



MPHys Advanced Cosmology 2010–2011

John Peacock and Catherine Heymans

Room C20, Royal Observatory; jap@roe.ac.uk

<http://www.roe.ac.uk/japwww/teaching/cos5.html>

Synopsis This course is intended to act as an extension of the current 4th-year course on Astrophysical Cosmology, which develops the basic tools for dealing with observations in an expanding universe, and gives an overview of some of the central topics in contemporary research. The aim here is to revisit this material at a level of detail more suitable as a foundation for understanding current research. Cosmology has a standard model for understanding the universe, in which the dominant theme is the energy density of the vacuum. This is observed to be non-zero today, and is hypothesised to have been much larger in the past, causing the phenomenon of 'inflation'. An inflationary phase can not only launch the expanding universe, but can also seed irregularities that subsequently grow under gravity to create galaxies, superclusters and anisotropies in the microwave background. The course will present the methods for analysing these phenomena, leading on to some of the frontier issues in cosmology, particularly the possible existence of extra dimensions and many universes. It is intended that the course should be self contained; previous attendance at courses on cosmology or general relativity will be useful, but not essential.

Recommended books (in reserve section of ROE library)

Peacock: Cosmological Physics (CUP) Gives an overview of cosmology at the level of this course, but contains much more than will be covered here. More recent developments to be covered in the lectures are not in the book.

1

Dodelson: Modern Cosmology (Wiley) Concentrating on the details of relativistic perturbation theory, with applications to the CMB. Higher level than this course, but contains many useful things.

Other good books for alternative perspectives and extra detail:

Mukhanov: Physical Foundations of Cosmology (CUP)
Peebles: Principles of Physical Cosmology (Princeton)
Weinberg: Gravitation & Cosmology (Wiley)

- (1) **Review of Friedmann models** FRW spacetime; Dynamics; Observables; Horizons
- (2) **The hot big bang** Thermal history; Freezeout; Relics; Recombination and last scattering
- (3) **Inflation – I** Initial condition problems; Planck era; Physics beyond the SM; Scalar fields; Noether's theorem
- (4) **Inflation – II** The zoo of inflation models; Equation of motion; Slow-roll; Ending inflation
- (5) **Fluctuations from inflation** Gauge issues; Power spectra; Basics of fluctuation generation; Tilt; Tensor modes; Eternal inflation
- (6) **Structure formation – I** Newtonian analysis neglecting pressure; Perturbation modes; Coupled perturbations; matter transfer functions
- (7) **Structure formation – II** Nonlinear development; Spherical model; Lagrangian approach; N-body simulations; Dark-matter haloes & mass function; Gas cooling; Brief overview of galaxy formation
- (8) **Gravitational lensing** Basics of light deflection; strong lensing and mass measurement; weak lensing and mapping dark matter
- (9) **CMB anisotropies – I** Anisotropy mechanisms; Overview of Boltzmann approach; Power spectrum; Properties of the temperature field
- (10) **CMB anisotropies - II** Geometrical degeneracies; Reionization; Polarization and tensor modes; The cosmological standard model
- (11) **Frontiers** Measuring dark energy; Extra dimensions and modified gravity; anthropics and the multiverse

1 Review of Friedmann models

Topics to be covered:

- Cosmological spacetime and RW metric
- Expansion dynamics and Friedmann equation
- Calculating distances and times

1.1 Cosmological spacetime

One of the fundamentals of a cosmologist's toolkit is to be able to assign coordinates to events in the universe. We need a large-scale notion of space and time that allows us to relate observations we make here and now to physical conditions at some location that is distant in time and space. The starting point is the relativistic idea that spacetime must have a **metric**: the equivalence principle says that conditions around our distant object will be as in special relativity (if it is freely falling), so there will be the usual idea of the **interval** or **proper time** between events, which we want to rewrite in terms of our coordinates:

$$(1) \quad -ds^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2 = g^{\mu\nu} dx_\mu dx_\nu.$$

Here, dashed coordinates are local to the object, undashed are the global coordinates we use. As usual, the Greek indices run from 0 to 3. Note the ambiguity in defining the sign of the squared interval. The matrix $g^{\mu\nu}$ is the **metric tensor**, which is found in principle by solving Einstein's gravitational field equations. A simpler alternative, which fortunately matches the observed universe pretty well, is to consider the most symmetric possibilities for the metric.

ISOTROPIC EXPANSION Again according to Einstein, any spacetime with non-zero matter content must have some spacetime curvature, i.e. the metric cannot have the special relativity form $\text{diag}(+1, -1, -1, -1)$. This curvature is something intrinsic to the spacetime, and does not need to be associated with extra spatial dimensions; these are nevertheless a useful intuitive way of understanding curved spaces such as the 2D surface of a 3D sphere. To motivate what is to come, consider the higher-dimensional analog of this surface: something that is almost a 4D (hyper)sphere in Euclidean 5D space:

$$(2) \quad x^2 + y^2 + z^2 + w^2 - v^2 = \mathcal{R}^2$$

$$(11) \quad t_H \equiv H^{-1} = 9.78 \text{ Gyr} \times (H/100 \text{ km s}^{-1} \text{ Mpc}^{-1})^{-1}.$$

The characteristic time of the expansion is called the **Hubble time**, and takes the value

$$(10) \quad H(t) = \dot{R}(t)/R(t).$$

or a velocity proportional to distance, independent of origin, with

$$(9) \quad \dot{\mathbf{x}}(t) = \dot{R}(t)\mathbf{x}(t_0) = R(t)/R(t_0) \mathbf{x}(t),$$

where $R(t)$ is the **scale factor**. Differentiating this with respect to t gives

$$(8) \quad \dot{\mathbf{x}}(t) = \dot{R}(t)\mathbf{x}(t_0),$$

t_0 :

where, locally, all position vectors at time t are just scaled versions of their values at a reference time expanding spacetime, and we need to make the idea general. What we are interested in is a situation

THE ROBERTSON-WALKER METRIC The de Sitter model is only one example of an isotropically expanding spacetime, and we need to make the idea general. What we are interested in is a situation where, locally, all position vectors at time t are just scaled versions of their values at a reference time more sense to use an expanding coordinate system defined by the locations of such a set of particles. It becomes clear that a static distribution of test particles is impossible in general, so that it makes a local (and thus Newtonian, thanks to the equivalence principle) look at particle dynamics. Then expanding. That this is the correct way to think about things only becomes apparent when we take In short, it takes more than just the appearance of $R(t)$ in a metric to prove that something is

(although we now know that his sample was too shallow to be able to detect it reliably). the expanding universe was explicitly motivated by the possibility of finding the 'de Sitter effect' Sipher had already established that most galaxies were redshifted. Hubble's 1929 'discovery' of this was interpreted as measuring the constant radius of curvature of spacetime, \mathcal{R} . By this time, that one would expect a redshift that increased linearly with distance in de Sitter's model, but immediately clear that the model is non-static. It was eventually concluded (in 1923, by Weyl) that one would expect a redshift that increased linearly with distance in de Sitter's model, but Historically, de Sitter space was extremely important in cosmology, although it was not

origin, and must thus suffer a Doppler redshift. This of course presumes that constant r corresponds to the actual trajectory of a free particle, which we have not proved – although it is true.

5

There is a universal **cosmological time**, which is the ticking of clocks at constant **comoving radius** r and constant angle on the sky. The spatial part of the metric expands with time, according to a universal **scale factor** $R(t) = \mathcal{R} \cosh(ct/\mathcal{R})$, so that particles at constant r recede from the

$$(7) \quad (\text{interval})^2 = (\text{time interval})^2 - (\text{scale factor})^2 (\text{comoving interval})^2.$$

Now we have a completely different interpretation of the metric:

$$(6) \quad c^2 dt^2 = c^2 dt^2 - \mathcal{R}^2 \cosh^2(ct/\mathcal{R}) (dr^2 + \sin^2 r [d\theta^2 + \sin^2 \theta d\phi^2]).$$

which by an obvious change of notation becomes

$$(5) \quad ds^2 = -\mathcal{R}^2 da^2 + \mathcal{R}^2 \cosh^2 a (\beta^2 + \sin^2 \beta) [d\gamma^2 + \sin^2 \gamma d\delta^2],$$

$|\mathbf{e}_a|^2 da^2$, which makes the metric into continuing the process to still more dimensions). The squared length of the vector is just the sum of $\partial(x, y, z, w, a)/\partial a$ is orthogonal to all the other \mathbf{e}_i (most simply seen by considering \mathbf{e}_β and imagining This has the advantage that it is an orthogonal coordinate system: a vector such as $\mathbf{e}_a =$

$$(4) \quad \begin{aligned} v &= \mathcal{R} \sinh a \\ w &= \mathcal{R} \cosh a \cos \beta \\ z &= \mathcal{R} \cosh a \sin \beta \cos \gamma \\ y &= \mathcal{R} \cosh a \sin \beta \sin \gamma \cos \delta \\ x &= \mathcal{R} \cosh a \sin \beta \sin \gamma \sin \delta. \end{aligned}$$

choose. Suppose we re-express things using the analogues of polar coordinates: This maximally symmetric spacetime is known as **de Sitter space**. It looks like a static spacetime, but relativity can be deceptive, as the interpretation depends on the coordinates you

Effectively, we have made one coordinate imaginary because we know we want to end up with the 4D spacetime signature.

$$(3) \quad ds^2 = dx^2 + dy^2 + dz^2 + dw^2 - dv^2.$$

where the metric is

$$\ddot{R} = -4\pi G R(\rho + 3p/c^2)/3. \quad (18)$$

This energy-like equation can be turned into a force-like equation by differentiating with respect to time:

This is almost obviously true, since the Newtonian result that the gravitational field inside a uniform shell is zero does still hold in general relativity, and is known as **Birkhoff's theorem**. For the present course, we will accept this quasi-Newtonian 'derivation', and merely attempt to justify the form of the rhs.

$$R\ddot{R} - \frac{8\pi G}{3}\rho R^2 = -kc^2. \quad (17)$$

THE FRIEDMANN EQUATION The equation of motion for the scale factor resembles Newtonian conservation of energy for a particle at the edge of a uniform sphere of radius R :

1.2 Cosmological dynamics

In terms of the normalized scale factor $a(t)$ we have simply $a(t) = (1+z)^{-1}$. So just by observing shifts in spectral lines, we can learn how big the universe was at the time the light was emitted. This is the key to performing observational cosmology.

$$\frac{t_{\text{obs}}}{R(t_{\text{obs}})} \equiv 1+z = \frac{t_{\text{emit}}}{R(t_{\text{emit}})}. \quad (16)$$

The comoving distance is constant, whereas the domain of integration in time extends from t_{emit} to t_{obs} ; these are the times of emission and reception of a photon. Thus $dt_{\text{emit}}/dt_{\text{obs}} = R(t_{\text{emit}})/R(t_{\text{obs}})$, which means that events on distant galaxies time-dilate. This dilation also applies to frequency, so

$$r = \int c dt/R(t). \quad (15)$$

LIGHT PROPAGATION AND REDSHIFT Light follows trajectories with zero proper time (**null geodesics**). The radial equation of motion therefore integrates to

7

$$a(t) \equiv \frac{R_0}{R(t)}, \quad (14)$$

so that $a = 1$ at the present.

The comoving radius r is dimensionless, and the scale factor R really is the spatial radius of curvature of the universe. Both are required in order to give a comoving distance dimensions of length – e.g. the combination $R_0 S_k(r)$. Nevertheless, it is often convenient to make the scale factor dimensionless, via

We only saw the $k = +1$ case of this in the de Sitter example, but mathematically we can then generate the $k = -1$ case by letting R and r both become imaginary.

$$S_k(r) \equiv \begin{cases} \sin r & (k = +1) \\ \sinh r & (k = -1) \\ r & (k = 0). \end{cases} \quad (13)$$

The angle $d\psi$ separates two points on the sky, so that $d\psi^2 = d\theta^2 + \sin^2\theta d\phi^2$ in spherical polars. The function $S_k(r)$ allows for positive and negative curvature of the comoving part of the metric:

$$c^2 d\tau^2 = c^2 dt^2 - R^2(t) [d\tau^2 + S_k^2(r) d\psi^2]. \quad (12)$$

By analogy with the de Sitter result, we now guess that the spatial metric will factorize into the scale factor times a comoving part that includes curvature. This overall Robertson–Walker metric (**RW metric**), can be written as:

As with de Sitter space, we assume a **cosmological time** t , which is the time measured by the clocks of these observers – i.e. t is the proper time measured by an observer at rest with respect to the local matter distribution. It makes sense that such a universal time exists if we accept that we are looking for models that are **homogeneous**, so that there are no preferred locations. This is obvious in de Sitter space: because it derives from a 4-sphere, all spacetime points are manifestly equivalent: the spacetime curvature and hence the matter density must be a constant. The next step is to weaken this so that conditions can change with time, but are uniform at a given time. A cosmological time coordinate can then be defined and synchronized by setting clocks to a reference value at some standard density.

$$(29) \quad d \propto a^{-3(m+1)}$$

so

$$(28) \quad \Lambda^2 d^2 c^2 \Lambda^2 = -d^2 \Lambda^2 d^2 c^2 \Lambda^2 = -w d^2 c^2 \Lambda^2 d^2 \ln a = -3(w+1),$$

then conservation of energy says

$$(27) \quad w \equiv d/d^2 c^2,$$

How this varies with redshift depends on its equation of state. If we define the parameter and wonder if the universe might contain another form of matter that we have not yet considered, expansion, with photons also having their energy reduced by the redshift. We can be more general, as a^{-3} and a^{-4} . The first two relations just say that the number density of particles is diluted by the nonrelativistic matter, and radiation-dominated matter. These have densities that scale respectively to specify the matter content of the universe, and there are two obvious candidates: pressureless models with general equations of state. To solve the Friedmann equation, we need

$$(26) \quad \rho_0 = 1.878 \times 10^{-26} \Omega h^2 \text{ kg m}^{-3} = 2.775 \times 10^{11} \Omega h^2 M_{\odot} \text{Mpc}^{-3},$$

in terms of which the current density of the universe is

$$(25) \quad h \equiv \frac{H_0}{100 \text{ km s}^{-1} \text{Mpc}^{-1}},$$

dimensionless (current) Hubble parameter as symbol in cosmological formulae, that it is normal to omit the zero subscript. We can also define a which diverges as the universe approaches the flat state with $\Omega = 1$. In practice, Ω_0 is such a common

$$(24) \quad R_0 = \frac{H_0}{c} [k/(\Omega_0 - 1)]^{1/2},$$

The current value of such parameters should be distinguished by a zero subscript. In these terms, the Friedmann equation gives the present value of the scale factor:

9

$$(23) \quad \Omega \equiv \frac{\rho}{\rho_c} = \frac{8\pi G \rho}{3H^2}.$$

density:

It is common to define a dimensionless **density parameter** as the ratio of density to critical

$$(22) \quad \rho_c = \frac{3H^2}{8\pi G}.$$

There is thus a **critical density** that will yield a flat universe,

we have in effect the $k = 0$ case. $R \rightarrow \infty$ with Rr fixed implies $r \rightarrow 0$, so the difference between $S_k(r)$ and r becomes negligible and the rhs of the Friedmann equation indistinguishable from zero. Looking at the metric with $k \neq 0$, Now imagine holding the observables H and ρ constant, but let $R \rightarrow \infty$: this has the effect of making

$$(21) \quad H^2 - \frac{8\pi G}{\text{const}} \rho = \frac{3}{R^2}.$$

through by R^2 :

is readily understood. Write the energy-conservation equation with an arbitrary rhs, but divide metric. This cannot be completely justified without the Field Equations, but the **flat** $k = 0$ case related to the curvature of spacetime, and $k = 0, \pm 1$ is the same integer that is found in the RW But the greatest astonishment in the Friedmann equation is the term on the rhs. This is

the extra term from the pressure is important. As an example, consider a **radiation-dominated fluid** $-i.e.$ one whose equation of state is the same as that of pure radiation: $p = u/3$, where u is the energy density. For such a fluid, $\rho + 3p/c^2 = 2\rho$, so its gravity is twice as strong as we might have expected.

$$(20) \quad \Delta^2 \Phi = 4\pi G(\rho + 3p/c^2).$$

The surprising factor here is the occurrence of the **active mass density** $\rho + 3p/c^2$. This is here because the weak-field form of Einstein's gravitational field equations is

$$(19) \quad d[\rho^2 R^3] = -pd[R^3].$$

To deduce this, we need to know p , which comes from conservation of energy:

CURVED MODELS We will not be very strongly concerned with highly curved models in this course, but it is worth knowing some basic facts, as shown in figure 1 (neglecting radiation). On a plot of the $\Omega_m + \Omega_v = 1$ always separates open and closed models. If $\Omega_v > 0$, recollapse always occurs – whereas a positive vacuum density does not always guarantee expansion to infinity, especially when the matter density is high. For closed models with sufficiently high vacuum density, there was no big bang in the past, and the universe must have emerged from a ‘bounce’ at some finite minimum radius. All these statements can be deduced quite simply from the Friedmann equation.

