmological constant corresponds to $w = -1$. Since all cosmological measurements to date are in agreement with the assumption of a cosmological constant, we need to arrange things such that $w \to -1$ today.

- Suppose we have some function $w(a)$, which could either be obtained from a phenomenological choice, a model for the self-interaction potential $V(\phi)$ through (8.21) or from a simple *ad-hoc* parameterisation. Then, (8.18) implies
$$\frac{\dot\rho}{\rho} = -3(1+w)\frac{\dot a}{a} \,, \tag{8.28}$$
or
$$\rho(a) = \rho_0 \exp\left\{-3\int_1^a [1+w(a')]\frac{\mathrm{d}a'}{a'}\right\} \equiv \rho_0 f(a) \,. \tag{8.29}$$

- If $w = $ const., this simplifies to
$$\rho(a) = \rho_0 \exp\left[-3(1+w)\ln a\right] = \rho_0 a^{-3(1+w)} \,. \tag{8.30}$$
If $w = -1$, we recover the cosmological-constant case $\rho = \rho_0 =$ const., for pressure-less material, $w = 0$ and $\rho \propto a^{-3}$, and for radiation, $w = 1/3$ and $\rho \propto a^{-4}$.

- Therefore, we can take account of the dynamical dark energy by replacing the term $\Omega_{\Lambda 0}$ in the Friedmann equation (2.7) by $\Omega_{\mathrm{DE}0} f(a)$, and the expansion function $E(a)$ turns into
$$E(a) = \left[\Omega_{\mathrm{r}0}a^{-4} + \Omega_{\mathrm{m}0}a^{-3} + \Omega_{\mathrm{DE}0}f(a) + \Omega_{\mathrm{K}0}a^{-2}\right]^{1/2} \,, \tag{8.31}$$
where $\Omega_{\mathrm{K}0} = 1 - \Omega_{\mathrm{r}0} - \Omega_{\mathrm{m}0} - \Omega_{\mathrm{DE}0}$ is the curvature density parameter.

- We thus see that the equation-of-state parameter enters the expansion function in integrated form. Since all cosmological observables are integrals over the expansion function, including the growth factor $D_+(a)$ satisfying (2.19), this implies that cosmological observables measure integrals over the integrated equation-of-state function $w(a)$. Needless to say, the dependence of cosmological measurements on the exact form of $w(a)$ will be extremely weak, which in turn implies that extremely accurate measurements will be necessary for constraining the nature of the dark energy.



Logarithmic derivatives of the angular-diameter distance and the growth factor with respect to the equation-of-state parameter.

- In order to illustrate the required accuracies, let us consider by how much the angular-diameter distance and the growth factor change compared to ΛCDM upon changes in w away from -1,

$$\frac{\mathrm{d}\ln D_{\mathrm{ang}}(z)}{\mathrm{d}w}\,, \qquad \frac{\mathrm{d}\ln D_{+}(z)}{\mathrm{d}w}\,, \tag{8.32}$$

  as a function of redshift z. Assuming $\Omega_{\mathrm{m0}} = 0.3$ and $\Omega_{\Lambda 0} = 0.7$, we find typical values between -0.1 and -0.2 at most. Since we currently expect deviations of w from -1 at most at the $\sim 10\%$ level, accurate constraints on the dark energy require relative accuracies of distances and the growth factor at the per-cent level.

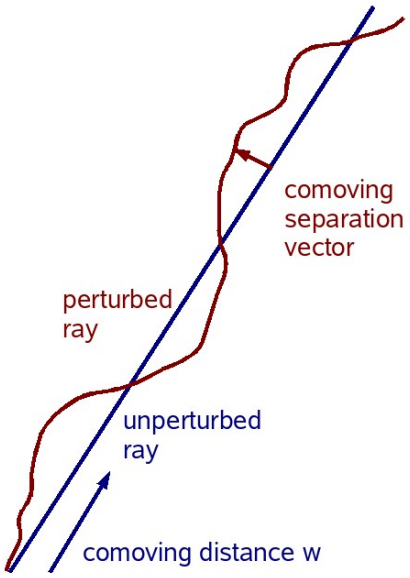
- It seems clear that all suitable cosmological information will need to be combined in order to make any progress. In the next chapters, we will study weak lensing, supernovae, and the growth of cosmic structures as probes of dark energy. All of these are complementary and powerful ways to measure and exploit the expansion history and structure formation history.

Chapter 10

Cosmological Weak Lensing

10.1 Cosmological light deflection

10.1.1 Deflection angle, convergence and shear



Density inhomogeneities along the way deflect light rays.

- Gravitational lensing was mentioned two times before: first in Sect. 3.2 as a means for measuring the Hubble constant through the time delay caused by gravitational light deflection, and second as a means for measuring cluster masses in Sect. 5.2.3. For cosmology as a whole, gravitational lensing has also developed into an increasingly important tool.

- Matter inhomogeneities deflect light. Working out this effect in the limit that the Newtonian gravitational potential is small, $\Phi \ll c^2$ leads to the deflection angle

$$\vec{\alpha}(\vec{\theta}) = \frac{2}{c^2} \int_0^w \mathrm{d}w' \, \frac{f_k(w - w')}{f_k(w)} \vec{\nabla}_\perp \Phi[f_k(w')\vec{\theta}] \; . \tag{10.1}$$

  It is determined by the weighted integral over the gradient of the Newtonian gravitational potential Φ perpendicular to the line-of-sight into direction θ on the observer's sky, and the weight is given by the comoving angular-diameter distance $f_k(w)$ defined in (2.3). The integral extends along the comoving radial distance w' along the line-of-sight to the distance w of the source.

- Equation (10.1) can be intuitively understood. Light is deflected due to the pull of the dimension-less Newtonian gravitational field $\vec{\nabla}_\perp \Phi / c^2$ perpendicular to the otherwise unperturbed line-of-sight, and the effect is weighted by the ratio between the angular-diameter distances from the deflecting potential to the source, $f_k(w - w')$, and from the observer to the source, $f_k(w)$. Thus, a lensing mass distribution very close to the observer gives rise to a large deflection, while a lens near the source, $w' \approx w$, has very

little effect. The factor of two is a relic from general relativity and is due to space-time curvature, which is absent from Newtonian gravity.

- It is important to realise that the deflection itself is not observable. If all light rays emerging from a source would be deflected by the same angle on their way to the observer, no noticeable effect would remain. What is important, therefore, is not the deflection angle itself, but its change from one light ray to the next. This is quantified by the derivative of the deflection angle with respect to the direction $\vec{\theta}$,

$$\frac{\partial\alpha_i}{\partial\theta_j} = \frac{2}{c^2}\int_0^w \mathrm{d}w'\,\frac{f_k(w-w')f_k(w')}{f_k(w)}\,\frac{\partial^2\Phi}{\partial x_i\partial x_j}[f_k(w')\vec{\theta}]\;. \qquad (10.2)$$

  The additional factor $f_k(w')$ in the weight function arises because the derivative of the potential is taken with respect to comoving coordinates x_i rather than the angular components θ_i.

- Obviously, the complete weight function

$$W(w',w) \equiv \frac{f_k(w-w')f_k(w')}{f_k(w)} \qquad (10.3)$$

  vanishes at the observer, $w' = 0$, and at the source, $w' = w$, and peaks approximately half-way in between.

- For applications of gravitational lensing, it is important to distinguish between the trace-free part of the matrix (10.2) and its trace,

$$\mathrm{tr}\frac{\partial\alpha_i}{\partial\theta_j} = \frac{2}{c^2}\int_0^w \mathrm{d}w'\,W(w',w)\,\frac{\partial^2\Phi}{\partial x_i^2}[f_k(w')\vec{\theta}]\;, \qquad (10.4)$$

  where the sum over i is implied. Therefore, the derivatives of Φ can be combined to the two-dimensional Laplacian, which can then be replaced by the three-dimensional Laplacian because the derivatives along the line-of-sight do not contribute to the integral (10.4). Thus, we find

$$\mathrm{tr}\frac{\partial\alpha_i}{\partial\theta_j} = \frac{2}{c^2}\int_0^w \mathrm{d}w'\,W(w',w)\,\Delta\Phi\;. \qquad (10.5)$$

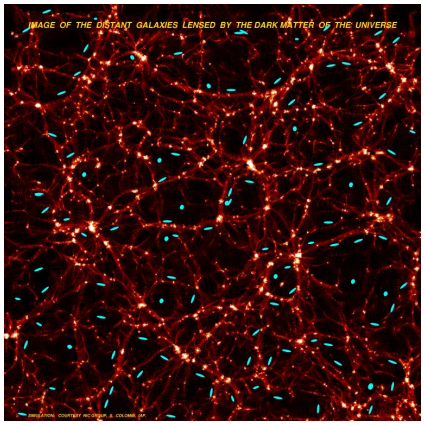
- Next, we can use Poisson's equation to replace the Laplacian of Φ by the density. In fact, we have to take into account that light deflection is caused by density *perturbations*, and that we need the Laplacian in terms of *comoving* rather than *physical* coordinates. Thus,

$$\frac{1}{a^2}\Delta\Phi = 4\pi G\bar{\rho}\delta\;, \qquad (10.6)$$

where δ is the density contrast and

$$\bar{\rho} = \bar{\rho}_0 a^{-3} = \rho_{\rm cr}\Omega_{\rm m0} = \frac{3H_0^2}{8\pi G}\Omega_{\rm m0}a^{-3} \tag{10.7}$$

is the mean matter density.



The gravitational tidal field (shear) of large-scale structures distorts the images of background galaxies (exaggerated).

- Thus, Poisson's equation reads

$$\Delta\Phi = \frac{3}{2}H_0^2\Omega_{\rm m0}\frac{\delta}{a}\,, \tag{10.8}$$

and (10.5) becomes

$$\mathrm{tr}\frac{\partial\alpha_i}{\partial\theta_j} = \frac{3H_0^2\Omega_{\rm m0}}{c^2}\int_0^w \mathrm{d}w'\, W(w',w)\,\frac{\delta}{a} \equiv 2\kappa\,, \tag{10.9}$$

where we have introduced the (effective) *convergence* κ.

- The trace-free part of the matrix (10.2) is

$$\frac{\partial\alpha_i}{\partial\theta_j} - \frac{1}{2}\delta_{ij}\mathrm{tr}\frac{\partial\alpha_i}{\partial\theta_j} = \frac{\partial\alpha_i}{\partial\theta_j} - \delta_{ij}\kappa \equiv \begin{pmatrix}\gamma_1 & \gamma_2\\ \gamma_2 & -\gamma_1\end{pmatrix}\,, \tag{10.10}$$

which defines the so-called *shear* components γ_i. Specifically,

$$\begin{aligned}\gamma_1 &= \frac{1}{c^2}\int_0^w \mathrm{d}w'\, W(w',w)\left(\frac{\partial^2\Phi}{\partial x_1^2} - \frac{\partial^2\Phi}{\partial x_2^2}\right)\,,\\ \gamma_2 &= \frac{2}{c^2}\int_0^w \mathrm{d}w'\, W(w',w)\left(\frac{\partial^2\Phi}{\partial x_1\partial x_2}\right)\,.\end{aligned} \tag{10.11}$$

- Combining the results, we can write the matrix of deflection-angle derivatives as

$$\frac{\partial\alpha_i}{\partial\theta_j} = \begin{pmatrix}\kappa+\gamma_1 & \gamma_2\\ \gamma_2 & \kappa-\gamma_1\end{pmatrix}\,. \tag{10.12}$$

This matrix contains the important information on how an image is magnified and distorted. In the limit of weak gravitational lensing, the *size* of a lensed image is changed by the relative magnification

$$\delta\mu = 2\kappa\,, \tag{10.13}$$

while the image *distortion* is given by the shear components.

- In fact, an originally circular source with radius r will appear as an ellipse with major and minor axes

$$a = \frac{r}{1-\kappa-\gamma}\,,\quad b = \frac{r}{1-\kappa+\gamma}\,, \tag{10.14}$$

where $\gamma \equiv (\gamma_1^2+\gamma_2^2)^{1/2}$. The *ellipticity* of the observed image of a circular source thus provides an estimate for the shear,

$$\epsilon \equiv \frac{a-b}{a+b} = \frac{\gamma}{1-\kappa} \approx \gamma\,. \tag{10.15}$$

10.1.2 Power spectra

- Of course, the exact light deflection expected along a particular line-of-sight cannot be predicted because the mass distribution along that light path is unknown. However, we can predict the *statistical properties* of weak lensing from those of the density-perturbation field.
- We are thus led to the following problem: Suppose the power spectrum $P(k)$ of a Gaussian random density-perturbation field δ is known, what is the power spectrum of any weighted projection of δ along the line-of-sight?
- The answer is given by Limber's equation. Suppose the weight function is $q(w)$ and the projection is

$$g(\vec{\theta}) = \int_0^w \mathrm{d}w'\, q(w')\delta[f_k(w')\vec{\theta}] \,. \tag{10.16}$$

  If $q(w)$ is smooth compared to δ, i.e. if the weight function changes on scales much larger than typical scales in the density contrast, then the power spectrum of g is

$$P_g(l) = \int_0^w \mathrm{d}w'\, \frac{q^2(w')}{f_k^2(w')}\, P\left(\frac{l}{f_k(w')}\right) , \tag{10.17}$$

  where $\vec{l}$ is a two-dimensional wave vector which is the Fourier conjugate variable to the two-dimensional position $\vec{\theta}$ on the sky.
- Strictly speaking, Fourier transforms are inappropriate because the sky is not an infinite, two-dimensional plane. The appropriate set of orthonormal base functions are the spherical harmonics instead. However, lensing effects are usually observed in areas whose solid angle is very small compared to the full sky. If this is so, the survey area can be approximated by a section of the local tangential plane to the sky, and then Fourier transforms can be used. This is the so-called flat-sky approximation.
- Equation (10.9) is clearly of the form (10.16) with the weight function

$$q(w') = \frac{3}{2}\Omega_{\mathrm{m}0}\frac{H_0^2}{c^2}\,\frac{W(w',w)}{a} , \tag{10.18}$$

  thus the power spectrum of the convergence is, according to Limber's equation,

$$P_\kappa(l) = \frac{9\Omega_{\mathrm{m}0}^2}{4}\frac{H_0^4}{c^4}\int_0^w \mathrm{d}w'\, \bar{W}^2(w',w)\, P\left(\frac{l}{f_k(w')}\right) , \tag{10.19}$$

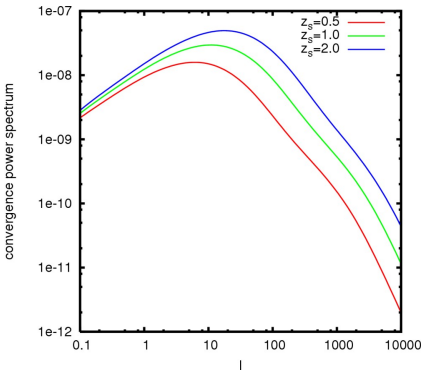
  with a new weight function

$$\bar{W}(w',w) \equiv \frac{W(w',w)}{a f_k(w')} \,. \tag{10.20}$$

- While it is generally difficult or impossible to observe the differential magnification $\delta\mu$ or the convergence κ, image distortions can in principle be measured. With a brief excursion through Fourier space, it can easily be shown that the power spectrum of the shear is exactly identical to that of the convergence,

$$P_\gamma(l) = P_\kappa(l) \; . \tag{10.21}$$

  Thus, the statistics of the image distortions caused by cosmological weak lensing contains integral information on the power spectrum of the matter fluctuations.



The power spectrum of the weak-lensing convergence κ for three different source redshifts.

- Since the shear is defined on the two-dimensional sphere (the observer's sky), its power spectrum is related to its correlation function ξ_γ through the two-dimensional Fourier transform

$$\xi_\gamma(\phi) = \int \frac{\mathrm{d}^2 l}{(2\pi)^2} P_\gamma(l) \mathrm{e}^{\mathrm{i}\vec{\phi}\vec{l}} = \int_0^\infty \frac{l\mathrm{d}l}{2\pi} P_\gamma(l) \mathrm{J}_0(l\phi) \; , \tag{10.22}$$

  where J_ν is the ordinary Bessel function of order ν.

10.1.3 Correlation functions

- In principle, shear correlation functions are measured by comparing the ellipticity of one galaxy with the ellipticity of other galaxies at an angular distance ϕ from the first.