This transition from the flat matter-dominated $a \propto t^{2/3}$ to de Sitter space seems to be the one that describes our actual universe (apart from the radiation era at $z \gtrsim 10^4$).

$$(35) \quad a = \left(\sinh(3\tau/2\sqrt{2}) \right)^{2/3}.$$

MATTER AND VACUUM Here, $a' = (a^{-1} + a^2)/2$, which can be tackled via the substitution $y = a^{3/2}$, to yield

Here, we move from $a \propto t^{1/2}$ at early times to an exponential behaviour characteristic of vacuum-dominated **de Sitter space**. This would be an appropriate model for the onset of a phase of inflation following a big-bang singularity.

$$(34) \quad a = \left(\sinh(\sqrt{2}\tau) \right)^{1/2}.$$

RADIATION AND VACUUM Now we have $a' = (a^{-2} + a^2)/2$, which is easily solved in the form $(a^2)/\sqrt{2} = \sqrt{1 + (a^2)^2}$, and simply inverted:

Both these powers are shallower than t , reflecting the decelerating nature of the expansion. so the universe expands as $t^{1/2}$ in the radiation era, which becomes $t^{2/3}$ once matter dominates.

$$(33) \quad \begin{aligned} \tau \gg 1 : a &= (\sqrt{2}\tau)^{1/2}, \\ \tau \ll 1 : a &= 1 : \tau = 3\tau/2\sqrt{2} \end{aligned}$$

This can be inverted to yield $a(\tau)$, but the full expression is too ugly to be much use. It will suffice to note the limits:

11

$$(32) \quad \tau = \frac{3}{2\sqrt{2}} \left(2 + (a^{-2} + a^2)\sqrt{1 + a} \right).$$

MATTER AND RADIATION Using dashes to denote $d/d(t/\tau)$, we have $a'^2 = (a^{-2} + a^{-1})/2$, which is simply integrated to yield

dimensionless version $\tau \equiv tH^*$. the Hubble parameter at that point (H^*) sets a characteristic time, from which we can make a $a = 1$ to be at a convenient point where the densities of two main components cross over. Also, look simplest if we appreciate that normalization to the current era is arbitrary, so we can choose postulates that the vacuum density was very much higher in the very distant past). The solutions be neglected at sufficiently early times, as can vacuum density (except that the theory of inflation worth knowing. Mostly these refer to the **flat universe** with total $\Omega = 1$. Curvature can always analytically in all cases, nor can we always invert to get $a(t)$, but there are some useful special cases This equation is in a form that can be integrated immediately to get $t(a)$. This is not possible

$$(31) \quad H^2(a) = \left[\Omega_v H_0^2 + \Omega_m a^{-3} + \Omega_r a^{-4} - (1 - \Omega) a^{-2} \right].$$

the form $H^2 = 8\pi G\rho/3 - kc^2/R^2$. Inserting the expression for $\rho(a)$ gives Friedmann equation in a manner useful for practical solution. Start with the Friedmann equation in except where we want to focus explicitly on this parameter. This expression allows us to write the (using the normalized scale factor $a = R/R_0$). We will generally set $w = -1$ without comment,

$$(30) \quad \frac{8\pi G\rho}{3} = \Omega_v H_0^2 a^{-3(w+1)} + \Omega_m a^{-3} + \Omega_r a^{-4}$$

In terms of observables, this means that the density is written as

But this may not be an exhaustive list, and the universe could contain substances with less familiar equations of state. Inventing new forms of matter may seem like a silly game to play, but cosmology can be the only way to learn if something unexpected exists. As we will see in more detail later, modern data force us to accept a contribution that is approximately independent of time with $w \simeq -1$: a **vacuum energy** that is simply an invariant property of empty space. A general name for this contribution is **dark energy**, reflecting our ignorance of its nature (although the name is not very good, since it is too similar to dark matter: ‘dark tension’ would better reflect its unusual equation of state with negative pressure).

If w is constant. Pressureless nonrelativistic matter has $w = 0$ and radiation has $w = 1/3$.

At high redshift, where H increases, this tends to zero. The onset of radiation domination does not change this: even though the presently visible universe was once very small, it expanded so quickly

$$(41) \quad R_{0r} \approx \frac{H_0}{2c} \Omega_m^{-0.4}.$$

matter-dominated models, full distance a particle can have travelled since the big bang – the **horizon distance**. For flat the integral for comoving radius is taken from $z = 0$ to ∞ , we get the

Remember that non-flat models need the combination $R_0 S_k(r)$, so one has to divide the above integral by $R_0 = (c/H_0) |\Omega - 1|^{-1/2}$, apply the S_k function, and then multiply by R_0 again. Once more, this process is not analytic in general.

$$(40) \quad R_{0r} = \int \frac{H(z)}{c} dz.$$

DISTANCE-REDSHIFT RELATION The equation of motion for a photon is $R dr = c dt$, so $R_0 dr/dz = (1+z)c dt/dz$, or

For a flat universe, the current age is $H_0 t_0 \approx (2/3) \Omega_m^{-0.3}$. For many years, estimates of this product were around unity, which is hard to understand without vacuum energy; unless the density is very low ($H_0 t_0$ is exactly 1 in the limit of an empty universe). This was one of the first astronomical motivations for a vacuum-dominated universe.

$$(39) \quad t \approx (2/3) H^{-1} \Omega_m^{-0.3} (1+z)^{-3/2}.$$

At $10 < z < 1000$, where matter dominates, this is

$$(38) \quad H(z)t(z) \approx \frac{3}{2} (0.7 \Omega_m(z) - 0.3 \Omega_v(z) + 0.3)^{-0.3}.$$

This can't be done analytically in general, but the following simple approximate formula is accurate to a few % for cases of practical interest:

$$(37) \quad H_2(a) = \Omega_v a^{-3} + \Omega_m a^{-3} + \Omega_r a^{-4} - (1-a^{-2}).$$

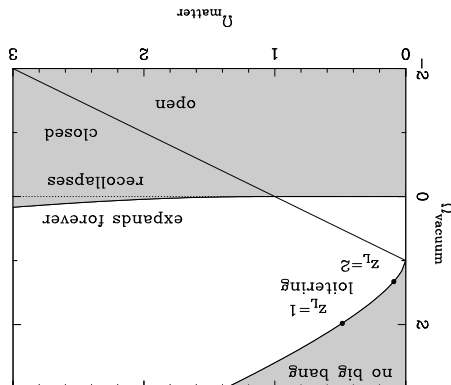
so $t(z) = \int_{\infty}^z H^{-1}(z) dz / (1+z)$, where

$$(36) \quad \frac{dz}{R_0} = -\frac{R_0}{R} \frac{dR}{R} = -(1+z) H(z),$$

AGE OF THE UNIVERSE Since $1+z = R_0/R(z)$, we have

1.3 Observational cosmology

Figure 1. This plot shows the different possibilities for the cosmological expansion as a function of matter density and vacuum energy. Models with total $\Omega < 1$ are always spatially closed (open for $\Omega > 1$), although closed models can still expand to infinity if $\Omega_v \neq 0$. If the cosmological constant is negative, recollapse always occurs; recollapse is also possible with a positive Ω_v if $\Omega_m \gg \Omega_v$. If $\Omega_v < 1$ and Ω_m is small, there is the possibility of a 'loitering' solution with some maximum redshift and infinite age (top left); for even larger values of vacuum energy, there is no big bang singularity.



$$dN = g \frac{V^{2\pi^3}}{V} d^3k \quad (47)$$

Although the timescale for expansion of the early universe is very short, the density is also very high, so it is normally sensible to assume that conditions are close to thermal equilibrium. Also the fluids of interest are simple enough that we can treat them as perfect gases. The thermodynamics of such a gas is derived starting with a box of volume $V = L^3$, and expanding the fields inside into periodic waves with **harmonic boundary conditions**. The density of states in k space is

2.1 Thermal history

- Thermal history
- Freezeout & relics
- Recombination and last scattering

Topics to be covered:

2 The hot big bang

Some example distance-redshift relations are shown in figure 2. Notice how a high matter density tends to make high-redshift objects brighter: stronger deceleration means they are closer for a given redshift.

$$\text{angular - diameter distance : } D_A = (1+z)^{-1} R_0 S_k(r) \quad \text{luminosity distance : } D_L = (1+z) R_0 S_k(r). \quad (46)$$

EFFECTIVE DISTANCES The angle and flux relations can be made to look Euclidean:

where B_ν is **surface brightness** (luminosity emitted into unit solid angle per unit area of source). This $(1+z)^3$ dimming makes it hard to detect extended objects at very high redshift. The factor becomes $(1+z)^4$ if we integrate over frequency to get a bolometric quantity.

15

$$I_\nu(\nu_0) = \frac{B_\nu([1+z]\nu_0)}{(1+z)^3}, \quad (45)$$

SURFACE BRIGHTNESS The flux density is the product of the **specific intensity** I_ν and the solid angle subtended by the source: $S_\nu = I_\nu d\Omega$. Combining the angular size and flux-density relations gives a relation that is independent of cosmology:

$$S_\nu(\nu_0) = \frac{L_\nu([1+z]\nu_0)}{L_\nu(\nu_0)} = \frac{4\pi R_0^2 S_k^2(r)(1+z)}{4\pi R_0^2 S_k^2(r)(1+z)^{1+\alpha}}, \quad (44)$$

where the second expression assumes a power-law spectrum $L \propto \nu^{-\alpha}$.

LUMINOSITY AND FLUX DENSITY Imagine a source at the centre of a sphere, on which we sit. The photons from the source pass through a proper surface area $4\pi[R_0 S_k(r)]^2$. But redshift still affects the flux density in four further ways: (1) photon energies are redshifted, reducing the flux density by a factor $1+z$; (2) photon arrival rates are time dilated, reducing the flux density by a further factor $1+z$; (3) opposing this, the bandwidth $d\nu$ is reduced by a factor $1+z$, which increases the energy flux per unit bandwidth by one power of $1+z$; (4) finally, the observed photons at frequency ν_0 were emitted at frequency $[1+z]\nu_0$. Overall, the flux density is the luminosity at frequency $[1+z]\nu_0$, divided by the total area, divided by $1+z$:

If we know r , we can therefore convert the angle subtended by an object into its physical extent perpendicular to the line of sight.

$$d\ell_\perp = d\psi R(z) S_k(r) = d\psi R_0 S_k(r) / (1+z). \quad (43)$$

The spatial parts of the metric give the *proper* transverse size of an object seen by us as its comoving size $d\psi S_k(r)$ times the scale factor at the time of emission:

$$c^2 dt^2 = c^2 dr^2 - R^2(t) [dr^2 + S_k^2(r) d\psi^2]. \quad (42)$$

ANGULAR DIAMETERS Recall the RW metric:

that causal contact was not easy. The observed large-scale near-homogeneity is therefore something of a puzzle.

The thermodynamic properties of Fermions can be obtained from those of Bosonic black-body radiation by the following trick: $1/(e^x + 1) = 1/(e^x - 1) - 2/(e^{2x} - 1)$. Thus, a gas of fermions looks like a mixture of bosons at two different temperatures. Knowing that boson number density and energy scale as $n \propto L^{-3}$ and $u \propto L^{-4}$, we find $n^{\text{F}} = (3/4)n^{\text{B}}$; $u^{\text{F}} = (3/8)u^{\text{B}}$.

$$n = \frac{\pi^2}{30} g(kT)^3 \quad (\text{bosons}). \quad (52)$$

In the ultrarelativistic limit, $\epsilon(d) = pc$, this becomes

$$n = \int_0^\infty \frac{4\pi p^2 dp}{1} g \frac{(2\pi\hbar)^3}{1} e^{\epsilon(d)/kT} \epsilon(d). \quad (51)$$

The above thermodynamics also gives the energy density of the background, since it is only necessary to multiply the integrand by a factor $\epsilon(d)$ for the energy in each mode:

(2) Non-relativistic limit. Here we can neglect the ± 1 in the occupation number, in which case the number is suppressed by a dominant $\exp(-mc^2/kT)$ factor. This shows us that the background switches on at about $kT \sim mc^2$; at this energy, known as a **threshold**, photons and other species in equilibrium will have sufficient energy to create particle-antiparticle pairs.

$$n = \left(\frac{c}{kT}\right)^3 \int_0^\infty \frac{4\pi p^2 dp}{y^2} g \frac{(2\pi\hbar)^3}{1} e^{\epsilon(d)/kT}. \quad (50)$$

(1) Ultrarelativistic limit. For $kT \gg mc^2$ the particles behave as if they were massless, and we get

where we have changed to momentum space; $\epsilon = \sqrt{m^2c^4 + p^2c^2}$ and g is the degeneracy factor. There are two interesting limits of this expression.

$$n = \frac{1}{V} \int f dN = \int_0^\infty \frac{4\pi p^2 dp}{1} g \frac{(2\pi\hbar)^3}{1} e^{\epsilon(d)/kT}. \quad (49)$$

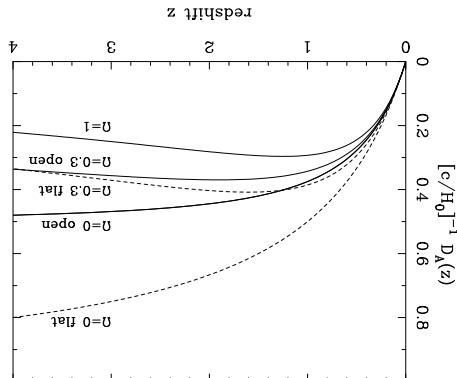
as N adjusts to its equilibrium value, we expect that the system will be stationary with respect to small changes in N . The thermal equilibrium **background number density** of particles is

(+ for fermions, - for bosons). Now, for a thermal radiation background, the **chemical potential**, μ is always zero. The reason for this is quite simple: μ appears in the first law of thermodynamics as the change in energy associated with a change in particle number, $dE = TdS - PdV + \mu dN$. So,

$$\langle f \rangle = \left[e^{\epsilon - \mu/kT} \pm 1 \right]^{-1} \quad (48)$$

(where g is a degeneracy factor for spin *etc.*). The equilibrium **occupation number** for a quantum state of energy ϵ is given generally by

Figure 2. A plot of dimensionless angular-diameter distance versus redshift for various cosmologies. Solid lines show models with zero vacuum energy; dashed lines show flat models with $\Omega_m + \Omega_v = 1$. In both cases, results for $\Omega_m = 1, 0.3, 0$ are shown; higher density results in lower distance at high z , due to gravitational focusing of light rays.