- Ellipticities are oriented, of course, and one has to specify against what other direction the direction of, say, the major axis of a given ellipse is to be compared to. Since correlation functions are measured by counting pairs, a preferred direction is defined by the line connecting the two galaxies of the pair under consideration.

- Let α be the angle between this direction and the major axis of the ellipse, then the *tangential* and *cross* components of the shear are defined by

$$\gamma_+ \equiv \gamma \cos 2\alpha \; , \quad \gamma_\times \equiv \gamma \sin 2\alpha \; . \tag{10.23}$$

  The factor two is important because it accounts for the fact that an ellipse is mapped onto itself when rotated by an angle π. This illustrates that the shear is a spin-2 field: It returns into its original orientation when rotated by π rather than 2π.

- The correlation functions of the tangential and cross components of the shear are

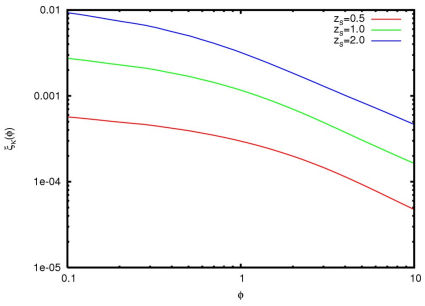
$$\xi_{++}(\phi) = \langle \gamma_+(\theta)\gamma_+(\theta+\phi)\rangle = \frac{1}{2}\int_0^\infty \frac{l\mathrm{d}l}{2\pi} P_\kappa(l) \left[\mathrm{J}_0(l\phi) + \mathrm{J}_4(l\phi)\right] \tag{10.24}$$

and

$$\xi_{\times\times}(\phi) = \langle\gamma_\times(\theta)\gamma_\times(\theta+\phi)\rangle = \frac{1}{2}\int_0^\infty \frac{l\mathrm{d}l}{2\pi} P_\kappa(l)\left[\mathrm{J}_0(l\phi) - \mathrm{J}_4(l\phi)\right] , \tag{10.25}$$

while the cross-correlation between the tangential and cross components must vanish,

$$\xi_{+\times}(\phi) = 0 . \tag{10.26}$$



The convergence (or shear) correlation function for three different source redshifts.

- This suggests to define the correlation functions $\xi_\pm = \xi_{++} \pm \xi_{\times\times}$, which are related to the power spectrum through

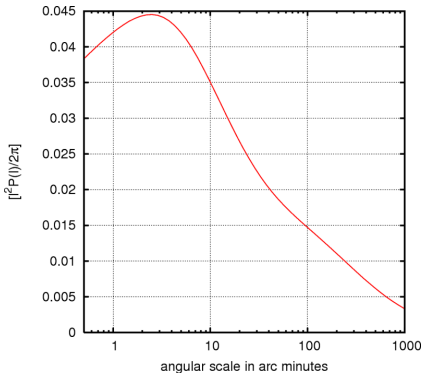
$$\xi_+ = \int_0^\infty \frac{l\mathrm{d}l}{2\pi} P_\kappa(l)\mathrm{J}_0(l\phi) ,$$
$$\xi_- = \int_0^\infty \frac{l\mathrm{d}l}{2\pi} P_\kappa(l)\mathrm{J}_4(l\phi) . \tag{10.27}$$

- Yet another measure for cosmological weak lensing is given by the absolute value of the shear averaged within a circular mask (or *aperture*) of radius θ,

$$\bar{\gamma}(\theta) \equiv \int_0^\theta \frac{\mathrm{d}^2\vartheta}{\pi\theta^2}\gamma(\vec{\vartheta}) , \tag{10.28}$$

which is related to the power spectrum by

$$\langle|\bar{\gamma}(\theta)|^2\rangle = \int_0^\infty \frac{l\mathrm{d}l}{2\pi} P_\kappa(l)\left[\frac{2\mathrm{J}_1(l\theta)}{l\theta}\right]^2 . \tag{10.29}$$



The power of cosmological weak lensing as a function of angular scale.

- The principle of all these measures for cosmic shear is the same: They are integrals of the weak-lensing power spectrum times so-called filter functions which describe the detailed response of the measurement to the underlying power spectrum of density fluctuations. The width of the filter functions controls the range of density-perturbation modes $\vec{k}$ that contribute to one specific mode $\vec{l}$ of weak-lensing on the sky.

- We can now estimate typical numbers for the cosmological weak-lensing effect. The power Δ_κ in the weak-lensing quantities such as the cosmic shear is given by the power spectrum $P_\kappa(l)$ found in (10.19), times the volume in l-space,

$$\Delta_\kappa(l) = l^2 P_\kappa(l) . \tag{10.30}$$

- Assuming a cosmological model with $\Omega_{\mathrm{m}0} = 0.3$ and $\Omega_{\Lambda 0} = 0.7$, the CDM power spectrum and a reasonable source redshift distribution, $\Delta_\kappa(l)^{1/2}$ is found to peak on scales l corresponding to angular scales $2\pi/l$ of $2' \ldots 3'$, and the peak reaches values of $0.04 \ldots 0.05$. This shows that cosmological weak lensing will typically cause source ellipticities of a few per cent, and they have a typical angular scale of a few arc minutes. Details depend on the measure chosen through the filter function.

10.2 Cosmic-shear measurements

10.2.1 Typical scales and requirements

- How can cosmic gravitational lensing effects be measured? As shown in (10.15), the ellipticity of a hypothetic circular source is a direct measure, a so-called *unbiased estimator* for the shear. But typical sources are not circular, but to first approximation elliptical themselves. Thus, measuring their ellipticities yields their intrinsic ellipticities in the first place.

- Let $\epsilon^{(s)}$ be the intrinsic source ellipticity. It is a two-component quantity because an ellipse needs two parameters to be described (e.g. an axis ratio and an orientation), and it is a spin-2 quantity because it is mapped onto itself upon a rotation by $2\pi/2 = \pi$. The cosmic shear adds to that ellipticity, such that the observed ellipticity is
$$\epsilon \approx \epsilon^{(s)} + \gamma \tag{10.31}$$
in the weak-lensing approximation. What is observed is therefore the sum of the signal, γ, and the intrinsic noise component $\epsilon^{(s)}$.

- On sufficiently deep observations, some 30 galaxies per square arc minute are detected. Since the full moon has half a degree diameter, it covers a solid angle of $15^2\pi = 700$ square arc minutes, or $21,000$ of such distant, faint galaxies! From this point of view, the sky is covered by densely patterned "wall paper" of distant galaxies.

- Thus, it is possible to average observed galaxy ellipticities. Assuming their shapes are intrinsically independent, the intrinsic ellipticities will average out, and the shear will remain,
$$\langle\epsilon\rangle \approx \langle\epsilon^{(s)}\rangle + \langle\gamma\rangle \approx \langle\gamma\rangle \,. \tag{10.32}$$

- It is a fortunate coincidence that the typical angular scale of cosmic lensing, which we found to be of order a few arc minutes, is large compared to the mean distance between background galaxies, which is of order $\sqrt{1/30} \approx 0.2'$. This allows averaging over background galaxies without cancelling the cosmic shear signal. If γ varied on scales comparable to or smaller than the mean galaxy separation, any average over galaxies would remove the lensing signal.

- The intrinsic ellipticities of the faint background galaxies have a distribution with a standard deviation of $\sigma_\epsilon \approx 0.3$. Averaging over N of them, and assuming Poisson statistics, gives expectation values of
$$\langle\epsilon^{(s)}\rangle = 0 \,, \quad \delta\epsilon = \langle(\epsilon^{(s)})^2\rangle^{1/2} = \frac{\sigma_\epsilon}{\sqrt{N}} \tag{10.33}$$

for the mean and its intrinsic fluctuation.

- A rough estimate for the signal-to-noise ratio of a cosmic shear measurement can proceed as follows. Suppose the correlation function ξ is measured by counting pairs of galaxies with a separation within $\delta\theta$ of θ. As long as θ is small compared to the side length of the survey area A, the number of pairs will be

$$N_\mathrm{p} = \frac{1}{2}\, 2\pi n^2 A\theta\delta\theta \;, \tag{10.34}$$

and thus the Poisson noise due to the intrinsic ellipticities will be

$$\text{noise} \approx \frac{2\sigma_\epsilon}{n\sqrt{\pi A\theta\delta\theta}} \;, \tag{10.35}$$

where the factor of two arises because of the two galaxies involved in each pair.

- The signal is the square root of the correlation function ξ, which we can approximate as

$$\xi \approx l^2 P_\kappa(l)\delta \ln l \approx l^2 P_\kappa(l)\frac{\delta l}{l} \approx l^2 P_\kappa(l)\frac{\delta\theta}{\theta} \;, \tag{10.36}$$

where we have used in the last step that $\theta = 2\pi/l$.

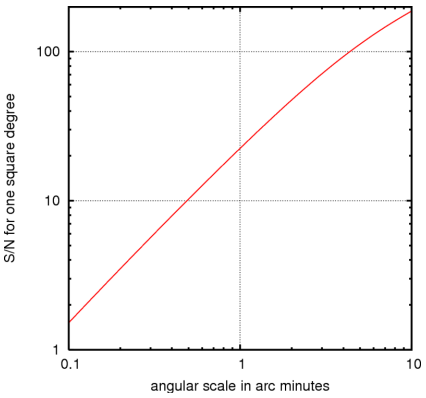
- Thus, the signal-to-noise ratio turns out to be

$$\frac{S}{N} \approx \frac{\sqrt{\xi}}{\text{noise}} \approx \frac{l n \delta\theta\sqrt{\pi A P_\kappa}}{2\sigma_\epsilon} = \frac{n\sqrt{\pi^3 A P_\kappa}}{\sigma_\epsilon}\frac{\delta\theta}{\theta} \;. \tag{10.37}$$

Evidently, the signal-to-noise ratio, and thus the significance of any cosmic-lensing detection, grows with the survey area and decreases with the intrinsic ellipticity of the source galaxies.

- In evaluating (10.37) numerically, we have to take into account that $l^2 P_\kappa(l)$ must be a dimension-less number, which implies that the power spectrum P_κ must have the dimension steradian. Therefore, either the survey area A and the number density n in (10.37) must be converted to steradians, or P_κ must be converted to square arc minutes first.

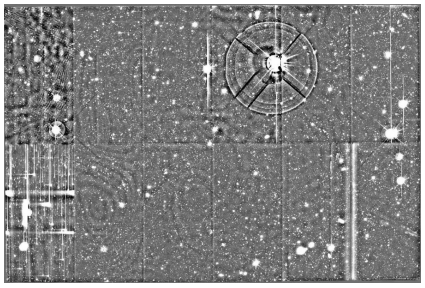
- The signal-to-noise ratio increases approximately linearly with scale. Assuming $\delta\theta/\theta = 0.1$, it is $S/N \approx 1.5$ on a scale of $0.1'$ for a survey of one square degree area. This shows that, if the cosmic shear should be measured on such small scales with an accuracy of, say, five per cent, a survey area of $A \approx (20/1.5)^2 \approx 180$ square degrees is needed since the signal-to-noise ratio scales like the square root of the survey area. On such an area, the ellipticities of $180 \times 3600 \times 30 \approx 2 \times 10^7$ background galaxies have to be accurately measured.



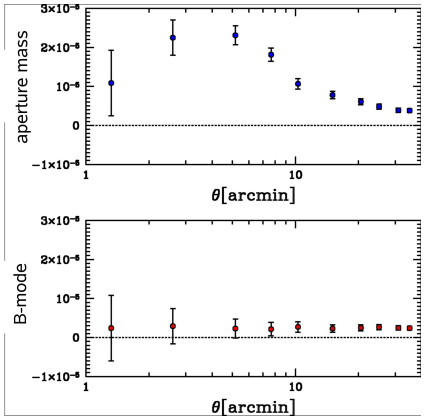
The estimated signal-to-noise ratio of weak-lensing measurements for a hypothetical survey on an area of one square degree.

- Matters are more complicated in reality, but the orders-of-magnitude are well represented by this rough estimate. Bearing in mind that typical fields-of-view of telescopes which are large enough to detect sufficiently many faint background galaxies reach one to ten per cent of a square degree, and that typical exposure times are of order half an hour for that purpose, the total amount of telescope time for a weak-lensing survey like that is estimated to be several thousand telescope hours. With perhaps eight hours of telescope time per night, and perhaps half of the nights per year usable, it is easy to see that the time needed for such surveys is measured in years.

- Since the faint background galaxies have typical sizes of arc seconds, shape measurements require a pixel resolution of, say, 0.1″. The total survey area of 180 square degrees must therefore be resolved into $180 \times 3600 \times 3600/0.1^2 \approx 2.3 \times 10^{11}$ pixels. Storing only one 4-byte number per pixel (i.e. the photon count), this amounts to $4.6 \times 10^{11}/2^{40} = 0.8$ TBytes.



The sobering appearance of real data.



The compatibility of the lower data points signals the almost complete absence of systematic effects in the data show above.

10.2.2 Ellipticity measurements

- The determination of image ellipticities is straightforward in principle, but difficult in practice. Usually, the surface-brightness quadrupole

$$Q_{ij} = \frac{\int I(\vec{x}) x_i x_j \mathrm{d}^2 x}{\int I(\vec{x}) \mathrm{d}^2 x} \tag{10.38}$$

  is measured, from whose principal axes the ellipticity can be read off.

- Real galaxy images, however, are typically far from ideally elliptical. They are structured or otherwise irregular. In addition, if they are small, they are coarsely resolved into just a few pixels, so that only a crude approximation to the integral in (10.38) can be found.

- Even if the surface-brightness quadrupole of the image on the detector can be accurately determined, the image appears affected by imperfections of the telescope optics and by the turbulence in the atmosphere, the so-called *seeing*.

- Due to the wave nature of light and the finite size of the telescope mirror, the telescope will have finite resolution. The angular resolution limit is given by

$$\Delta\theta \approx 1.44 \frac{\lambda}{D} \tag{10.39}$$

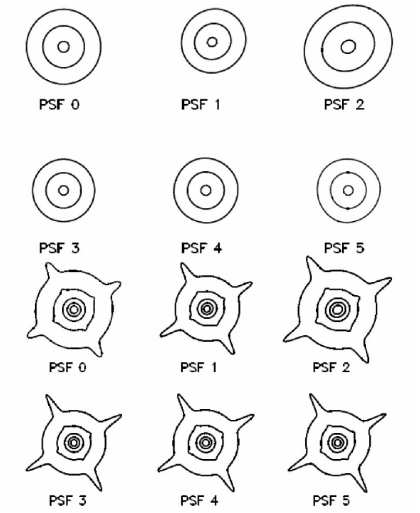
as mentioned in (7.41) before. With $\lambda \approx 6 \times 10^{-5}$ cm and $D = 400$ cm, the angular resolution is $\Delta\theta \approx 0.04''$, much smaller than needed for our purposes.

- The turbulence of the Earth's atmosphere effectively convolves images with a Gaussian whose width depends on the site, the weather and other conditions. Typical seeing ranges around $1''$. Under very good conditions, it can shrink to $\sim 0.5''$ or less. Clearly, if an image of approximately one arc second size is convolved with a Gaussian of similar width, any ellipticity is substantially reduced.

- How the image of a point-like source, such as a star, appears on the detector is described by the so-called *point-spread function* (PSF). The PSF may be anisotropic if the telescope optics is slightly astigmatic, and this anisotropy may, and will in general, depend on the location on the focal plane. The image is a convolution of the ideal image shape before any distortion by the atmosphere and the telescope optics and the PSF. Any accurate measurement of image ellipticities requires a PSF deconvolution, for which the PSF must of course be known. It is measured by fitting elliptical Gaussians to stellar images on the exposure.

- Many other effects may distort images in systematic ways. For instance, if the CCD chips are not exactly perpendicular to the optical axis of the telescope, or if the individual chips of a CCD mosaic are not exactly in the same plane, or if the telescope is slightly out of focus, systematic image deformations may result which typically vary across the focal plane. They have to be measured and corrected. This is commonly achieved by fitting the parameters of a model PSF to a low-order, two-dimensional polynomial on the focal plane. Since part of the image distortions may depend on time due to thermal deformation, changing atmospheric conditions and such, PSF corrections will also typically depend on time and have to be determined and applied with much care.