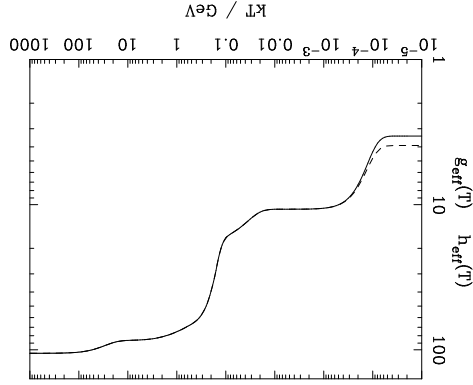
$$T = 2.725 \pm 0.002 \text{ K.} \quad (58)$$

This is independent of the present-day temperature of the photon background, which manifests itself as the **cosmic microwave background (CMB)**,

$$t/\text{seconds} = g_{\text{eff}}^{-1/2} (T/10^{10.26} \text{ K})^{-2}. \quad (57)$$

dominated universe, $t = (32\pi G\rho/3)^{-1/2}$, we can deduce the time-temperature relation—
 TIME AND TEMPERATURE. This temperature-dependent equilibrium density sets the timescale
 for expansion in the early universe. Using the relation between time and density for a flat radiation—

Figure 3. The number of relativistic degrees of freedom as a function of photon temperature. g_{eff} measures the energy density; h_{eff} the entropy (dashed line). The two depart significantly at low temperatures, when the neutrinos are cooler than the photons. For a universe consisting only of photons, we would expect $g = 2$. The main features visible are (1) The electroweak phase transition at 100 GeV; (2) The QCD phase transition at 200 MeV; (3) the e^{\pm} annihilation at 0.3 MeV.



nonrelativistic. model of particle physics), and fall as the temperature drops and more species of particles become total number of distinct elementary particles that exist (of order 100, according to the standard entropy density: $s = [2\pi^2 k/45(hc)^3] h_{\text{eff}} (kT)^3$. In equilibrium, $h_{\text{eff}} = g_{\text{eff}}$, but this ceases to be true at late times, when the neutrinos and photons have different temperatures. The g_{eff} functions are plotted against photon temperature in figure 3. They start at a number determined by the expressing the fermion contribution as an effective number of bosons. A similar relation holds for

$$\rho c^2 = \frac{\pi^2}{30(hc)^3} g_{\text{eff}} (kT)^4; \quad g_{\text{eff}} \equiv \sum_{\text{bosons}} g_i + \frac{8}{7} \sum_{\text{fermions}} g_j, \quad (56)$$

DEGREES OF FREEDOM Overall, the equilibrium relativistic density is

called the **entropy per baryon**. This justifies a common piece of terminology, in which the ratio of the number density of photons in the universe to the number density of **baryons** (protons plus neutrons) is number of particles. Now, we saw earlier that the number density of an ultrarelativistic background also scales as T^3 —therefore we have the simple result that entropy just counts the and 7/8 of this for fermions. Now, we saw earlier that the number density of an ultrarelativistic

$$s = \frac{2\pi^2 k}{45(hc)^3} g (kT)^3 \quad (\text{bosons}), \quad (55)$$

These results take an interesting and simple form in the ultrarelativistic limit. The energy density, u , obeys the usual black-body scaling $u \propto T^4$. In the ultrarelativistic limit, we also know that the pressure is $P = u/3$, so that the entropy density is

$$S = \frac{E + PV}{T} \quad (54)$$

Equating the dV and dT parts gives the familiar $\partial E/\partial T = T \partial S/\partial T$ and

$$dE = T dS - P dV \Leftrightarrow \left(\frac{\partial E}{\partial T} dT + \frac{\partial E}{\partial V} dV \right) = \left(T \frac{\partial S}{\partial T} dT + T \frac{\partial S}{\partial V} dV \right) - P dV. \quad (53)$$

It will also be useful to know the **entropy of the background**. This is not too hard to work out, because energy and entropy are extensive quantities for a thermal background. Thus, writing the first law for $\mu = 0$ and using $\partial S/\partial V = S/V$ etc. for extensive quantities,

$$S = \langle \sigma v \rangle n_T^2. \quad (63)$$

To complete the Boltzmann equation, we need the source term S . This term can be fixed by a thermodynamic equilibrium argument: for a non-expanding universe, n will be constant at the equilibrium value for that temperature, n_T , showing that

It allows for a whole set of **relics** to exist from different stages of the hot big bang. This phenomenon of freezeout is essential to the understanding of the present-day nature of the 'snapshot' of the properties of the universe at the time the particle was last in thermal equilibrium. The situation therefore changes from one of thermal equilibrium at early times to a state of **freezeout or decoupling** at late times. Once the interaction timescale becomes much longer than the age of the universe, the particle has effectively ceased to interact. It thus preserves a

$$\text{expansion timescale} = H(z)^{-1} \quad \text{interaction timescale} = \langle \sigma v \rangle n^{-1} \quad (62)$$

where σ is the reaction cross-section, v is the particle velocity, and S is a source term that represents thermal particle production. The $3Hn$ term just represents dilution by the expansion of the universe. Leaving aside the source term for the moment, we see that the change in n involves two timescales:

$$n + 3Hn = -\langle \sigma v \rangle n^2 + S, \quad (61)$$

Nevertheless, the annihilations only proceed at a finite rate: each member of the pair has to find a partner to interact with. We can express this by writing a simple differential equation for the electron density, called the **Boltzmann equation**:

So far, we have assumed that thermal equilibrium will be followed in the early universe, but this is far from obvious. Equilibrium is produced by reactions that involve individual particles, e.g. $e^+e^- \leftrightarrow 2\gamma$ converts between electron-positron pairs and photons. When the temperature is low, typical photon energies are too low for this reaction to proceed from right to left, so there is nothing to balance annihilations.

2.2 Freezeout and relics

Event	T	kT	g_{eff}	redshift	time
Now	2.73 K	0.0002 eV	3.3	0	13 Gyr
Distant galaxy	16 K	0.001 eV	3.3	5	1 Gyr
Recombination	3000 K	0.3 eV	3.3	1100	10 ^{5.6} years
Radiation domination	9500 K	0.8 eV	3.3	3500	10 ^{4.7} years
Electron pair threshold	10 ^{9.7} K	0.5 MeV	11	10 ^{9.5}	3 s
Nucleosynthesis	10 ¹⁰ K	1 MeV	11	10 ¹⁰	1 s
Nucleon pair threshold	10 ¹³ K	1 GeV	70	10 ¹³	10 ^{-6.6} s
Electroweak unification	10 ^{15.5} K	250 GeV	100	10 ¹⁵	10 ⁻¹² s
Grand unification	10 ²⁸ K	10 ¹⁵ GeV	100(?)	10 ²⁸	10 ⁻³⁶ s
Quantum gravity	10 ³² K	10 ¹⁹ GeV	100(?)	10 ³²	10 ⁻⁴³ s

The following table shows some of the key events in the history of the universe. Note that, for very high temperatures, energy units for kT are often quoted instead of T . The conversion is $kT = 1 \text{ eV}$ for $T = 10^{4.06} \text{ K}$. Some of the numbers are rounded, rather than exact; also, some of them depend a little on Ω and H_0 . Where necessary, a flat model with $\Omega = 0.3$ and $h = 0.7$ has been assumed.

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.

$$1 + z_{\text{eq}} = 24074 \Omega h^2 (T/2.725 \text{ K})^{-4}. \quad (60)$$

The energy density in photons is supplemented by that of the neutrino background. Because they have a lower temperature, as shown below, they contribute an energy density 0.68 times that from the photons (if the neutrinos are massless and therefore relativistic). If there are no other contributions to the energy density from relativistic particles, then the total effective radiation density is $\Omega_r h^2 \simeq 4.2 \times 10^{-5}$ and the redshift of **matter–radiation equality** is

which is justified informally by arguing that photon energies scale as $E \propto 1/a$ and saying that the typical energy in black-body radiation is $\sim kT$. Being more careful, we should conserve entropy, so that $s \propto a^{-3}$. Since $s \propto T^3$ while h_{eff} is constant, this requires $T \propto 1/a$. But clearly this does not apply near a threshold. At these points, h_{eff} changes rapidly and the universe will expand at nearly constant temperature for a period.

$$T(z) = 2.725(1+z), \quad (59)$$

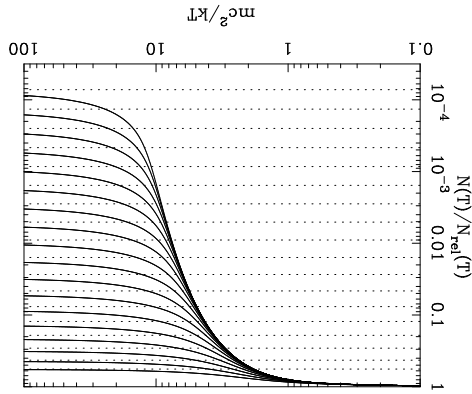
Frequently, we will assume This temperature was of course higher in the past, owing to the adiabatic expansion of the universe.

where the 'picobarn' is $1 \text{ pb} = 10^{-40} \text{ m}^2$. Thus only a small range of annihilation cross-sections will be of observational interest. The steps needed to get this formula are as follows. (1) From $\Gamma/H = 1$, the number density of relics at freezeout is $n_f = H^2/\langle\sigma v\rangle$; (2) $H = 8\pi G\rho/3^{1/2}$, where $\rho c^2 = (\pi^2/30)h^3 c^3 g^{\text{eff}}(kT)_4^4$; (3) $\Omega_{\text{relic}} = 8\pi G m m_0/3H_0^2$. The only missing ingredient here is how

$$\Omega_{\text{relic}} h^2 \approx 0.03 (\sigma/\text{pb})^{-1}, \quad (66)$$

THE RELIC DENSITY The above freezeout criterion can be used to deduce a simple and very important expression for the present-day density of a non-relativistic relic:

Figure 4. Solution of the Boltzmann equation for freezeout of a single massive fermion. We set $\Gamma/H = \epsilon(kT/mc^2)N/N_{\text{rel}}$, as appropriate for a radiation-dominated universe in which $\langle\sigma v\rangle$ is assumed to be independent of temperature. The solid lines show the case $\epsilon = 1$ and increasing by powers of 2. A high value of ϵ leads to freezeout at increasingly low abundances. The dashed lines show the abundance predicted by the simple recipe of the thermal density for which $\Gamma/H = 1$.



Exact numerical solutions of the rate equation almost always turn out very close to this simple rule, as shown in figure 4.

$$N(a \rightarrow \infty) = N_f(\Gamma/H = 1). \quad (65)$$

Now consider the opposite case, where the thermal solution would be nonrelativistic, with $N_f \propto T^{-3/2} \exp(-mc^2/kT)$. If the background stays at the equilibrium value, the lhs of the rate equation will therefore be negative and $\gg 1$ in magnitude. This is consistent if $\Gamma/H \gg 1$, because then the $(N_f/N)^2$ term on the rhs can still be close to unity. However, if $\Gamma/H \ll 1$, there must be a deviation from equilibrium. When N_f changes sufficiently fast with a , the actual abundance cannot keep up, so that the $(N_f/N)^2$ term on the rhs becomes negligible and $d \ln N/d \ln a \approx -\Gamma/H$, which is therefore a critical time at which the reaction rate drops low enough that particles are simply conserved as the universe expands – the population has **frozen out**. This provides a more detailed justification for the intuitive rule-of-thumb used above to define decoupling.

Unfortunately, this equation must be solved numerically. The main features are easy enough to see, however. Suppose first that the universe is sustaining a population in approximate thermal equilibrium, $N \approx N_f$. If the population under study is relativistic, N_f does not change with time, because $n_f \propto T^3$ and $T \propto a^{-1}$. This means that it is possible to keep $N = N_f$ exactly, whatever Γ/H . It would however be grossly incorrect to conclude from this that the population stays in thermal equilibrium: if $\Gamma/H \gg 1$, a typical particle suffers no interactions even while the universe doubles in size, halving the temperature. A good example is the microwave background, whose photons last interacted with matter at $z \approx 1100$.

where $\Gamma = n\langle\sigma v\rangle$ is the interaction rate experienced by the particles.

$$\frac{d \ln N}{d \ln a} = -\frac{H}{\Gamma} \left[1 - \left(\frac{N}{N_f} \right)^2 \right], \quad (64)$$

If we define comoving number densities $N \equiv a^3 n$ (effectively the ratio of n to the relativistic density for that temperature, n_{rel}), the rate equation can be rewritten in the simple form

Based on this, even the electron neutrino could be of great cosmological significance. But in practice, we will see later that studies of cosmological large-scale structure limit the sum of the masses to a maximum of about 0.5 eV. This is becoming interesting, since it is known that neutrino masses must be non-zero. In brief, this comes from studies of **neutrino mixing**, in which each neutrino type

$$(72) \quad \nu_e \lesssim 2.2 \text{ eV} \quad \nu_\mu \lesssim 0.17 \text{ MeV} \quad \nu_\tau \lesssim 15 \text{ MeV}.$$

The current direct laboratory limits to the neutrino masses are

The more complicated case of neutrinos that decouple when they are already nonrelativistic is studied below.

$$(71) \quad \Omega_\nu h^2 = \frac{\sum m_\nu}{94.1 \text{ eV}}.$$

MASSIVE NEUTRINOS Theoretical progress in understanding the origin of masses in particle physics means that there is no reason for the neutrino to be completely devoid of mass. Also, there is now clear experimental evidence that neutrinos have a small non-zero mass. The consequences of this for cosmology could be quite profound, as relic neutrinos are expected to be very abundant. The above section showed that $n(\nu + \bar{\nu}) = (3/4)n(\gamma)$; $T = 1.945 \text{ K}$. That yields a total of 113 relic neutrinos in every cm^3 for each species. Suppose these neutrinos were ultrarelativistic at decoupling: as the universe expands to $\gamma T > m_\nu c^2$, the total number of neutrinos is preserved, so the present-day mass density in neutrinos is just the zero-mass number density times m_ν , and the consequence for the cosmological density in light neutrinos is easily worked out to be

there are three **leptons**: the μ and τ particles are heavy analogues of the electron). density in relativistic particles by a factor 1.68 (there are three different kinds of neutrinos, just as for $T_\gamma = 2.725 \text{ K}$. These relativistic relic neutrinos contribute an energy density that is a factor $(7/8) \times (4/11)^{4/3}$ times that of the photons. For three neutrino species, this enhances the energy

$$(70) \quad T_\nu = \left(\frac{4}{11}\right)^{1/3} T_\gamma = 1.945 \text{ K},$$

Equating this to photon entropy at a new temperature gives the factor by which the photon temperature is enhanced with respect to that of the neutrinos. Thus we infer the existence of a neutrino background with a temperature

25

$$(69) \quad s(\gamma + e^+ + e^-) = \frac{11}{4} s(\gamma).$$

The effect of the electron-positron annihilation is therefore to enhance the numbers of photons relative to neutrinos. Strictly, what is conserved in this process is the *entropy*. The entropy of an $e^\pm + \gamma$ gas is easily found by remembering that it is proportional to the number density, and that all three particle species have $g = 2$ (polarization or spin). The total is then

NEUTRINO DECOUPLING The best case for application of this freezeout apparatus is to relic neutrinos. At the later stages of the big bang, energies are such that only light particles survive in equilibrium: photons (γ), neutrinos (ν) and e^+e^- pairs. As the temperature falls below $T_e = 10^9 \text{ K}$, interaction, so in principle the annihilations might yield pairs of photons or neutrinos. However, in practice the weak reactions freeze out earlier, at $T \approx 10^{10} \text{ K}$.

We see from figure 4 that $m_e^2/kT_f \sim 10$ with only a logarithmic dependence on reaction rate, which roughly cancels the last factor on the rhs. Finally, since particles are nearly relativistic at freezeout, we set $\langle \sigma v \rangle = \sigma c$ to get our final estimate of the typical cross-section for an interesting relic abundance. The eventual conclusion makes sense: the higher the cross-section, the longer the particle can stay in equilibrium, and the more effective annihilations can be in suppressing the number density. Note that, in detail, we need to worry about whether the particle is a **Majorana particle** (i.e. its own antiparticle) or a **Dirac particle** where particles and antiparticles are distinct.