- Systematic effects may remain which need to be detected and quantified. Any coherent image distortions caused by gravitational lensing must be describable by the tidal gravitational field, i.e. by second-order derivatives of a scalar potential. In analogy to the $\vec{E}$-field in electromagnetism, such distortion patterns are called E-modes. Similarly, distortion patterns which are the curl of a vector field are called B-modes. They cannot be due to gravitational lensing and thus signal systematic effects remaining in the data. Such B-mode contaminations could recently be strongly reduced or suppressed by improved algorithms for PSF correction.



The point-spread function of the Canada-France-Hawaii telescope.

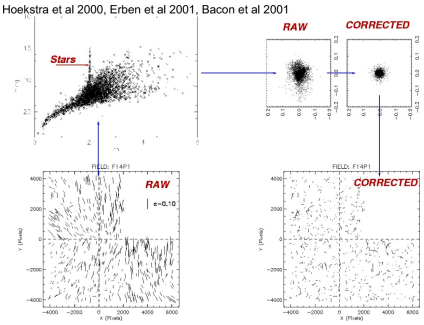
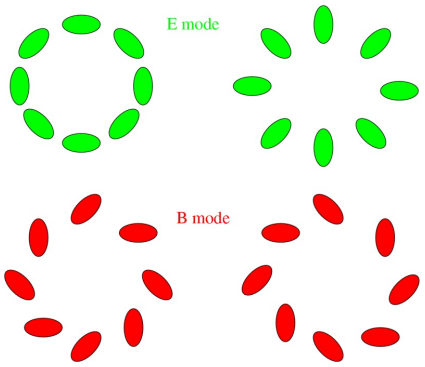


Illustration of systematic image distortions in the CFHTLS and their correction.



E- and B-mode distortion patterns.

10.2.3 Results

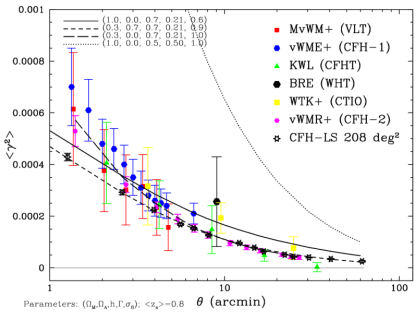
- Despite the smallness of the effect and the many difficulties in measuring it, much progress in cosmic-shear observations has been achieved in the past few years. Current and ongoing surveys, in particular the Canada-France-Hawaii Legacy Survey, combined with well-developed, largely automatic data-analysis pipelines, have managed to produce cosmic-shear correlation functions with very small error bars covering angular scales from below an arc minute to several degrees. The best correlation functions could be shown to be at most negligibly contaminated by B-modes.

- The power spectrum $P_\kappa(l)$ depends crucially on the non-linear evolution of the dark-matter power spectrum. This, and the exact redshift distribution of the background galaxies, are the major uncertainties now remaining in the interpretation of cosmic-shear surveys. Apart from that, the measured cosmic-shear correlation functions agree very well with the theoretical expectation from CDM density fluctuations in a spatially-flat, low-density universe.

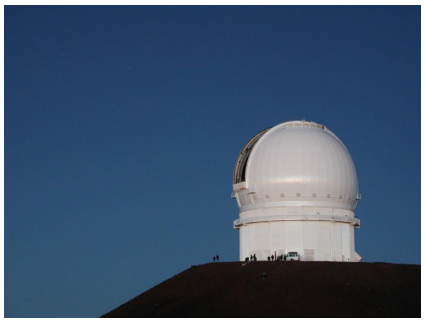
- As (10.19) shows, the weak-lensing power spectrum $P_\kappa(l)$ depends on the product of a factor $\Omega_{\rm m0}^2$ due to the Poisson equation, times the amplitude A of the matter power spectrum. An additional weak dependence on cosmological parameters is caused by the geometric weight function $\bar{W}(w', w)$, but this is not very important. By and large, therefore, the cosmic-shear correlation function measures the product $A\Omega_{\rm m0}^2$, which means that the amplitude of the power spectrum is (almost) precisely degenerate with the matter density parameter. Only if it is possible to constrain $\Omega_{\rm m0}$ or A in any other way can the degeneracy be broken.

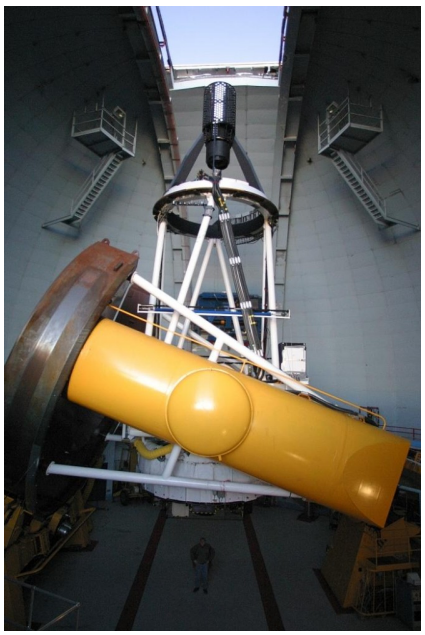
- We shall see later how this may work. The amplitude of the power spectrum A is conventionally described by a parameter σ_8^2 which will be defined and described in more detail later. Weak lensing thus measures the product $\sigma_8\Omega_{\rm m0}$, and current measurements find $\sigma_8\Omega_{\rm m0} \approx 0.2$.

- Weak gravitational lensing is a fairly new field of cosmological research. Within a few years, it has considerably matured and returned cosmologically interesting constraints. Considerable potential is expected from weak lensing in wide-area surveys in particular when combined with photometric redshift information.



The first published measurements of the cosmic-shear correlation function.





The CFHT dome (top) and the Mega-Prime Camera in its prime focus (bottom).

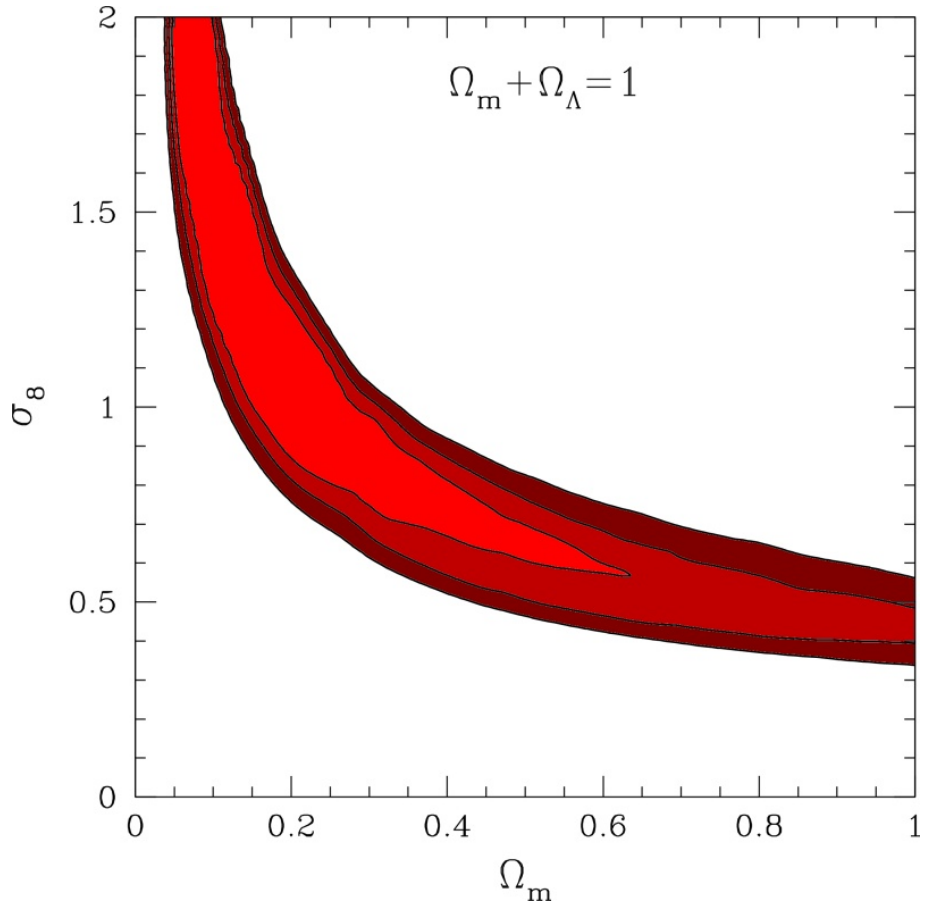


Figure 10.1: Recent constraints in the $\Omega_{m0} - \sigma_8$ plane obtained from weak-lensing measurements. The Universe is assumed spatially flat here.

Chapter 11

The Normalisation of the Power Spectrum

11.1 Introduction

- We saw in Chapter 9 that the measured power spectrum of the galaxy distribution follows the CDM expectation in the range of wave numbers where current large surveys allow it to be determined. This range can be extended to some degree towards smaller scales by measuring the autocorrelation of hydrogen absorption lines in the spectra of distant quasars. Such observations of the power spectrum of the so-called Lyman-α forest lines show that the power spectrum does indeed turn towards the asymptotic behaviour $\propto k^{-4}$. In addition, we have seen that the peak location agrees with the expectation for universe with $\Omega_{\rm m0} \approx 0.3$ and $h \approx 0.72$. This indicates that the CDM expectation for the dark-matter power spectrum is indeed at least very close to its real shape, which is a remarkable success.

- Although the shape of the power spectrum could thus be quite well established, its amplitude still poses a surprisingly obstinate problem. We shall see in this section why it is so difficult to measure. For this purpose, we shall discuss three ways of measuring σ_8; the amplitude of large-scale temperature fluctuations in the CMB, the cosmic-shear autocorrelation function, and the abundance and evolution of the galaxy-cluster population.

- For historical reasons, the amplitude of the dark-matter power spectrum is characterised by the density-fluctuation variance within spheres of $8\,h^{-1}$Mpc radius. This is because in the first measurement of the fluctuation amplitude in the galaxy distribution, Davis & Peebles found that it reached unity in such spheres.

- More generally, one imagines randomly placing spheres of radius

R and measuring the density-contrast variance within them. Since the variance in Fourier space is characterised by the power spectrum, it can be written as

$$\sigma_R^2 = \int_0^\infty \frac{\mathrm{d}^3k}{(2\pi)^3} P_\delta(k) W_R^2(k) \,, \tag{11.1}$$

where $W_R(k)$ is a *window function* selecting the k modes contributing to the variance.

- Imagining spheres of radius R in real space, the window function should be the Fourier transform of a step function, which is, however, inconvenient because it extends to infinite wave numbers. It is thus more common to use either Gaussians, since they Fourier transform into Gaussians, or step functions in Fourier space. For simplicity of the illustrative calculations that will follow, we use the latter choice, thus

$$W_R(k) = \Theta(k_R - k) = \Theta\left(\frac{2\pi}{R} - k\right) \,. \tag{11.2}$$

This is a step function dropping to zero for $k > 2\pi/R$.

- Inserting this into (11.1), we find

$$\sigma_R^2 = \int_0^{2\pi/R} \frac{k^2 \mathrm{d}k}{2\pi^2} P_\delta(k) \,. \tag{11.3}$$

In other words, all modes *larger* than R contribute to the density fluctuations in spheres of radius R because all smaller modes average to zero.

- The normalisation of the power spectrum is usually expressed in terms of σ_8.

11.2 Fluctuations in the CMB

11.2.1 The large-scale fluctuation amplitude

- We saw in Chapter 6 that the long-wavelength (low-k) tail of the CMB power spectrum is caused by the Sachs-Wolfe effect, giving rise to relative temperature fluctuations of

$$\frac{\delta T}{T} \equiv \tau = \frac{\Phi}{3c^2} \tag{11.4}$$

in terms of the Newtonian potential fluctuations Φ; see also Eq. (7.22).

- The three-dimensional temperature-fluctuation power spectrum is then

$$P_\tau(k) = \frac{1}{9c^4} P_\Phi(k) \; . \tag{11.5}$$

The Poisson equation in its form (10.8) implies that the power spectra of potential- and density fluctuations are related through

$$P_\Phi(k) = \frac{9H_0^4}{3}\Omega_{\mathrm{m}0}^2 \left(\frac{D_+(a)}{a}\right)^2 \frac{P_\delta(k)}{k^4} \; , \tag{11.6}$$

where the linear growth factor $D_+(a)$ was introduced to relate the potential-fluctuation power spectrum at the time of decoupling to the *present* density-fluctuation power spectrum $P_\delta(k)$.

- Now, we need to account for projection effects. A three-dimensional mode with wave number k and wavelength $\lambda = 2\pi/k$ appears under an angle $\theta = \lambda/D$, where D is the angular-diameter distance to the CMB. We saw in (7.29) that

$$D \approx \frac{2ca}{H_0\sqrt{\Omega_{\mathrm{m}0}}} \propto \frac{1}{H_0\sqrt{\Omega_{\mathrm{m}0}}} \tag{11.7}$$

to first order in $\Omega_{\mathrm{m}0}$. Thus, the *angular* wave number under which the mode appears is

$$l \approx \frac{2\pi}{\theta} \approx Dk \; . \tag{11.8}$$

- Expressing now the power spectrum (11.5) in terms of the *angular* wave number l yields

$$P_\tau(l) \propto \left(\frac{H_0}{c}\right)^4 \Omega_{\mathrm{m}0}^2 \left(\frac{D_+(a)}{a}\right)^2 \frac{1}{D^2}\frac{D^4}{l^4} P_\delta\left(\frac{l}{D}\right) \; , \tag{11.9}$$

where the factor D^{-2} arises because of the transformation from *spatial* to *angular* wave numbers l, and the factor D^4/l^4 expresses the factor k^{-4} from the squared Laplacian.

- Let us now insert a highly simplified model for the power spectrum,

$$P_\delta(k) = A\begin{cases} k^n & (k < k_0) \\ k^{n-4} & \text{else} \end{cases} \; . \tag{11.10}$$

Inserting its long-wave limit, $P_\delta(k) = Ak^n$, into (11.9) yields

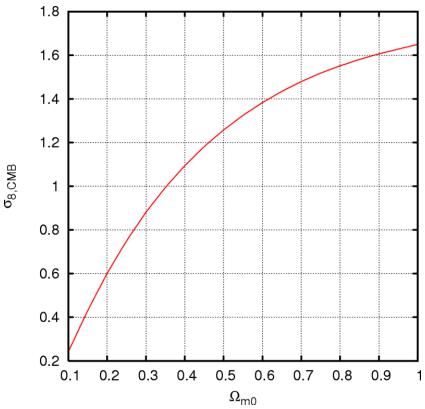
$$P_\tau(l) \propto A\left(\frac{H_0}{c}\right)^4 \Omega_{\mathrm{m}0}^2 \left(\frac{D_+(a)}{a}\right)^2 \frac{1}{D^2}\left(\frac{D}{l}\right)^{4-n} \; . \tag{11.11}$$

- This shows that the temperature-fluctuation power spectrum depends on the cosmological parameters in various subtle ways; through the Poisson equation, the projection, the angular-diameter distance, the growth factor and the power-spectrum exponent n.

- Taking all dependences on H_0 and $\Omega_{\rm m0}$ into account shows that the amplitude A of the dark-matter power spectrum depends on the cosmological parameters through

$$A \propto \Omega_{\rm m0}^{-1-n/2} h^{-2-n} \left(\frac{D_+(a)}{a} \right)^{-2} P_\tau(l) \; . \tag{11.12}$$

  In other words, the *measured* power $P_\tau(l)$ in the CMB temperature fluctuations can only be translated into the amplitude of the dark-matter power spectrum A if the cosmological parameters are known well enough.



The translation of the CMB temperature fluctuations depends on cosmological parameters, e.g. on $\Omega_{\rm m0}$.

11.2.2 Translation to σ_8

- Regarding σ_8, we are not done yet. Inserting the model power spectrum (11.10) into the definition (11.3) gives

$$\sigma_8^2 = \frac{A}{2\pi^2} \left[\frac{k_0^{n+3}}{n+3} + \begin{cases} \frac{k_8^{n-1}}{n-1} - \frac{k_0^{n-1}}{n-1} & (n \neq 1) \\ \ln \frac{k_8}{k_0} & (n = 1) \end{cases} \right] , \tag{11.13}$$

  where $k_8 = 2\pi/(8\,h^{-1}{\rm Mpc})$.