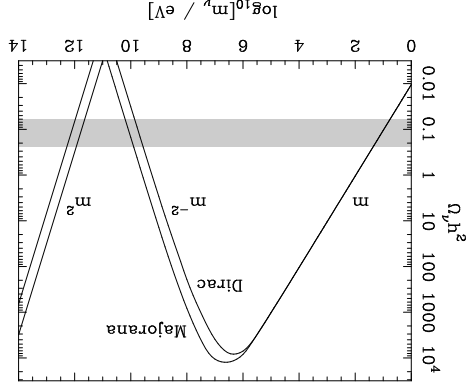
$$(68) \quad \Omega_{\text{relic}} h^2 \approx \frac{10^{-33.0} m^2}{m_e^2} \frac{\langle \sigma v \rangle}{kT_f} g_{\text{eff}}^{-1/2}.$$

Today, $h_0^{\text{eff}} = 43/11$, and $h_{\text{eff}}^2 = g_{\text{eff}}$ at high redshift. This allows us to deduce the relic density, given the mass, cross-section and temperature of freezeout:

$$(67) \quad n_f/n_0 = (h_{\text{eff}}^3 T_3^f)/(h_0^{\text{eff}} T_3^0).$$

to relate the present number density n_0 to the density n_f at temperature T_f . Since the relics are conserved, the number density must have fallen by the same factor as the entropy density:

Figure 6. The contribution to the density parameter produced by relic neutrinos (or neutrino-like particles) as a function of their rest mass. The shaded band shows a factor of 2 either side of the observed CDM density. At low masses, the neutrinos are highly relativistic when they decouple: their abundance takes the zero-mass value, and the density is just proportional to the mass. Above about 1 MeV, the neutrinos are non-relativistic at decoupling, and their relic density is reduced by annihilation. Above the mass of the Z boson, the cross-section falls, so that annihilation is less effective and the relic density rises again.



RELIC PARTICLES AS DARK MATTER. Many other particles exist in the early universe, so there are a number of possible relics in addition to the massive neutrino. A common collective term for these particles is **WIMP** – standing for weakly interacting massive particle. There are really three generic types to consider, as follows.

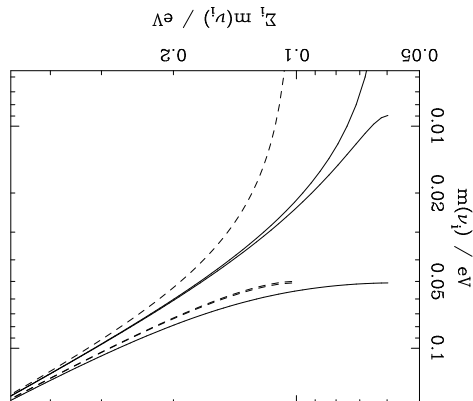
situation is a normal hierarchy with m_1 negligibly small, in which case the mass is dominated by m_3 , which is around 0.05 eV. The cosmological limits are within a power of 10 of this interesting point.

where m_1 , m_2 and m_3 are the three mass eigenstates. This information does not give the absolute mass scale, nor does it tell us whether there is a **normal hierarchy** with $m_3 \gg m_2 \gg m_1$, or an **inverted hierarchy** in which states 1 & 2 are a close doublet lying well above state 3. Cosmology can settle both these issues by measuring the total density in neutrinos. The absolute minimum

$$(73) \quad \begin{aligned} \Delta(m_{21})^2 &= 8.0 \times 10^{-5} \text{ eV}^2 \\ \Delta(m_{32})^2 &= 2.5 \times 10^{-3} \text{ eV}^2 \end{aligned}$$

is a mixture of energy eigenstates. The energy differences can be measured, which yields a measure of the difference in the square of the masses (consider the relativistic relation $E^2 = m^2 + p^2$, and expand to get $E \approx m + m^2/2p$). These mixings are known from wonderfully precise experiments detecting neutrinos generated in the sun and the Earth's atmosphere:

Figure 5. The masses of the individual neutrino mass eigenstates, plotted against the total neutrino mass for a normal hierarchy (solid lines) and an inverted hierarchy (dashed lines). Current cosmological data set an upper limit on the total mass of light neutrinos of around 0.5 eV.



(1) **Hot Dark Matter (HDM)** These are particles that decouple when relativistic, and which have a number density roughly equal to that of photons; eV-mass neutrinos are the archetype. The relic density scales linearly with the particle mass.

(2) **Warm Dark Matter (WDM)** If the particle decouples sufficiently early, the relative abundance of photons can then be boosted by annihilations other than just e^\pm . In modern particle physics theories, there are of order 100 distinct particle species, so the critical particle mass to make $\Omega = 1$ can be boosted to around $1 - 10$ keV.

(3) **Cold Dark Matter (CDM)** If the relic particles decouple while they are nonrelativistic, the number density can be exponentially suppressed. If the interactions are like those of neutrinos, then the freezeout temperature is about 1 MeV, and the relic mass density then falls with increasing mass (see figure 6). For weak interactions, cross-sections scale as (energy)², so that the relic density falls as $1/m^2$. Interesting masses then lie in the ~ 10 GeV range, this cannot correspond to the known neutrinos, since such particles would have been seen in accelerators. But beyond about 90 GeV (the mass of the Z boson), the strength of the weak interaction is reduced, with cross-section going as (energy)⁻². The relic density now rises as m^2 , so that the observed dark matter density is attained at $m \sim 1$ TeV. Plausible candidates of this sort are found among so-called **supersymmetric theories**, which predict many new weakly-interacting particles. The favoured particle for a CDM relic is called the **neutralino**.

Since these particles exist to explain galaxy rotation curves, they must be passing through us right now. There is therefore a huge effort in the direct laboratory detection of dark matter, mainly via cryogenic detectors that look for the recoil of a single nucleus when hit by a DM particle (in deep mines, to shield from cosmic rays). Well-constructed experiments with low backgrounds are starting to set interesting limits, as shown in figure 7. There is no unique target to aim for, since even the simplest examples of supersymmetric models contain a variety of free parameters. These allow models that are optimistically close to current limits, but also some that will be hard to verify. The public-domain package **DARKSUSY** is available at www.phy.sto.se/~edsjo/darksusy to make these detailed abundance calculations.

This subject saw a lot of publicity at the end of 2009, when the **CDMS** experiment announced events that were consistent with relic WIMPs (see <http://arxiv.org/abs/0912.3592>). In brief, cryogenic Ge and Si detectors are examined for evidence of nuclear recoil, which manifests itself in two distinct ways: heat (phonons) and ionization (electrons). The double signature allows rejection of many non-WIMP background events, although high-energy neutrons from cosmic ray events or radioactivity are a fundamental limit. CDMS estimate that these processes should cause on average 0.8 WIMP-like events during their 2 years of data; 2 events were actually seen. This is thus not so far inconsistent with background, but it is equally possible that there is a signal at a level of up

What is particularly exciting is that the properties of these relic particles can also be observed via new examples manufactured in particle accelerators. The most wonderful outcome would be for the same particle to be found in these two different ways. The chances of success in this enterprise are hard to estimate, and some models exist in which detection would be impossible for many decades. But it would be a tremendous scientific achievement if dark matter particles were to be detected in this way, and a good part of the plausible parameter space will be covered over the next decade.

BARYOGENESIS It should be emphasised that these freezeout calculations predict equal numbers of particles and antiparticles. This makes a critical contrast with the case of normal or **baryonic** material. The number density of baryons is low (roughly 10^{-9} that of the CMB photons), so one's first thought might be that baryons are another frozen-out relic. But as far as is known, there is a negligible cosmic density of antibaryons; even if antimatter existed, freezeout applied to protons-antiproton pairs predicts a density far below what is observed. The inevitable conclusion is that the universe began with a very slight asymmetry between matter and antimatter: at high temperatures there were $1 + O(10^{-9})$ protons for every antiproton. If baryon number is conserved, this imbalance cannot be altered once it is set in the initial conditions; but what generates it? This is clearly one of the big challenges in cosmology, but our ideas are less well formed here than in many other areas.

2.3 Recombination

Moving closer to the present, and passing through matter-radiation equality at $z \sim 10^4$, the next critical epoch in the evolution of the universe is reached when the temperature drops to the point ($T \sim 1000$ K) where it is thermodynamically favourable for the ionized plasma to form neutral atoms. This process is known as **recombination**: a complete misnomer, as the plasma has always been completely ionized up to this time.

THE RATE EQUATION A natural first thought is that the ionization of the plasma may be treated by a thermal-equilibrium approach, but such an approach is almost always invalid. This is not because electromagnetic interactions are too slow to maintain equilibrium: rather, they are too fast. Consider a single recombination; if this were to occur directly to the ground state, a photon with

$$x(z) \propto \frac{\Omega_b h^2}{\Omega_m h^2)^{1/2}}. \quad (76)$$

for a matter-dominated model at large redshift. Putting these together, the fractional ionization must scale as

$$\frac{d}{dt} \simeq -3.09 \times 10^{17} \Omega_m h^2)^{-1/2} z^{-5/2} \text{ s}, \quad (75)$$

An important point about the rate equation is that it is only necessary to solve it once, and the results can then be scaled immediately to some other cosmological model. Consider the this: both R and $\Lambda^{\nu}(T)$ are functions of temperature, and thus of redshift only, so that any parameter dependence is carried just by n^2 , which scales $\propto (\Omega_b h^2)^2$, where Ω_b is the baryonic density parameter. Similarly, the lhs depends on $\Omega_b h^2$ through n ; the other parameter dependence comes if we convert time derivatives to derivatives with respect to redshift:

where n is the number density of protons, x is the fractional ionization, R is the recombination coefficient ($R \simeq 3 \times 10^{-17} T^{-1/2} m^3 s^{-1}$), Λ^{ν} is the two-photon decay rate, and $\Lambda^{\nu}(T)$ is the stimulated transition rate upwards from the $2S$ state. This equation just says that recombinations are a two-body process, which create excited states that cascade down to the $2S$ level, from whence a competition between the upward and downward transition rates determines the fraction that make the downward transition.

$$\frac{d(nx)}{dt} = -R(nx)^2 \frac{\Lambda^{\nu} + \Lambda^{\nu}(T)}{\Lambda^{\nu}}, \quad (74)$$

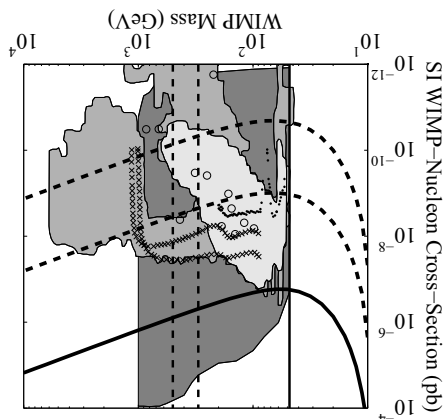
A highly stripped-down analysis of events simplifies the hydrogen atom to just two levels ($1S$ and $2S$). Any chain of recombinations that reaches the ground state can be ignored through the above argument: these reactions produce photons that are immediately re-absorbed elsewhere, so they have no effect on the ionization balance. The main chance of reaching the ground state comes through the recombinations that reach the $2S$ state, since some fraction of the atoms that reach that state will suffer two-photon decay before being re-excited. The rate equation for the fractional ionization is thus

transition by emitting a *pair* of photons. Because of this slow bottleneck, the ionization at low redshift is well above the equilibrium level.

There is a way out, however, using **two-photon emission**. The $2S \rightarrow 1S$ transition is strictly forbidden at first order and one can only conserve energy and angular momentum in the

escape). **scattering** and are very hard to get rid of (unlike a finite HII region, where the Ly α photons can efficiently. This is a common phenomenon in astrophysics: the Lyman α photons undergo **resonant** energetic as the $2P \rightarrow 1S$ spacing (Lyman α , with $\lambda = 1216\text{\AA}$), and these also are re-absorbed very neutral atom. Similarly, reaching the ground state requires the production of photons at least as $h\nu > \chi$ would be produced. Such photons are almost immediately destroyed by ionizing another

Figure 7. A plot of the dark-matter experimentalists' space: cross-section for scattering off nucleons (in wonderfully baroque units: the 'picobarn' is 10^{-40} m^2) against WIMP mass. The shaded areas and points indicate various supersymmetric models that match particle-physics constraints and have the correct relic density. The upper curve indicates direct (non-detection) limits, and dashed curves are where we might be in about a decade. Vertical lines are current collider limits, and predictions for the LHC and a future linear collider.



QUANTUM GRAVITY LIMIT In principle, $T \rightarrow \infty$ as $R \rightarrow 0$, but there comes a point at which this extrapolation of classical physics breaks down. This is where the thermal energy of typical particles is such that their de Broglie wavelength is smaller than their Schwarzschild radius: quantum black holes clearly cause difficulties with the usual concept of background spacetime. Equating $2\pi h/(mc)$

in our description of the early stages of cosmological history. The expanding universe of the big-bang model is surprising in many ways: (1) What caused the expansion? (2) Why is the expansion so close to flat - i.e. $\Omega \sim 1$ today? (3) Why is the universe close to isotropic (the same in all directions)? (4) Why does it contain structure? Some of these problems may seem larger than others, but when examined in detail all point to something missing

3.1 Initial condition problems

- Noether's theorem
- Dynamics of scalar fields
- Initial condition problems

Topics to be covered:

3 Inflation - I

(cf: Jones & Wyse 1985).

$$\tau(z) \approx \left(\frac{1+z}{1080} \right)^{13} \tag{79}$$

Also, $dt_{\text{proper}} = c dt$, which brings in a factor of $(\Omega_m h^2)^{-1/2}$. These two density terms automatically cancel the principal dependence of $x(z)$, so we predict that the optical depth should be very largely a function of redshift only. For standard parameters, a good approximation around $\tau = 1$ is

33

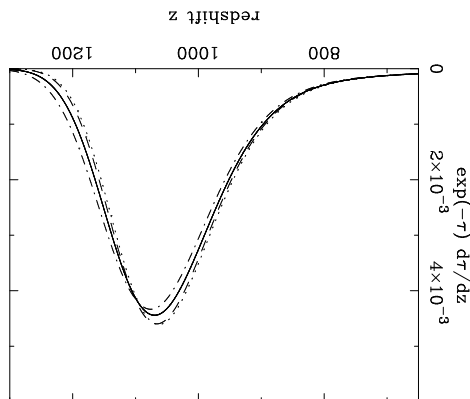
$$n_{\text{tot}}^e(z) = 9.83 \Omega_b h^2 (1+z)^3 m^{-3}. \tag{78}$$

For a fully ionized plasma with 25% He by mass, the total electron number density is

$$\tau = \int n_{\text{tot}}^e x \sigma_T dt_{\text{proper}}; \quad dt_{\text{proper}} = R(z) dr = R_0 dr / (1+z). \tag{77}$$

LAST SCATTERING Recombination is important observationally because it marks the first time that photons can travel freely. When the ionization is high, Thomson scattering causes them to proceed in a random walk, so the early universe is opaque. The interesting thing from our point of view is to work out the maximum redshift from which we can receive a photon without it suffering scattering. To do this, we work out the optical depth to Thomson scattering.

Figure 8. The 'visibility function' governing where photons in the CMB undergo their final scattering. This is very nearly independent of cosmological parameters, as illustrated by the effect of a 50% increase in Ω_b (dotted line), Ω_m (dot-dashed line) and h (dashed line), relative to the standard model (solid line).

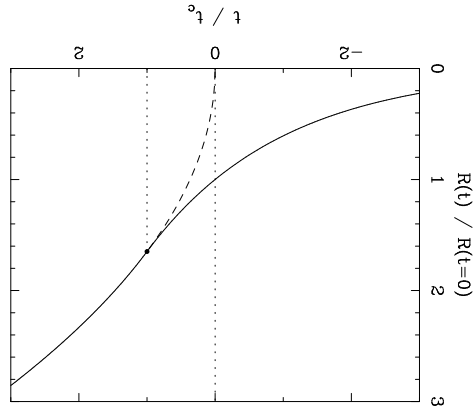


In a vacuum-dominated phase, ρR^2 increases as the universe expands. This term can therefore always be made to dominate over the curvature term, making a universe that is close to being flat

$$R^2 = \frac{8\pi G \rho R^2}{3} - k c^2. \quad (91)$$

however. Before trying to plug this critical gap, we can note that an early phase of vacuum-dominated expansion can also solve the flatness problem. Consider the Friedmann equation,

Figure 9. Illustrating the true history of the scale factor in the simplest possible inflationary model. Here, the universe stays in an exponential de Sitter phase for an infinite time until its equation of state abruptly changes from vacuum dominated to radiation dominated at time t_c . This must occur in such a way as to match R and \dot{R} , leading to the solid curve, where the plotted point indicates the join. For $0 < t < t_c$, the dashed curve indicates the time dependence we would infer if vacuum energy was ignored. This reaches $R = 0$ at $t = 0$: the classical 'big bang'. The inflationary solution clearly removes this feature, placing any singularity at large negative time. The universe is much older than we would expect from observations at $t > t_c$, which is one way of seeing how the horizon problem can be evaded.



E.B. Gliner, in 1969. He suggested no mechanism by which the vacuum energy could change its level, This idea of a non-singular origin to the universe was first proposed by the Soviet cosmologist

popular description of 'an extraordinarily rapid phase of expansion'. Inflationary start to the expansion would in reality be a very slow one – as compared to the common time, but one could imagine the de Sitter phase being of infinite duration. In a sense, then, an now answered: there was no big bang. There might have still been a singularity at large negative easy to show that $t_c = 1/2H$. In principle, the question 'what happened before the big bang?' is t_c , we have $R \propto \exp Ht$; after t_c , $R \propto t^{1/2}$. We have to match R and \dot{R} at the join; it is then equation of state abruptly from vacuum dominated to radiation dominated at some time t_c . Before This is illustrated by in figure 9, where we assume that the universe can be made to change its

where $H = \sqrt{8\pi G \rho_{vac}/3}$. This is the basic idea of the **inflationary universe**: vacuum repulsion can cause the universe to expand at an ever-increasing rate. This launches the Hubble expansion, and solves the horizon problem by stretching a small causally-connected patch to a size large enough to cover the whole presently-observable universe.

$$R \propto \exp Ht, \quad (90)$$

DE SITTER SPACE The familiar example of negative pressure is vacuum energy, and this is therefore a hint that the universe may have been vacuum-dominated at early times. The Friedmann equation in the $k = 0$ vacuum-dominated case has the **de Sitter solution**:

$$\rho c^2 + 3p > 0. \quad (89)$$

and realize that $R \propto t^\alpha$, with $\alpha > 1$ implies an accelerating expansion, we see that what is needed is negative pressure:

$$R = -4\pi G R(\rho + 3p/c^2)/3, \quad (88)$$

The standard radiation-dominated $R \propto t^{1/2}$ law makes this integral converge near $t = 0$. To solve the horizon problem and allow causal contact over the whole of the region observed at last scattering requires a universe that expands 'faster than light' near $t = 0$: $R \propto t^\alpha$, with $\alpha > 1$. It is tempting to assert that the observed homogeneity *proves* that such causal contact must once have occurred, but this means that the equation of state at early times must have been different. Indeed, if we look at Friedmann's equation in its second form,

We will show below that, not only does $V(\phi)$ play the role of a potential energy in the equation of motion, it acts as a physical energy density in space. This potential energy density is equivalent to a vacuum density: its gravitational properties are repulsive and can cause an inflationary phase of exponential expansion. In this simple model, the universe is started in a potential-dominated state, and inflates until the field falls enough that the kinetic energy becomes important. In practical models, this stage will be associated with **reheating**: although weakly interacting, the field does couple to other particles, and its oscillations can generate other particles – thus transforming the scalar-field energy into energy of a normal radiation-dominated universe.

See this, write the fields in terms of the potentials: $\mathbf{B} = \nabla \wedge \mathbf{A}$ and $\mathbf{E} = -\nabla\phi - \dot{\mathbf{A}}$. We don't see energy density in electromagnetism ($\epsilon_0 E^2/2 + B^2/2\mu_0$) is entirely kinetic energy in this analogy (to electromagnetic field: if the field is homogeneous, it does not oscillate. This is because the familiar with kinetic energy: $T = \phi^2/2$. This behaviour is rather different to the familiar oscillations of the potential, and we can see that the field will oscillate in the potential, and we can see that the field will oscillate in the potential, and we can see that the field will oscillate in the potential,

$$\ddot{\phi} = -\frac{d}{dt}V(\phi); \quad V(\phi) = (m^2 c^4/\hbar^2)\phi^2/2. \quad (98)$$

which is just the standard wave equation if $m = 0$. This is easy to derive just by substituting the de Broglie relations $\mathbf{p} = -i\hbar\nabla$ and $E = i\hbar\partial/\partial t$ into $E^2 = p^2 c^2 + m^2 c^4$. To apply this to cosmology, we neglect the spatial derivatives, since we imagine some initial domain in which we have a **homogeneous scalar field**. This synchronizes the subsequent dynamics of $\phi(t)$ throughout the observable universe (*i.e.* the patch that we inflate). The differential equation is now

$$\ddot{\phi} - \Delta^2 \phi + (m^2 c^2/\hbar^2)\phi = 0, \quad (97)$$

Since 1981, these ideas have been set on a more specific foundation using models for a variable vacuum energy that come from particle physics. There are many variants, but the simplest concentrate on **scalar fields**. These are fields like the electromagnetic field, but differing in a number of respects. First, the field has only one degree of freedom: just a number that varies with position, not a vector like the EM field. The wave equation obeyed by such a field in flat space is the **Klein–Gordon equation**:

3.2 Dynamics of scalar fields

39

The GUT energy corresponds to a time of about 10^{-35} s in the conventional radiation-dominated model, and we have seen that this switchover time should be of order $H_{\text{inflation}}^{-1}$. Therefore, the whole inflationary episode need last no longer than about 10^{-33} s.

This expands the Planck length to 0.4 mm today. This is far short of the present horizon ($\sim 6000 h^{-1} \text{Mpc}$), by a factor of nearly 10^{30} , or e^{69} . It is more common to assume that inflation happened at a safer distance from quantum gravity, at about the GUT energy of 10^{15} GeV. The GUT-scale horizon needs to be stretched by 'only' a factor e_{60} in order to be compatible with observed homogeneity. This tells us a minimum duration for the inflationary era:

$$z_p \approx 10^{31.6}. \quad (95)$$

HOW MUCH INFLATION DO WE NEED? To be quantitative, we have to decide when inflation is to happen. The earliest possible time is at the Planck era, $t \approx 10^{-43}$ s, at which point the causal scale was $ct \approx 10^{-35}$ m. What comoving scale is this? The redshift is roughly (ignoring changes in g_{eff}) the Planck energy (10^{19} GeV) divided by the CMB energy ($kT \approx 10^{-3.6}$ eV), or

of inflation (although, as we have seen, the expectation of flatness is fairly general). completion and yield $\Omega = 1$ to measurable accuracy today. This is one of the most robust predictions needed to solve the horizon problem. It then seems almost inevitable that the process should go to e-foldings of expansion are needed; it will be proved below that this is also exactly the number

$$\ln f \gtrsim 60 \quad (94)$$

This can solve the flatness problem, provided f is large enough. To obtain Ω of order unity today requires $|\Omega - 1| \lesssim 10^{-52}$ at the GUT epoch, and so

$$\Omega = 1 + O(f^{-2}). \quad (93)$$

where $H = \sqrt{8\pi G\rho_{\text{vac}}/3}$. Note that H is not the Hubble parameter at an arbitrary time (unless $k = 0$), but it becomes so exponentially fast as the hyperbolic trigonometric functions tend to the exponential. If we assume that the initial conditions are not fine tuned (*i.e.* $\Omega = O(1)$ initially), then maintaining the expansion for a factor f produces

$$R \propto \begin{cases} \exp Ht & (k = 0), \\ \cosh Ht & (k = +1), \\ \sinh Ht & (k = -1) \end{cases} \quad (92)$$

(the curvature scale has increased exponentially). In more detail, the Friedmann equation in the vacuum-dominated case has three solutions:

$$(106) \quad \sum_i \frac{\partial L}{\partial p_i} = \sum_i \left(\frac{\partial L}{\partial \dot{q}_i} + \frac{\partial L}{\partial p_i} \right) = \sum_i \frac{\partial L}{\partial \dot{q}_i}$$

If L has no explicit dependence on t , then

$$(105) \quad \sum_i \frac{\partial L}{\partial p_i} = \sum_i \frac{\partial L}{\partial \dot{q}_i} = 0.$$

Using Euler's equation, this gives conservation of total momentum. The Lagrangian as a whole will be unchanged by a $q_i \rightarrow q_i + dq_i$, where dq_i is a constant. More realistically, the potential will depend on the q_i , but homogeneity of space says that the Lagrangian is independent of all the position coordinates q_i , then we obtain conservation of momentum (or angular momentum, if q_i is an angular coordinate): $\frac{\partial L}{\partial q_i} = 0$ for each particle. If L is independent of all the position coordinates q_i , then we obtain conservation of momentum (or angular momentum, if q_i is an angular coordinate): $\frac{\partial L}{\partial q_i} = 0$ for each particle. More realistically, the potential will depend on the q_i , but homogeneity of space says that the Lagrangian is independent of all the position coordinates q_i , then we obtain conservation of momentum (or angular momentum, if q_i is an angular coordinate): $\frac{\partial L}{\partial q_i} = 0$ for each particle. More realistically, the potential will depend on the q_i , but homogeneity of space says that the Lagrangian is independent of all the position coordinates q_i , then we obtain conservation of momentum (or angular momentum, if q_i is an angular coordinate): $\frac{\partial L}{\partial q_i} = 0$ for each particle.

$$(104) \quad \frac{\partial L}{\partial p_i} = \frac{\partial L}{\partial \dot{q}_i} = 0,$$

The final ingredient we need before applying scalar fields to cosmology is to understand that they can be treated as a fluid with thermodynamic properties like pressure. These properties are derived by a profound argument that relates the existence of global symmetries to conservation laws in physics. In classical mechanics, conservation of energy and momentum arise by considering Euler's equation

Again, we will be content with checking that this does the right thing in a simple case: the homogeneous model, where $\mathcal{L} = \int dV \phi^2/2 - V(\phi)$. This is now just like the earlier example, and gives $\dot{\phi} = -\partial V/\partial \phi$, as required.

$$(103) \quad \mathcal{L} = \int dV \phi^2/2 - V(\phi); \quad \frac{\delta \mathcal{L}}{\delta \phi} = 0$$

For quantum mechanics, we want a Lagrangian that will yield the Klein-Gordon equation. If ϕ is a single real scalar field, then the required Lagrangian is

therefore generally equivalent, although the Lagrangian arguably seems more fundamental: we can obtain the field equations given the Lagrangian, but inverting the process is less straightforward.

where we use the shorthand $\partial^\mu \phi \equiv \partial \phi / \partial x_\mu$. Note the downstairs index for consistency: in special relativity, $x^\mu = (ct, \mathbf{x})$, $x_\mu = (ct, -\mathbf{x}) = g^{\mu\nu} x_\nu$. The Lagrangian \mathcal{L} and the field equations are

$$(102) \quad \frac{\partial \mathcal{L}}{\partial x^\mu} = \left[\frac{\partial \mathcal{L}}{\partial (\partial^\mu \phi)} \right] \partial^\mu \phi,$$

This suggests that we write the Lagrangian in terms of a **Lagrangian density** \mathcal{L} : $L = \int dV \mathcal{L}$. This quantity is of such central importance in quantum field theory that it is usually referred to (incorrectly) simply as 'the Lagrangian'. The equation of motion that corresponds to Euler's equation is now the **Euler-Lagrange equation**

$$(101) \quad H = \int dV \mathcal{H} = \int dV \left(\frac{1}{2} \dot{\phi}^2 + \frac{1}{2} \nabla^2 \phi^2 + V(\phi) \right)$$

A field may be regarded as a dynamical system, but with an infinite number of degrees of freedom. How do we handle this? A hint is provided by electromagnetism, where we are familiar with writing the total energy in terms of a density which, as we are dealing with generalized mechanics, we may formally call the Hamiltonian density:

which again is clearly sensible for Cartesian coordinates.

$$(100) \quad p_i \equiv \frac{\partial \mathcal{H}}{\partial \dot{q}_i}$$

As a sanity check, consider a single particle in a potential in 1D: $L = m\dot{x}^2/2 - V(x)$. $\partial L/\partial \dot{x} = m\dot{x}$, so we get $m\ddot{x} = -\partial V/\partial x$, as desired. The advantage of the Lagrangian formalism, of course, is that it is not necessary to use Cartesian coordinates. In passing, we note that the formalism also supplies a general definition of momentum:

$$(99) \quad \frac{\partial L}{\partial \dot{q}_i} = \frac{\partial L}{\partial \dot{q}_i}$$

To understand what scalar fields can do for cosmology, it is necessary to use some elements of the more powerful Lagrangian description of the dynamics. We will try to keep this fairly informal. Consider first a classical system of particles: the **Lagrangian** L is defined as the difference of the kinetic and potential energies, $L = T - V$, for some set of particles with coordinates $q_i(t)$. **Euler's equation** gives an equation of motion for each particle

$$(115) \quad \phi + 3H\dot{\phi} - \Delta^2\phi + p/V = 0.$$

We now want to revisit the equation of motion for the scalar field, but with the critical difference that we place the field in the expanding universe. Everything so far has been special relativity, so we don't have quite enough formalism to derive the full equation of motion, but it is

If the field is constant both spatially and temporally, the equation of state is then $p = -\rho$, as required if the scalar field is to act as a cosmological constant; note that derivatives of the field spoil this identification.

$$(114) \quad \begin{aligned} d = T_{00} &= \frac{1}{2}\dot{\phi}^2 + V(\phi) + \frac{1}{2}(\Delta\phi)^2 \\ d = T_{11} &= \frac{1}{2}\dot{\phi}^2 - V(\phi) - \frac{1}{2}(\Delta\phi)^2. \end{aligned}$$

From this, we can read off the energy density and pressure:

$$(113) \quad T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\mathcal{L}.$$

It turns out that we can get inflation with even the simple mass potential $V(\phi) = m^2\phi^2/2$, but it is easy to keep things general. Noether's theorem gives the energy-momentum tensor for the field as

$$(112) \quad \mathcal{T} = \frac{1}{2}\partial^\mu\phi\partial_\mu\phi - V(\phi) - \frac{1}{2}\Delta^2\phi - V(\phi).$$

Most of the main features of inflation can be illustrated using the simplest case of a single real scalar field, with Lagrangian

4.1 Equation of motion

- Models for inflation
- Slow roll dynamics
- Ending inflation

Topics to be covered:

4 Inflation – II

so now we can read off the density and pressure generated by a scalar field. Note immediately the important consequence for cosmology: a potential term $-V(\phi)$ in the Lagrangian produces $T_{\mu\nu} = V(\phi)g_{\mu\nu}$. This is the $p = -\rho$ equation of state characteristic of the cosmological constant. If we now follow the evolution of ϕ , the cosmological 'constant' changes and we have the basis for models of inflationary cosmology.

$$(111) \quad T_{\mu\nu} = \text{diag}(\rho, p, p, p),$$

where ρ is the charge density, \mathbf{j} is the current density, and J^μ is the 4-current. We have effectively four such equations (one for each value of ν) so there must be four conserved quantities: clearly energy and the four components of momentum. Conservation of 4-momentum is expressed by $T^{\mu\nu}$, which is the 4-current of 4-momentum. For a simple fluid, it is just

$$(110) \quad \partial^\mu J_\mu = 0,$$

This is a conservation law, as we can see by analogy with a simple case like the conservation of charge. There, we would write

$$(109) \quad \frac{d}{dt} \left[\frac{\partial \mathcal{L}}{\partial \phi} \frac{\partial \phi}{\partial t} - \mathcal{L} \right] = 0.$$

Using the Euler-Lagrange equation to replace $\partial\mathcal{L}/\partial\phi$ and collecting terms results in

$$(108) \quad \frac{d}{dt} \left[\frac{\partial \mathcal{L}}{\partial \dot{\phi}} \dot{\phi} + \frac{\partial \mathcal{L}}{\partial \phi} \phi \right] = 0,$$

Something rather similar happens in the case of quantum (or classical) field theory: the existence of a global symmetry leads directly to a conservation law. The difference between discrete dynamics and field dynamics, where the Lagrangian is a *density*, is that the result is expressed as a **conserved current** rather than a simple constant of the motion. Suppose the Lagrangian has no explicit dependence on spacetime (*i.e.* it depends on x^μ only implicitly through the fields and their 4-derivatives). As above, we write

$$(107) \quad H \equiv \sum_i p_i \dot{q}_i - L = \text{constant}.$$

which leads us to define the **Hamiltonian** as a further constant of the motion

If we think about a single field, models can be divided into two basic classes, as illustrated in figure 10. The simplest are **large-field inflation** models, in which the field is strongly displaced from the origin. There is nothing to prevent the scalar field from reaching the minimum of the potential – but it can take a long time to do so, and the universe meanwhile inflates by a large factor. In this case, inflation is realized by means of ‘inertial confinement’. The opposite is when the potential is something like the Higgs potential, where the gradient vanishes at the origin: this is a model of **small-field inflation**. In principle, the field can stay at $\phi = 0$ forever if it is placed exactly there. One would say that the universe then inhabited a state of **false vacuum**, as opposed to the true vacuum at $V = 0$ (but it is important to be clear that there is no fundamental reason why the minimum should be at zero density exactly; we will return to this point).