- Since $n \approx 1$, the second term is close to logarithmic and thus weakly dependent on the cosmological parameters in k_0. Then, we see by combining (11.13) with (11.12) that

$$\sigma_8 \propto \Omega_{\rm m0}^{1+n/4} h^{2+n/2} \, \frac{D_+(a)}{a} \; . \tag{11.14}$$

  Note that this is an *approximate* result which is meant to illustrate the principle. It shows that a measurement of the temperature fluctuations in the CMB can only be translated into σ_8 if the matter-density parameter, the Hubble constant, the growth factor and the shape of the power spectrum are accurately known.

- Of course, one could also use the small-scale part of the CMB power spectrum for normalising the dark-matter power spectrum. Due to the acoustic oscillations, however, this part depends in a much more complicated way on additional cosmological parameters, such as the baryon density. Reading σ_8 off the low-order multipoles is thus a safer procedure.

- Even if the cosmological parameters are now known well enough to translate the low-order CMB multipoles to σ_8, an additional uncertainty remains. We know that, although the Universe became neutral $\sim 400,000$ years after the Big Bang, it must have been reionised after the first stars and other sources of UV radiation formed. Since then, CMB photons are travelling through ionised material again and experience Thomson (or Compton) scattering.

- The *optical depth* for Thomson scattering is

$$\tau = \int \mathrm{d}x\, n_\mathrm{e} \sigma_\mathrm{T} \,, \tag{11.15}$$

  where n_e is the number density of free electrons and σ_T is the Thomson scattering cross section. After propagating through the optical depth τ, the CMB fluctuation amplitude is reduced by $\exp(-\tau)$.

- Of course, the CMB photons cannot disappear through Thomson scattering, thus its overall intensity cannot be changed in this way, but the fluctuation amplitudes are lowered in this diffusion process.

- The optical depth τ depends on the path length through ionised material. In view of the CMB, this means that the degree of fluctuation damping depends on the *reionisation redshift*, i.e. the redshift after which the cosmic baryons were transformed back into a plasma. Unless the reionisation redshift is known, we cannot know by how much the CMB fluctuations were suppressed.

- So far, the reionisation redshift can be estimated in two ways. First, as discussed in Sect. 6.2.4, Thomson scattering creates linear polarisation. Of course, the polarisation due to *reionised* material appears superposed on the primordial polarisation, but on different angular scales. The characteristic scale for *secondary* polarisation is the horizon size at the reionisation redshift, which is much larger than the typical scales of the primordial polarisation. Thus, the reionisation redshift can be inferred from large-scale features in the CMB polarisation, provided the cosmological parameters are known well enough to translate angular scales into physical scales.

- Unfortunately, this is aggravated by the polarised microwave radiation from the Milky Way. Synchrotron and dust emission can be substantially polarised and mask the CMB polarisation, which can only be measured reliably if the foregrounds of Galactic origin can be accurately subtracted. Thus, the degree to which the foreground polarisation is known directly determines the accuracy of the σ_8 parameter derived from the CMB fluctuations. This is the main reason for a considerable remaining uncertainty in the σ_8 derived from the 3-year WMAP data given in the table in Sect. 6.2.7.

- The other way to constrain the reionisation redshift uses the spectra of distant quasars. Light with wavelengths shorter than the Lyman-α wavelength cannot propagate through neutral hydrogen because it is immediately absorbed. Therefore, quasar spectra

released before the reionisation redshift must be completely absorbed blueward of the Lyman-α emission line. The appearance of this so-called *Gunn-Peterson effect* at high redshift thus signals the transition from ionised into neutral material. Using this technique, the reionisation redshift was found to be $\sim 6.5 \ldots 7$, which now agrees well with the estimates from the secondary polarisation of the CMB.

11.3 Cosmological weak lensing

- Compared to the outlined procedure to obtain σ_8 from the CMB, it appears completely straightforward to derive it from the cosmic-shear measurements. As we have seen in (10.19), the cosmic-shear power spectrum is proportional to $\Omega_{\rm m0}^2$ times the amplitude A of the dark-matter power spectrum, which leads to the approximate degeneracy $\Omega_{\rm m0}\sigma_8 \approx$ const. between σ_8 and the matter-density parameter $\Omega_{\rm m0}$.

- A more subtle dependence on $\Omega_{\rm m0}$ and to some degree also on other cosmological parameters is introduced by the geometrical weight function $\bar{W}(w', w)$ shown in (10.20), and by the growth of the power spectrum along the line-of-sight. This slightly modifies the form of the σ_8-$\Omega_{\rm m0}$ degeneracy, but does not lift it.

- However, knowing $\Omega_{\rm m0}$ well enough, we should be able to read σ_8 off the cosmic-shear correlation function. However, there are three problems associated with that.

- First, the cosmic shear measured on angular scales below $\sim 10'$ is heavily influenced by the onset of non-linear structure growth and the effect this has on the dark-matter power spectrum. While the linear growth factor can be straightforwardly calculated analytically, non-linear growth can only be quantified by means of large numerical simulations and recipes derived from them. Insufficient knowledge of the non-linear dark-matter power spectrum is a major uncertainty in the cosmological interpretation of cosmic shear.

- Second, the amplitude of cosmological weak-lensing effects depends on the redshift distribution of the sources used for measuring ellipticities. Since these background galaxies are typically very faint, it is demanding to measure their redshifts. Two methods have typically been used. One adapts the known redshift distribution of sources in small, very deep observations such as the *Hubble Deep Field* to the characteristics of the observation to be analysed. The other relies on photometric redshifts, i.e. redshift

estimates based on multi-band photometry. Yet, the precise redshift distribution of the background sources adds additional uncertainty to estimates of σ_8.

- Third, it is possible that systematic effects remain in weak-lensing measurements because the effect is so small, and many corrections have to be applied to measured ellipticities before the cosmic shear can be extracted. Advanced correction methods have been developed which made the *B*-mode contamination almost or completely disappear. This is good news, but it does not yet guarantee the absence of other systematic effects in the data.

- Still, cosmic lensing, combined with estimates of the matter-density parameter, is perhaps the most promising method for precisely determining σ_8. Table 11.1 lists values of σ_8 derived from some cosmic-shear measurements under the assumption of $\Omega_{\rm m0} = 0.3$ in a spatially-flat universe.

σ_8	data	reference
$0.86^{+0.09}_{-0.13}$	RCS	Hoekstra et al. 2002
$0.71^{+0.12}_{-0.16}$	CTIO	Jarvis et al. 2003
0.72 ± 0.09	Combo-17	Brown et al. 2003
0.97 ± 0.13	Keck-II	Bacon et al. 2003
1.02 ± 0.16	HST/STIS	Rhodes et al. 2004
0.83 ± 0.07	Virmos-Descart	van Waerbeke et al. 2005
0.68 ± 0.13	GEMS	Heymans et al. 2005
0.85 ± 0.06	CFHTLS	Hoekstra et al. 2006

Table 11.1: Values for σ_8 derived from cosmic-shear measurements under the assumption of a spatially-flat universe with $\Omega_{\rm m0} = 0.3$.

11.4 Galaxy clusters

11.4.1 The mass function

- Based on the assumption that the density contrast is a Gaussian random field and the spherical-collapse model, Press & Schechter in 1974 derived a mass function for dark-matter halos. It compares the standard deviation σ_R of the density-fluctuation field to the linear density-contrast threshold $\delta_{\rm c} \approx 1.686$ for collapse in the spherical-collapse model. The mean mass contained in spheres of radius R sets the halo mass, which brings the mean (dark-) matter density $\bar{\rho}$ into the game.

- The standard deviation σ_R is related to the power spectrum. For convenience, we introduce an effective slope

$$n = \frac{\mathrm{d}\ln P(k)}{\mathrm{d}\ln k} \tag{11.16}$$

for the power spectrum, which will of course be scale-dependent. On large scales, $n \approx 1$, while $n \to -3$ on small scales, i.e. for small halo masses. For galaxy clusters, $n \approx -1$.