We therefore consider simple classes of possible example models, with varying degrees of physical motivation.

4.3 Inflationary models

These two criteria make perfect intuitive sense: the potential must be flat in the sense of having small derivatives if the field is to roll slowly enough for inflation to be possible.

$$(121) \quad \eta \equiv \frac{8\pi}{m_p^2} (V''/V) \gg 1$$

Also, we can differentiate this expression to obtain the criterion $V'' \gg V'/m_p$, or $m_p V''/V \gg V'/V \sim \sqrt{\epsilon}/m_p$. This gives a requirement for the second derivative of V to be small, which we can write as

$$(120) \quad \epsilon \equiv \frac{m_p^4}{16\pi} (V'/V)^2 \gg 1.$$

The conditions for inflation can be cast into useful dimensionless forms. The basic condition $V \gg \dot{\phi}^2$ can now be rewritten using the slow-roll relation as

This gives a powerful but simple apparatus for deducing the expansion history of any inflationary model.

45

$$(119) \quad H^2 = \frac{8\pi}{3} \frac{m_p^2}{V} (\dot{\phi}^2/2 + V) \approx \frac{3m_p^2}{8\pi} V,$$

In combination with Friedmann's equation in the natural-unit form

$$(118) \quad 3H\dot{\phi} = -dV/d\phi.$$

differentiating this gives $\dot{\phi} \ll |dV/d\phi|$, as required. We therefore have a simple slow-rolling equation for homogeneous fields:

$$(117) \quad |\dot{\phi}^2/2| \ll |V(\phi)|;$$

The solution of the equation of motion becomes tractable if we both ignore spatial inhomogeneities in ϕ and make the **slow-rolling approximation** that the $\dot{\phi}$ term is negligible. The physical motivation here is to say that we are most interested in behaviour close to de Sitter space, so that the potential dominates the energy density. This requires

4.2 The slow-roll approximation

following which the relations $H = d \ln a/dt$ and $V = \dot{\phi} V'$ can be used to change variables to t , and

$$(116) \quad \frac{d \ln a}{d \ln V} = -3(1+w) = -3\dot{\phi}^2/(\dot{\phi}^2/2 + V),$$

This is not hard to prove in the homogeneous case, which is the main one of interest for inflationary applications. This is because $\Delta \phi = \Delta_{\text{comoving}} \phi/R$. Since R increases exponentially, these perturbations are damped away: assuming V is large enough for inflation to start in the first place, inhomogeneities rapidly become negligible. In the homogeneous limit, we can simply appeal to energy conservation:

This is a wave equation similar to the one in flat space. The **Hubble drag** term $3H\dot{\phi}$ is the main new feature: loosely, it reflects the fact that the redshifting effects of expansion will drain energy from the field oscillations.

Hybrid inflation therefore has the ability to make some of the features of the simplest inflation models seem more plausible, while introducing sufficient extra complexity that one can try to test is particularly high, keeping the field from rolling away from $\psi = 0$ until near to $\phi = 0$.

We can think of this as being primarily $V(\psi)$, but with the form of V controlled by the second field, ϕ . For $\phi = 0$, we have the standard symmetry-breaking potential; but for large ϕ , $\phi > M/g$, the dependence on ψ becomes parabolic. Evolution in this parabolic trough at large ϕ can thus naturally lower ψ close to $\psi = 0$. If this happens, we have inflation driven by ϕ as the inflaton, with $V(\phi) = U(\phi) + \lambda M^2/4$. This extra constant in the potential raises H , so the Hubble damping term

$$V(\phi, \psi) = \frac{1}{4} M^2 (\lambda \psi^2)^2 + U(\phi) + \frac{1}{2} g^2 \psi^2 \phi^2. \tag{123}$$

One way in which the symmetric nature of the initial condition for small-field inflation can be made more plausible is to go beyond the space of single-field inflation. The most popular model in this generalized class is **hybrid inflation**, in which there are two fields, with potential

This is an exact solution, not a slow-roll approximation.

$$V(\phi) \propto \exp \left(\sqrt{\frac{16\pi}{3}} \frac{p}{m_p^2} \phi \right). \tag{122}$$

Power-law inflation. On the other hand, $a(t) \propto t^p$ would suffice, provided $p > 1$. The potential required to produce this behaviour is

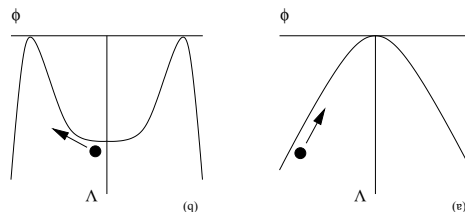
- (1) **Polynomial inflation.** If the potential is taken to be $V \propto \phi^\alpha$, then the scale-factor behaviour can be very close to exponential. This becomes less true as α increases, but investigations are usually limited to ϕ^2 and ϕ^4 potentials on the grounds that higher powers are nonrenormalizable.

A wide range of inflation models of this kind is possible, but it will suffice here to discuss two simple special cases:

CHAOTIC INFLATION MODELS Most attention is therefore currently paid to the large-field models where the field finds itself some way from its potential minimum. This idea is also termed **chaotic inflation**: there could be primordial chaos, within which conditions might vary. Some parts may leave a universe inside a single bubble – which could be the one we inhabit. In principle this bubble has an edge, but if inflation persists for sufficiently long, the distance to this nastiness is so much greater than the current particle horizon that its existence has no testable consequences.

The first inflation model (Guth 1981) was of the small-field type, but large-field models have tended to be considered more plausible, for two reasons. The first is to do with initial conditions. If inflation starts from anywhere near to thermal equilibrium at a temperature T_{GUT} , we expect thermal fluctuations in ϕ ; the potential should generally differ from its minimum by an amount $V \sim T_{\text{GUT}}^4$. How then is the special case needed to trap the potential near $\phi = 0$ to arise? We have returned to the sort of fine-tuned initial conditions from which inflation was designed to save us. The other issue with simple small-field models relates to the issue of how inflation ends. This can be viewed as a form of phase transition, which is continuous or second order in the case of large-field models. For small-field models, however, the transition to the true vacuum can come about by quantum tunnelling, so that the transition is effectively discontinuous and first order. As we will discuss further below, this can lead to a universe that is insufficiently homogeneous to be consistent with observations.

Figure 10. The two main classes of single-field inflation models: (a) large-field inflation; (b) small-field inflation. The former is motivated by a mass-like potential, the latter by something more like the Higgs potential.



An effective density perturbation is thus produced by a local alteration in the time coordinate: when we look at a universe with a fluctuating density, should we really think of a uniform model in which time is wrinkled? This ambiguity may seem absurd, and in the laboratory it could be resolved empirically by constructing the coordinate system directly – in principle by using light signals. This shows that the cosmological horizon plays an important role in this topic: perturbations with wavelength $\lambda \lesssim ct$ inhabit a regime in which gauge ambiguities can be resolved directly via common

$$(127) \quad \rho' = \rho - \epsilon^0 p.$$

Consider applying this to the case of a uniform universe; here ρ only depends on time, so that

$$(126) \quad S'(x_H) = S(x_H) - \epsilon^0 \partial S / \partial x^0.$$

Because relativistic physics equations are written in a covariant form in which all quantities are independent of coordinates, relativity does not distinguish between *active* changes of coordinate (e.g. a Lorentz boost) or *passive* changes (a mathematical change of variable, normally termed a gauge transformation). This generality is a problem, as we can see by asking how some scalar quantity S (which might be density, temperature etc.) changes under a gauge transformation $x_H \rightarrow x_H + \epsilon_H$. A gauge transformation induces the usual Lorentz transformation coefficients dx_H/dx^0 (which have no effect on a scalar), but also involves a translation that relabels spacetime points. We therefore have $S'(x_H) = S(x_H) + \epsilon_H$, or

method. The first issue we have to deal with is how to quantify departures from uniform density. Frequently, an intuitive Newtonian approach can be used, and we will adopt this whenever possible. But we should begin with a quick overview of the relativistic approach to this problem, to emphasise some of the big issues that are ignored in the Newtonian

We now need to consider the greatest achievement of inflation, which was not anticipated when the theory was first put forward: it provides a concrete mechanism for generating the seeds of structure in the universe. In essence, the idea is that the inevitable small quantum fluctuations in the inflation field ϕ are transformed into residual classical fluctuations in density when inflation is over. The details of this process can be technical, and could easily fill a lecture course. The following treatment is therefore simplified as far as possible, while still making contact with the full results.

5.1 The perturbed universe

- Description of inhomogeneity
- Mechanisms for fluctuation generation
- Tilt and tensor modes
- Eternal inflation

Topics to be covered:

5 Fluctuations from inflation

As well as being of interest for completing the picture of inflation, it is essential to realize that these closing stages of inflation are the *only* ones of observational relevance. Inflation might well continue for a huge number of e -foldings, all but the last few satisfying $\epsilon, n \gg 1$. However, the scales that left the de Sitter horizon at these early times are now vastly greater than our observable horizon, c/H_0 , which exceeds the de Sitter horizon by only a finite factor – about e^{60} for GUT-scale inflation, as we saw earlier. Realizing that the observational regime corresponds only to the terminal phases of inflation is both depressing and stimulating: depressing, because ϕ may well not move very much during the last phases – our observations relate only to a small piece of the potential, and we cannot hope to recover its form without substantial *a priori* knowledge; stimulating, because observations even on very large scales must relate to a period where the simple concepts of exponential inflation and scale-invariant density fluctuations were coming close to breaking down. This opens the possibility of testing inflation theories in a way that would not be possible with data relating to only the simpler early phases. These tests take the form of tilt and gravitational waves in the final perturbation spectrum, to be discussed further below.

However, this conclusion is incomplete, because we have so far neglected the couplings of particles, leading to **reheating**. Thus, even if the minimum of $V(\phi)$ is at $V = 0$, the universe is left containing roughly the same energy density as it started with, but now in the form of normal matter and radiation – which starts the usual FRW phase, albeit with the desired special ‘initial’ conditions.

Eventually, we will be left with a stationary field that either continues to inflate without end, if about the bottom of the potential, with the oscillations becoming damped by the $3H\dot{\phi}$ friction term. $V(\phi) > 0$, or which simply has zero density.

$$(134) \quad \delta(\mathbf{k}) = \int_{\mathcal{V}} \delta(\mathbf{x}) \exp(i\mathbf{k} \cdot \mathbf{x}) d^3x.$$

and the inverse Fourier relation would be

$$(133) \quad k_x = n \frac{T}{2\pi}, \quad n = 1, 2, \dots,$$

where \mathbf{k} is the comoving wavevector. What are the allowed modes? If the field were periodic within some box of side L , we would have the usual harmonic boundary conditions

$$(132) \quad \delta(\mathbf{x}) = \sum_{\mathbf{k}} \delta^{\mathbf{k}} e^{-i\mathbf{k} \cdot \mathbf{x}},$$

to making a Fourier expansion:

We are very often interested in asking how these fluctuations depend on scale, which amounts

for the relativistic active mass density ρ and the factor $(1 + 3w)$ accounts the factor of a^2 is there so we can use comoving length units in Δ^2

$$(131) \quad \delta \equiv \frac{d}{\rho}.$$

where we have defined a dimensionless fluctuation amplitude

$$(130) \quad \Delta^2 \Phi / a^2 = 4\pi G(1 + 3w) \rho \delta,$$

directly related to those in density via Poisson's equation. From the Newtonian point of view, potential fluctuations are

FLUCTUATION POWER SPECTRA From the Newtonian point of view, potential fluctuations are and made plausible in this way. The main results of the full theory can at least be understood to go beyond the Newtonian approach. But with the length restrictions of this course, it is hard

53

The conclusion is thus that the gravitational potential can for many purposes give an effectively gauge-invariant measure of cosmological perturbations. The advantage of this fact is that the

and this should be quite familiar. If we consider small scales, so that the spatial metric γ_{ij} becomes that of flat space, then this form matches, for example, the Schwarzschild metric with $\Phi = -GM/r$, in the limit $\Phi/c^2 \gg 1$.

$$(129) \quad d\tau^2 = (1 + 2\Phi)dt^2 - (1 - 2\Phi)\gamma_{ij}dx^i dx^j,$$

A second key result is that inserting the longitudinal metric into the Einstein equations shows that Ψ and Φ are identical in the case of fluid-like perturbations where off-diagonal elements of the energy-momentum tensor vanish. In this case, the longitudinal gauge becomes identical to the **Newtonian gauge**, in which perturbations are described by a single scalar field, which is the gravitational potential:

Health warning: there are different conventions, and the symbols for the potentials are sometimes swapped, or signs flipped.

$$(128) \quad d\tau^2 = (1 + 2\Psi)dt^2 - (1 - 2\Phi)\gamma_{ij}dx^i dx^j.$$

A key result is that scalar perturbations can be described by just two gauge-invariant potentials (functions of spacetime coordinates). Since these are gauge-invariant, we may as well write the perturbed metric in a particular gauge that makes things look as simple as possible. This is the **longitudinal gauge** in which the time and space parts of the RW metric are perturbed separately:

Firstly, metric perturbations can be split into three classes: **scalar perturbations**, which are described by scalar functions of spacetime coordinate, and which correspond to growing density perturbations; **vector perturbations**, which correspond to vorticity perturbations, and **tensor perturbations**, which correspond to gravitational waves. Here, we shall concentrate mainly on scalar perturbations.

The most direct general way of solving these difficulties is to construct perturbation variables that are explicitly independent of gauge. A comprehensive technical discussion of this method is given in chapter 7 of Mukhanov's book, and we summarize the essential elements here, largely without proof.

sense. The real difficulties lie in the super-horizon modes with $\lambda \gtrsim ct$. Within inflationary models, however, these difficulties can be overcome, since the true horizon is $\gg ct$.

Before going any further, we can immediately note that a natural prediction will be a spectrum of perturbations that are nearly *scale invariant*. This means that the metric fluctuations of spacetime receive equal levels of distortion from each decade of perturbation wavelength, and may be quantified for subsequent density fluctuations.

(2) 'Outside the horizon': $a/k \gg c/H$. Now the mode has a wavelength that exceeds the scale over which causal influences can operate. Therefore, it must now act as a 'frozen' quantity, which has the character of a classical disturbance. This field fluctuation can act as the seed

(1) 'Inside the horizon': $a/k \ll c/H$. Here the de Sitter expansion is negligible, just as we neglect the modern vacuum energy in the Solar system. The fluctuations in ϕ can be calculated exactly as in flat-space quantum field theory.

comoving Fourier modes, then there are two interesting limits for the mode wavelength: by c/H can never communicate with each other. If we imagine expanding the inflation, ϕ , using expansion literally makes distant regions of space move faster than light, so that points separated With $R \propto \exp(Ht)$, the proper radius of the horizon is given by $R_0 r_{\text{rh}} = c/H$. The exponential this is not to be confused with the particle horizon, where the integral would be between 0 and t_0 .

$$(140) \quad r_{\text{rh}} = \int_{t_0}^{\infty} \frac{c}{a} dt;$$

$t = \infty$ is finite, an **event horizon**, in that the comoving distance that particles can travel between a time t_0 and The essence of the idea can be seen in figure 12. This reminds us that de Sitter space contains

Cambridge. It was realized very quickly after the invention of inflation that the theory might also solve the other big puzzle with the initial condition of the universe. When we study gravitational instability, we will see that the present-day structure requires that the universe at even the Planck era would have had to possess a finite degree of inhomogeneity. Inflation suggests an audacious explanation for this structure, which is that it is an amplified form of the quantum fluctuations that are inevitable when the universe is sufficiently small. The present standard theory of this process was worked out by a number of researchers and generally agreed at a historic 1982 Nuffield conference in

5.2 Relic fluctuations from inflation

which absorbs the box size into the definition of a dimensionless power spectrum, which gives the contribution to the variance from each logarithmic range of wavenumber (or wavelength). Despite the attraction of a dimensionless quantity, one still frequently sees plots of $P(k)$ – and often in a dimensionally fudged form in which $L = 1$ is assumed, and P given units of volume.

$$(139) \quad \Delta_2^2(k) \equiv \frac{L^3}{(2\pi)^3} 4\pi k^3 P(k),$$

In the last step, we have defined the combination

$$(138) \quad \langle \delta^2 \rangle = \sum |\delta^k|^2 \leftarrow \frac{L^3}{(2\pi)^3} \int P(k) d^3k = \int \Delta_2^2(k) d \ln k.$$

over modes is replaced by an integral over k space times the usual density of states, $(L/2\pi)^3$ to appear. This happens naturally: as the box becomes big, the modes are finely spaced and a sum is statistically **isotropic**. In applying this apparatus, we would not want the (arbitrary) box size independent of direction of the wavevector in the limit of a large box: the fluctuating density field is called the **power spectrum**. Note that, in an isotropic universe, we assume that P will be

$$(137) \quad P(k) \equiv |\delta^k|^2$$

that all cross terms integrate to zero via the boundary conditions. For obvious reasons, the quantity To see this result, write the lhs instead as $\langle \delta^{*} \delta \rangle$ (makes no difference for a real field), and appreciate

$$(136) \quad \langle \delta^2 \rangle = \sum |\delta^k|^2.$$

The fluctuating density field can be described by its statistical properties. The mean is zero by construction; the variance is obtained by taking the volume average of δ^2 :

$$(135) \quad \Phi_k = -4\pi G(1 + 3w)a^2 \rho \delta^k/k^2.$$

Working in Fourier space in this way is powerful because it immediately gives a way of solving Poisson's equation and relating fluctuations in density and potential. For a single mode, $\Delta_2^2 \rightarrow -k^2$, and so

$$\delta^k(\mathbf{k}) = \int_3 \delta(\mathbf{x}) \exp(i\mathbf{k} \cdot \mathbf{x}) d^3x. \quad (143)$$

To get a feeling for what this means, consider the case of a matter distribution built up by the random placement of particles. It is not hard to show that this corresponds to **white noise**: a power spectrum that is independent of scale – i.e. $n = 0$. Recall the inverse Fourier relation:

$$P(k) \propto k^n; \quad n = 1. \quad (144)$$

So the density field is very strongly inhomogeneous on small scales. Another way of putting this is in terms of a standard power-law notation for the non-dimensionless spectrum:

$$\Delta_2(k) \propto k^4, \quad (145)$$

If we accept this rough argument, then the implied density power spectrum is interesting because of the relation between potential and density, it must be

The intuitive argument for scale invariance is that de Sitter space is invariant under time translation: there is no natural origin of time under exponential expansion. At a given time, the only length scale in the model is the horizon size c/H , so it is inevitable that the fluctuations exist on this scale are the same at all times. By our causality argument, these metric fluctuations must be copied unchanged to larger scales as the universe exponentiates, so that the appearance of the universe is independent of the scale at which it is viewed.

(where we have taken a $w = 0$ pressureless equation of state). This says that $\Phi_k/c^2 \sim \delta_k$ when the reciprocal of the physical wavenumber is c/H , i.e. is of order the horizon size.

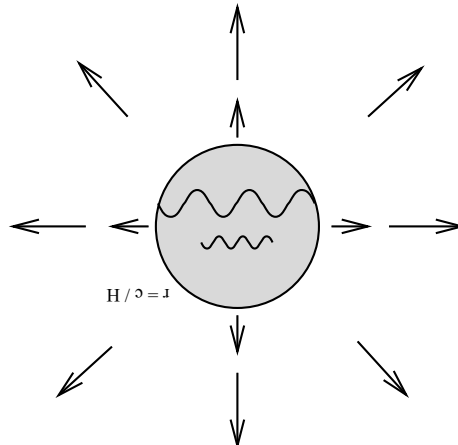
$$\Phi_k = -\frac{a^2}{2} 4\pi G \rho \delta_k = -(3/2) \frac{a^2}{2} \Omega_m H^2 \delta_k \quad (146)$$

The origin of the term 'scale-invariant' is clear: since potential fluctuations modify spacetime, this is equivalent to saying that spacetime must be a fractal: it has the same level of deviation from the exact RW form on each level of resolution. It is common to denote the level of metric fluctuations by δ_h – the **horizon-scale amplitude** (which we know to be about 10^{-5}). The justification for this name is that the potential perturbation is of the same order as the density fluctuation on the scale of the horizon at any given time. We can see this from Poisson's equation in Fourier space:

$$\Delta_2^2 \Phi \equiv \frac{d^2 \sigma^2(\Phi)}{d \ln k} = \text{constant} \equiv \delta_h^2. \quad (147)$$

in terms of the dimensionless power spectrum, Δ_2^2 , of the Newtonian gravitational potential, Φ ($c = 1$):

Figure 12. The event horizon in de Sitter space. Particles outside the sphere at $r = c/H$ can never receive light signals from the origin, nor can an observer at the origin receive information from outside the sphere. The exponential expansion quickly accelerates any freely falling observers to the point where their recession from the origin is effectively superluminal. The wave trains represent the generation of fluctuations in this spacetime. Waves with $\lambda \gg c/H$ effectively occupy flat space, and so undergo the normal quantum fluctuations for a vacuum state. As these modes (of fixed comoving wavelength) are expanded to sizes $\gg c/H$, causality forces the quantum fluctuation to become frozen as a classical amplitude that can seed large-scale structure.



Having seen that the inflation perturbation behaves in this way, it is not much work to obtain the quantum fluctuations that result in the field at late times (i.e. on scales much larger than the Sitter horizon). First consider the fluctuations in flat space on scales well inside the horizon. In principle, this requires quantum field theory, but the vacuum fluctuations in ϕ can be derived by a simple argument using the uncertainty principle. First of all, note that the sub-horizon equation

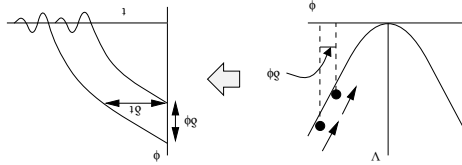
$$[\delta\phi] + 3H[\delta\phi] + (k/v)^2[\delta\phi] = 0, \quad (150)$$

and we seek the corresponding equation for the perturbation $\delta\phi$ obtained by starting inflation with slightly different values of ϕ in different places. Suppose this perturbation takes the form of a comoving plane-wave perturbation of comoving wavenumber k and amplitude A : $\delta\phi = A \exp(i\mathbf{k} \cdot \mathbf{x} - ikt/v)$. If the slow-roll conditions are also assumed, so that V' may be treated as a constant, then the perturbed field $\delta\phi$ obeys the first-order perturbation of the equation of motion for the main field:

$$\ddot{\phi} + 3H\dot{\phi} - \Delta^2\phi + V'(\phi) = 0, \quad (149)$$

The last step uses the crucial input of quantum field theory, which says that the rms $\delta\phi$ is given by $H/2\pi$, and we now sketch the derivation of this result. What we need to do is consider the equation of motion obeyed by perturbations in the inflaton field. The basic equation of motion is

Figure 13. This plot shows how fluctuations in the scalar field transform themselves into density fluctuations at the end of inflation. Different points of the universe inflate from different starting points. Inflation finishes at times separated by δt in time for these two points, inducing a density fluctuation $\delta = H\delta t$.



59

The $\delta_H \approx H \delta t$ argument relies on $R(t) \propto \exp(Ht)$ and that δ_H is of order the fractional change in R . We will not attempt here to do better than justify the order of magnitude.

$$\delta_H \approx H \delta t = \frac{\phi}{H} \delta\phi = \frac{\phi}{H} \times \frac{2\pi}{H} = \frac{2\pi\phi}{H^2}. \quad (148)$$

These universes will then finish inflation at different times, leading to a spread in energy densities (figure 13). The horizon-scale density amplitude is given by the different amounts that the universes have expanded following the end of inflation:

$$\delta t = \frac{\phi}{\delta\phi}. \quad (147)$$

which can be understood as follows. Imagine that the main effect of fluctuations is to make different parts of the universe have fields that are perturbed by an amount $\delta\phi$. In other words, we are dealing with various copies of the same rolling behaviour $\phi(t)$, but viewed at different times

$$\delta_H = \frac{2\pi\phi}{H^2}. \quad (146)$$

To anticipate the final answer, the inflationary prediction is of a horizon-scale amplitude

A MORE DETAILED TREATMENT We now need to give an outline of the exact treatment of inflationary fluctuations, which will allow us to calculate both the scale dependence of the spectrum and the absolute level of fluctuations. This can be a pretty technical subject, but it is possible to take a simple approach and still give a flavour of the main results and how they arise.

This $n = 1$ spectrum was considered a generic possibility long before inflation, and is also known as the **Zeldovich spectrum**. It is possible to alter this prediction of scale invariance only if the expansion is non-exponential; but we have seen that such deviations must exist towards the end of inflation. As we will see, it is natural for n to deviate from unity by a few %, and this is one of the predictions of inflation.

Here, the density field is a sum of spikes at the locations of particles. Because the placement is random, the contribution of each spike is a complex number of phase uniformly distributed between 0 and 2π , independent of k . Conversely, the $n = 1$ 'scale-invariant' spectrum thus represents a density field that is super-uniform on large scales, but with enhanced small-scale fluctuations.

This completes the argument. The initial quantum zero-point fluctuations in the field have been transcribed to a constant classical fluctuation that can eventually manifest itself as large-scale structure. The rms value of fluctuations in ϕ can be used as above to deduce the power in the field variance; $4\pi k^2 dk = 4\pi k^3 d \ln k$ comes from the k -space volume element).

(the factor $V/(2\pi)^3$ comes from the density of states in the Fourier transform, and cancels the $1/V$)

$$(159) \quad \frac{d(\delta\phi)_z^2}{H^2} = \frac{p \ln k}{4\pi k^3 V} \langle 0 | |\phi_k|^2 | 0 \rangle = \left(\frac{2\pi}{H} \right)^2$$

where we have emphasised that this is the vacuum expectation value. The fluctuations in ϕ depend on k in such a way that the fluctuations per decade are constant:

$$(158) \quad \langle 0 | |\phi_k|^2 | 0 \rangle = \frac{2\pi^3 V}{H^2}$$

At the opposite extreme, $aH/k \gg 1$, the squared fluctuation amplitude becomes frozen out at the value

$(d/dt)[k/aH] = -k/a$, we see that the oscillatory term has a leading dependence on t of the desired form. (remember that H is a constant, so that $(d/dt)[aH] = \dot{H} = aH^2$ etc.). At early times, when the horizon is much larger than the wavelength, $aH/k \gg 1$, and so this expression is the flat-space result, except that the time dependence looks a little odd, being $\exp(ik/aH)$. However, since $(d/dt)[k/aH] = -k/a$, we see that the oscillatory term has a leading dependence on t of the desired

$$(157) \quad \delta\phi = (a^3 V)^{-1/2} (2k/a)^{-1/2} e^{ik/aH} (1 + iaH/k)$$

With this boundary condition, it straightforward to check by substitution that the following expression satisfies the evolution equation:

This is the correct expression that results from a full treatment in quantum field theory.

$$(156) \quad \delta\phi = (a^3 V)^{-1/2} (2k/a)^{-1/2} e^{-ik/a}$$

so we adopt an rms field amplitude from quantum fluctuations of

61

$$(155) \quad \langle (\delta\phi)_z^2 \rangle = (2(a^3 V)(k/a))^{-1},$$

as

The uncertainty principle therefore gives us the variance of the zero-point fluctuations in ϕ

and the 'momentum' conjugate to ϕ is $p = a^3 \dot{\phi}$. In the current case, p is a momentum *density*, since \mathcal{L} is a Lagrangian density; we should therefore multiply p by a comoving volume V , so the analogue of the SHO mass is $m = a^3 V$.

$$(154) \quad \mathcal{L}^{\text{kinetic}} = a^3 \dot{\phi}^2 / 2,$$

For scalar field fluctuations, our 'coordinate' q is just the field ϕ , the oscillator frequency is $\omega = k/a$, and we now revert to $\eta = 1$. What is the analogue of the mass of the oscillator in this case? Recall that a Lagrangian, L , has a momentum $p = \partial L / \partial \dot{q}$ corresponding to each coordinate. For the present application, the kinetic part of the Lagrangian density is

which yields the required result.

$$(153) \quad m^2 \omega^2 q_{\text{rms}}^2 = \eta^2 / 4,$$

The uncertainty principle therefore says
but in any case the amplitudes in momentum and coordinate must be related by $p_{\text{rms}} = m\omega q_{\text{rms}}$.
uncertainty can be thought of as saying that we lack a knowledge of the amplitude of the oscillator, Quantum
For a classical oscillation with $q(t) \propto e^{i\omega t}$, the momentum is $p(t) = m\dot{q} = i\omega m q(t)$.

$$(152) \quad \langle (\delta p)_z^2 \rangle \langle (\delta q)_z^2 \rangle = \eta^2 / 4.$$

This can be derived immediately from the uncertainty principle, which says that the minimum uncertainty is

$$(151) \quad q_{\text{rms}} = \left(\frac{2m\omega}{\eta} \right)^{1/2}.$$

oscillator of mass m and position coordinate q , the rms uncertainty in q in the ground state is
of motion is just that for a simple harmonic oscillator: $[\delta\dot{\phi}]^2 + \omega^2 [\delta\phi]^2 = 0$, where $\omega = k/a$. For an

$$(167) \quad \delta_H \sim \left(\frac{m_{\text{GUT}}}{m_p} \right)^2.$$

We have argued that inflation will end with ϵ of order unity; if the potential were to have the characteristic value $V \sim E_{\text{GUT}}^4$ then this would give the simple result

$$(166) \quad \delta_H \sim \frac{m_p^2 \epsilon^{1/2}}{V^{1/2}}.$$

These constraints appear to suggest a defect in inflation, in that we should be able to use the theory to *explain* why $\delta_H \sim 10^{-5}$, rather than using this observed fact to constrain the theory. The amplitude of δ_H is one of the most important numbers in cosmology, and it is vital to know whether there is a simple explanation for its magnitude. One way to view this is to express the horizon-scale amplitude as

$$(165) \quad m \lesssim 10^{-5} m_p.$$

Alternatively, in the case of $V = m_p^2 \phi^2$, $\delta_H = 3H^3/(2m_p^2 \phi)$. Since $H \sim \sqrt{V}/m_p$, this gives $\delta_H \sim m_p^2/m_p^3 \sim 10^{-5}$. Since we have already seen that $\phi \gtrsim m_p$ is needed for inflation, this gives

$$(164) \quad \lambda \lesssim 10^{-15}.$$

Since $N \gtrsim 60$, the observed $\delta_H \sim 10^{-5}$ requires

$$(163) \quad \delta_H \sim \frac{\phi}{3H^2} = \frac{V'}{\lambda^{1/2} N^{3/2}}.$$

Suppose $V(\phi)$ takes the form $V = \lambda \phi^4$, so that $N = H^2/(2\lambda \phi^2)$. The density perturbations can then be expressed as

$$(162) \quad N = \int H dt = \int H d\phi / \dot{\phi} = \int 3H^2 d\phi / V'.$$

INFLATION COUPLING The calculation of density inhomogeneities sets an important limit on the inflation potential. From the slow-rolling equation, we know that the number of e -foldings of inflation is