- We introduce the *non-linear* mass scale M_* as the mass contained in spheres of radius R such that $\sigma_R = 1$. Since σ_R grows with the linear growth factor $D_+(a)$, the non-linear mass grows with time. It is convenient here to express the amplitude of the power spectrum, and thus σ_8, in terms of M_*. It is straightforward to show that

$$\sigma_R = \left(\frac{M_*}{M}\right)^{\alpha} , \tag{11.17}$$

with

$$\alpha \equiv \frac{1}{2}\left(1 + \frac{n}{3}\right) . \tag{11.18}$$

- In terms of the dimensionless mass $m \equiv M/M_*$, the Press-Schechter mass function can then be written in the form

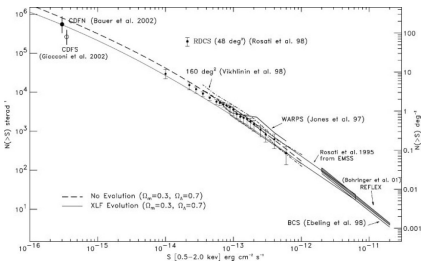
$$N(m,a)\mathrm{d}m = \sqrt{\frac{2}{\pi}}\frac{\bar{\rho}\delta_\mathrm{c}}{M_*^2 D_+(a)}\alpha m^{\alpha-2}\exp\left(-\frac{\delta_\mathrm{c}^2}{2D_+^2(a)}m^{2\alpha}\right)\mathrm{d}m . \tag{11.19}$$

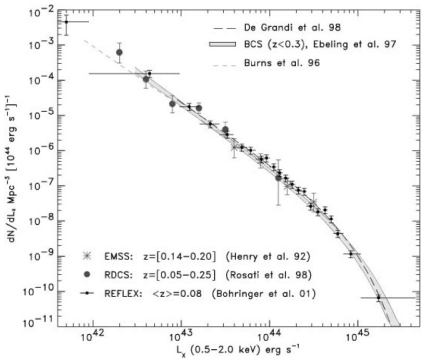
- The Press-Schechter mass function, and some improved variants of it, have been spectacularly confirmed by numerical simulations. It shows that the mass function is a power law with an exponential cut-off near the non-linear mass scale M_*. For galaxy clusters, $n \approx -1$, thus $\alpha \approx 1/3$, and

$$N(m,a)\mathrm{d}m \propto m^{-5/3}\exp\left(-\frac{\delta_\mathrm{c}^2}{2D_+^2(a)}m^{2/3}\right)\mathrm{d}m , \tag{11.20}$$

with an amplitude characterised by M_*, the mean dark-matter density $\bar{\rho}$, and the growth factor $D_+(a)$.

- This opens a way to constrain cosmological parameters as well as σ_8 with galaxy clusters: if the abundance and evolution of the cluster mass function can be measured, they can be determined from the mass scale of the exponential cut-off and the amplitude of the power-law end. Today, the non-linear mass scale is a few times $10^{13}\,M_\odot$. Therefore, the exponential cut-off in the halo mass will not be seen in the galaxy mass function. Clusters, however, show the exponential cut-off very well, and thus their population is very sensitive to changes in σ_8. In principle, therefore, σ_8 should be very well constrained by the cluster population.





The X-ray flux (top) or luminosity functions of galaxy clusters can be converted to a mass function *if* it is possible to measure cluster masses sufficiently accurately.

11.4.2 What is a cluster's mass?

- The main problem here is how observable cluster properties should be related to quantities used in theory. Strictly speaking, the cluster *mass*, as used in the theoretical mass function (11.20), is *not* an observable. Global cluster observables are the X-ray temperature and flux, the optical luminosity and the velocity distribution of their galaxies, and their gravitational-lensing effects. Before we discuss their relation to mass, let us first see what the "mass of a galaxy cluster" could be.

- It is easy to define masses of gravitationally bound, well localised objects, such as planets or stars. They have a well-defined boundary, e.g. the planetary surfaces or the stellar photospheres. This is markedly different for objects like galaxies and galaxy clusters. As far as we know, their densities drop smoothly towards zero like power laws, $\propto r^{-(2...3)}$. Thus, although they are gravitationally bound, it is less obvious what should be seen as their outer boundary. Strictly speaking, there is none.

- The only way out is then to *define* an outer boundary in such a way that it is well-defined in theory and identifiable in observational data. A common choice was introduced in Sect. 5.1.2: it defines the boundary by the mean overdensity it encloses. Although this is problematic as well, it may be as good as it gets. Three immdiately obvious problems created by this definition are that objects like galaxy clusters are often irregularly shaped rather than spherical, that the overdensity of 200 is quite arbitrary, even if it is inspired by virial equilibrium in the spherical-collapse model, and that its measurement requires a sufficiently accurate density profile to be known or assumed.

- How could standardised radii such as R_{200} be measured? This could for instance be achieved applying equations such as (6.33) after measuring the slope β and the core radius of the X-ray surface brightness profile together with the X-ray temperature, by calibrating an *assumed* density profile with galaxy kinematics based on the virial theorem, or by constraining the cluster mass profile with gravitational lensing.

- Obviously, all these measurements have their own problems. Being sensitive to all mass along the line-of-sight, gravitational lensing cannot distinguish between mass bound to a cluster or just projected onto it. Any measurement based on the virial theorem must of course rely on virial equilibrium, which takes time to be established and is often perturbed in real clusters because of merging and accretion. The common interpretation of X-ray measurements requires the assumption that the X-ray gas be in

hydrostatic equilibrium with the host cluster's gravitational potential.

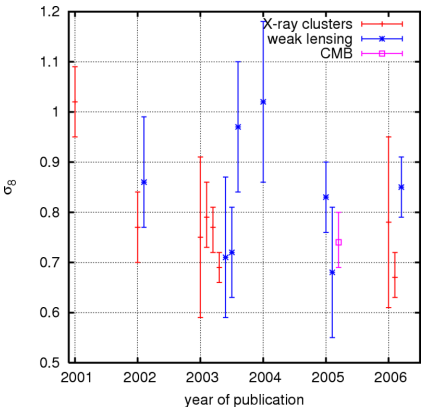
- This illustrates that it may be fair to say that *there is no such thing as the mass of a galaxy cluster*. Even if measurements of cluster "radii" were less dubious, it remained unclear whether they mean the same as those assumed in theory, which are related to the spherical-collapse model. Interestingly, but not surprisingly, cluster masses obtained from numerical simulations suffer from the same poor definition of the concept of a "cluster radius".

- How can we make progress then? Observables such as the cluster temperature T_X or its X-ray luminosity L_X should be related to the depth of the gravitational-potential well they are embedded in, which should in turn be related to some measure of the total mass. If we can calibrate such expected temperature-mass or luminosity-mass relations, e.g. using numerical simulations of galaxy clusters, a direct comparison between theory and observations seems possible. This is sometimes called an *external calibration* of the required relations.

- *Internal calibrations*, i.e. calibrations based on cluster data alone, have become increasingly fashionable over the past years. Here, empirical temperature-mass and luminosity-mass relations are obtained based on one or more estimates of the mass estimates sketched above.

σ_8	data	reference
1.02 ± 0.07	M-T relation	Pierpaoli et al. 2001
0.77 ± 0.07	M-T relation	Seljak 2002
0.75 ± 0.16	lensing masses	Smith et al. 2003
$0.79^{+0.06}_{-0.07}$	luminosity function	Pierpaoli et al. 2003
$0.77^{+0.05}_{-0.04}$	temperature function	Pierpaoli et al. 2003
0.69 ± 0.03	lensing masses	Allen et al. 2003
0.78 ± 0.17	optical richness	Eke et al. 2006
$0.67^{+0.04}_{-0.05}$	lensing masses	Dahle 2006

Table 11.2: Values of σ_8 derived from the galaxy-cluster population based on different observational data.



Several recent determinations of σ_8.

- The result of both procedures is qualitatively the same. It allows the conversion of observables to mass, and thus of the observed cluster temperature or luminosity functions to mass functions, which can then compared to theory. The shape and amplitude of the power spectrum and the growth factor can then be adapted to optimise the agreement between observed and expected mass

functions. Clusters at moderate or high redshift constrain the evolution of the mass function and allow an independent estimate of the matter-density parameter Ω_{m0}, as sketched in Sect. 5.3 before.

- In view of the many difficulties listed, it is an astonishing fact that, when applied not to cluster samples rather than individual clusters, the determination of the cluster mass function and its evolution seems to work very well. Values for σ_8 derived therefrom are given in Tab. 11.2.