But unlike fluctuations in the inflaton, the tensor fluctuation do not affect the progress of inflation: once generated, they play no further part in events and survive to the present day. Detection of these primordial tensor perturbations would not only give confidence in the basic inflationary picture, but would measure rather directly the energy scale of inflation.

$$(161) \quad h_{\text{rms}} \sim H/m_p.$$

What value of h does inflation predict? For scalar perturbations, small-scale quantum fluctuations lead to an amplitude $\delta\phi = H/2\pi$ on horizon exit, which transforms to a metric fluctuation $\delta h = H\delta\phi/\dot{\phi}$. Tensor modes behave similarly – except that h must be dimensionless, whereas δh has dimensions of mass. On dimensional grounds, then, the formula for the tensor fluctuations is plausible:

TENSOR PERTURBATIONS Later in the course, we will compare the predictions of this inflationary apparatus with observations of the fluctuating density field of the contemporary universe. It should be emphasised again just what an audacious idea this is: that all the structure around us was seeded by quantum fluctuations while the universe was of subnuclear scale. It would be nice if we could verify this radical assumption, and there is one basic test: if the idea of quantum fluctuations is correct, it should apply to every field that was present in the early universe. In particular, it should apply to the gravitational field. This corresponds to metric perturbations in the form of a tensor h_{TT} , whose coefficients have some typical amplitude h (not the Hubble parameter). This **spatial strain** is what is measured by gravity-wave telescopes such as LIGO: the separation between a pair of freely-suspended masses changes by a fractional amount of order h as the wave passes. These experiments can be fabulously precise, with a current sensitivity of around $h = 10^{-21}$.

where we have also written once again the slow-roll condition and the corresponding relation between H and V , since manipulation of these three equations is often required in derivations.

$$(160) \quad \delta_H^2 \equiv \Delta_{\frac{1}{2}}^{\Phi}(k) = \frac{H^4}{8\pi^2 V} \frac{3}{m_p^2}, \quad 3H\dot{\phi} = -V',$$

spectrum of mass fluctuations well after inflation is over. In terms of the variance per h in potential perturbations, the answer is

The predictions of simple chaotic inflation are thus very close to scale invariance in practice: $n = 0.97$ for $\alpha = 2$ and $n = 0.95$ for $\alpha = 4$. However, such a tilt has a significant effect over the several decades in k from CMB anisotropy measurements to small-scale galaxy clustering. These results are in some sense the default inflationary predictions: exact scale invariance would be surprising, as would large amounts of tilt. Either observation would indicate that the potential must have a more complicated structure, or that the inflationary framework is not correct.

$$(176) \quad 1 - n = (2 + \alpha)/120.$$

It is easy to see that $\phi_{\text{end}} \gg \phi$ and that $\epsilon = \alpha/240$, leading finally to

$$(175) \quad 60 = \frac{8\pi}{\phi} \int_{\phi}^{\phi_{\text{end}}} \frac{m_p^2}{\phi} d\phi = \frac{4\pi}{\phi^2} \frac{m_p^2}{\phi^2} (\phi^2 - \phi_{\text{end}}^2).$$

A better motivated choice than power-law inflation would be a power-law potential $V(\phi) \propto \phi^\alpha$; many chaotic inflation models concentrate on $\alpha = 2$ (mass-like term) or $\alpha = 4$ (highest renormalizable power). Here, $\epsilon = m_p^2 \alpha^2 / (16\pi \phi^2)$, $n = 1 - 1/\alpha$, and

$$(174) \quad 60 = \int H dt = \int_{\phi}^{\phi_{\text{end}}} \frac{8\pi}{\phi} \frac{m_p^2}{V'} d\phi.$$

In general, however, the inflation derivatives have to be evaluated explicitly on the largest scales, 60 e -foldings prior to the end of inflation, so that we need to solve

$$(173) \quad 1 - n = 2/p$$

It is interesting to put flesh on the bones of this general expression and evaluate the tilt for some specific inflationary models. This is easy in the case of power-law inflation with $a \propto t^p$ because the inflation parameters are constant: $\epsilon = \eta/2 = 1/p$, so that the tilt here is always

65

For most models in which the potential is a smooth polynomial-like function, $|\eta| \lesssim |\epsilon|$. Since ϵ has the larger coefficient and is positive by definition, the simplest inflation models tend to predict that the spectrum of scalar perturbations should be slightly tilted, in the sense that n is slightly less than unity.

$$(172) \quad 1 - n = 6\epsilon - 2\eta$$

and derivatives of V can be expressed in terms of the dimensionless parameters ϵ and η . The tilt of the density perturbation spectrum is thus predicted to be

$$(171) \quad \delta_2^2 = \frac{H^4}{128\pi^3} \left(\frac{m_p^2}{V^3} \right) \left(\frac{2\pi\phi^2}{V^2} \right),$$

We can now work out the tilt, since the horizon-scale amplitude is

$$(170) \quad \frac{d \ln k}{d} = a \frac{da}{d} = \frac{H}{\phi} \frac{dH}{d\phi} = -\frac{8\pi}{m_p^2} \frac{V'}{V} d\phi.$$

Since H is nearly constant during the inflationary evolution, we can replace $d/d \ln k$ by $d \ln a$, and use the slow-roll condition to obtain

$$(169) \quad a/k = H^{-1}.$$

We then want to express the tilt in terms of parameters of the inflationary potential, ϵ and η . These are of order unity when inflation terminates; ϵ and η must therefore be evaluated when the observed universe left the horizon, recalling that we only observe the last 60-odd e -foldings of inflation. The way to introduce scale dependence is to write the condition for a mode of given comoving wavenumber to cross the de Sitter horizon,

$$(168) \quad \text{tilt} \equiv 1 - n \equiv -\frac{d \ln \delta_2^2}{d \ln k}.$$

Finally, deviations from exact exponential expansion must exist at the end of inflation, and the corresponding change in the fluctuation power spectrum is a potential test of inflation. Define the **tilt** of the fluctuation spectrum as follows:

These fluctuations in the scalar field can affect the progress of inflation itself. They can be thought of as adding a random-walk element to the classical rolling of the scalar field down the trough defined by $V(\phi)$. In cases where ϕ is too close to the origin for inflation to persist for sufficiently long, it is possible for the quantum fluctuations to push ϕ further out – creating further inflation in a self-sustaining process. This is the concept of **stochastic inflation**.

Consider the scalar field at a given point in the inflationary universe. Each e -folding of the expansion produces new classical fluctuations, which add incoherently to those previously present. If the field is sufficiently far from the origin in a polynomial potential, these fluctuations produce a random walk of $\phi(t)$ that overwhelms the classical trajectory in which ϕ tries to roll down the potential, as follows. The classical amplitude from quantum fluctuations is $\delta\phi = H/2\pi$, and a new disturbance of the same rms will be added for every $\Delta t = 1/H$. The slow-rolling equation says that the trajectory is $\dot{\phi} = -V'/3H$; we also have $H^2 = 8\pi V/3m_p^2$, so that the classical change in ϕ is $\Delta\phi = -m_p^2 V'/8\pi V$ in a time $\Delta t = 1/H$. Consider $V = \lambda|\phi|^n/(m_p^{n-4})$, for which these two changes in ϕ will be equal at $\phi^* \sim m_p/\lambda^{1/(n+2)}$. For smaller ϕ , the quantum fluctuations will have a negligible effect on the classical trajectory; for larger ϕ , the equation of motion will become stochastic. The resulting random walk will send some parts of the universe to ever larger values of ϕ , so inflation never entirely ends. This **eternal inflation** is the basis for the concept of the **inflationary multiverse**: different widely-separated parts of the universe will inflate by different amounts, producing in effect separate universes with distinct formation histories.

6 Structure formation – I

6.1 Newtonian equations of motion

We have decided that perturbations will in most cases effectively be described by the Newtonian potential, Φ . Now we need to develop an equation of motion for Φ , or equivalently for the density fluctuation $\rho \equiv (1 + \delta)\rho$. In the Newtonian approach, we treat dynamics of cosmological matter exactly as we would in the laboratory, by finding the equations of motion induced by either pressure or gravity. We begin by casting the problem in comoving units:

$$(177) \quad \begin{aligned} \mathbf{x}(t) &= a(t)\mathbf{r}(t) \\ \delta\mathbf{v}(t) &= a(t)\mathbf{u}(t), \end{aligned}$$

67

so that \mathbf{x} has units of proper length, i.e. it is an **Eulerian coordinate**. First note that the comoving peculiar velocity \mathbf{u} is just the time derivative of the comoving coordinate \mathbf{r} :

$$(178) \quad \mathbf{x} = a\mathbf{r} + a\mathbf{r}' = H\mathbf{x} + a\mathbf{r}'$$

where the rhs must be equal to the Hubble flow $H\mathbf{x}$, plus the peculiar velocity $\delta\mathbf{v} = a\mathbf{u}$.

The equation of motion follows from writing the Eulerian equation of motion as $\ddot{\mathbf{x}} = \mathbf{g}_0 + \mathbf{g}$, where $\mathbf{g} = -\Delta\Phi/a$ is the peculiar acceleration, and \mathbf{g}_0 is the acceleration that acts on a particle in a homogeneous universe (neglecting pressure forces to start with, for simplicity). Differentiating $\mathbf{x} = a\mathbf{r}$ twice gives

$$(179) \quad \ddot{\mathbf{x}} = a\ddot{\mathbf{u}} + 2\dot{a}\dot{\mathbf{u}} + \frac{d}{dt}\mathbf{x} = \mathbf{g}_0 + \mathbf{g}.$$

The unperturbed equation corresponds to zero peculiar velocity and zero peculiar acceleration: $(a/a)\mathbf{x} = \mathbf{g}_0$; subtracting this gives the perturbed equation of motion

$$(180) \quad \ddot{\mathbf{u}} + 2(\dot{a}/a)\dot{\mathbf{u}} = \mathbf{g} = -\Delta\Phi/a.$$

This equation of motion for the peculiar velocity shows that \mathbf{u} is affected by gravitational acceleration and by the **Hubble drag** term, $2(\dot{a}/a)\dot{\mathbf{u}}$. This arises because the peculiar velocity falls with time as a particle attempts to catch up with successively more distant (and therefore more rapidly receding) neighbours. In the absence of gravity, we get $\delta v \propto 1/a$: momentum redshifts away, just as with photon energy.

The peculiar velocity is directly related to the evolution of the density field, through conservation of mass. If we restrict ourselves to a linear approximation where $\delta \ll 1$, the linearized continuity equation for conservation of momentum and matter as experienced by fundamental observers moving with the Hubble flow is:

$$(181) \quad \delta = -\Delta \cdot \mathbf{u}.$$

(see e.g. Section 15.3 of Peacock 1999 for the details).

The solutions of these equations can be decomposed into modes either parallel to \mathbf{g} or independent of \mathbf{g} (these are the homogeneous and inhomogeneous solutions to the equation of motion). The homogeneous case corresponds to no peculiar gravity – i.e. zero density perturbation. This is consistent with the linearized continuity equation, $\Delta \cdot \mathbf{u} = -\delta$, which says that it is possible

This is accurate, but still hard to remember. For flat models with $\Omega_m + \Omega_v = 1$, a simpler approximation is $f \approx \Omega_{0.23}^{0.65}$, which is less marked than $f \approx \Omega_{0.65}^{0.65}$ in the $\Lambda = 0$ case. This reflects the more rapid variation of Ω_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 Ω_m .

$$(188) \quad f(\Omega) \approx \frac{2}{5} \Omega_m \left[\Omega_{4/7}^m \Omega_v + (1 + \frac{2}{5} \Omega_m) (1 + \frac{7}{4} \Omega_v) \right]^{-1}.$$

where the factor f expresses a deviation from the simple growth law. The case of matter + cosmological constant is of the most common practical interest, and a very good approximation to the answer is given by Carroll et al. (1992):

$$(187) \quad \delta(a) \propto a f[\Omega(a)],$$

MODELS WITH NON-CRITICAL DENSITY We have solved the growth equation for the matter-dominated $\Omega = 1$ case. It is possible to cope with other special cases (e.g. matter + curvature) with some effort. In the general case (especially with a general vacuum having $w \neq -1$), it is necessary to integrate the differential equation numerically. At high z , we always have the matter-dominated $\delta \propto a$, and this serves as an initial condition. In general, we can write

in either case. In other words, the metric fluctuations resulting from potential perturbations are frozen, at least for perturbations with wavelengths greater than the horizon size.

$$(186) \quad \Phi = \text{constant}$$

The results for matter domination and radiation domination can be combined to say that gravitational potential perturbations are independent of time (at least while $\Omega = 1$). Poisson's equation tells us that $-k^2 \Phi / a^2 \propto \rho$; since $\rho \propto a^{-3}$ for matter domination or a^{-4} for radiation, that gives $\Phi \propto \delta / a$ or δ / a^2 respectively, so that

The growing mode during radiation domination ($a \propto t^{1/2}$) has $\delta(t) \propto a(t)^2$.

69

$$(185) \quad \delta(t) \propto t \quad \text{or} \quad t^{-1}.$$

For the flat case, a power-law solution gives:

$$(184) \quad \delta + 2 \frac{a}{\dot{a}} \dot{\delta} = \frac{32\pi}{3} G \rho_0 \delta,$$

RADIATION-DOMINATED UNIVERSE The analysis so far does not apply when the universe was radiation dominated ($c_s = c/\sqrt{3}$). For this period of the early Universe it is therefore common to resort to general relativity perturbation theory or use special relativity fluid mechanics and Newtonian gravity with a relativistic source term (see e.g. Section 15.2 of Peacock 1999). In the interests of brevity and completeness we simply quote the result of this analysis where the resulting evolution equation for δ has a driving term on the rhs that is a factor $8/3$ higher than in the matter-dominated case

The first solution, with $\delta(t) \propto a(t)$ is the growing mode, corresponding to the gravitational instability of density perturbations. Given some small initial seed fluctuations, this is the simplest way of creating a universe with any desired degree of inhomogeneity.

$$(183) \quad \delta(t) \propto t^{2/3} \quad \text{or} \quad t^{-1}.$$

works:

This is easily solved for the $\Omega_m = 1$ case, where $4\pi G \rho_0 = 3H^2/2 = 2/3t^2$, and a power-law solution

$$(182) \quad \delta + 2 \frac{a}{\dot{a}} \dot{\delta} = 4\pi G \rho_0 \delta.$$

GROWING MODE For the growing mode, it is most convenient to eliminate \mathbf{u} by taking the divergence of the equation of motion for \mathbf{u} , and the time derivative of the continuity equation. This requires a knowledge of $\nabla \cdot \mathbf{g}$, which comes via Poisson's equation: $\nabla \cdot \mathbf{g} = 4\pi G a \rho_0 \delta$. The resulting 2nd-order equation for δ is

in this case. The proper velocities of these vorticity modes decay as $v = a u \propto a^{-1}$, as with the kinematic analysis for a single particle.

So far, we have mainly considered the collisionless component. For the photon-baryon gas, all that changes is that the peculiar acceleration gains a term from the pressure gradients:

$$\mathbf{g} = -\nabla\Phi/a - \nabla p/(a\rho). \quad (189)$$

The pressure fluctuations are related to the density perturbations via the sound speed

$$c_s^2 \equiv \frac{\partial p}{\partial \rho}. \quad (190)$$

Now think of a plane-wave disturbance $\delta \propto e^{-i\mathbf{k}\cdot\mathbf{r}}$, where \mathbf{k} and \mathbf{r} are in comoving units. All time dependence is carried by the amplitude of the wave. The linearized equation of motion for δ then gains an extra term on the rhs from the pressure gradient:

$$\ddot{\delta} + 2\frac{v}{a}\dot{\delta} = \delta(4\pi G\rho_0 - c_s^2 k^2/a^2). \quad (191)$$

This shows that there is a critical proper wavelength, the **Jeans length**, at which we switch from the possibility of gravity-driven growth for long-wavelength modes to standing sound waves at short wavelengths. This critical length is

$$\lambda_{\text{proper}}^J = \frac{2\pi}{k_{\text{proper}}^J} = c_s \sqrt{\frac{G\rho}{\pi}}. \quad (192)$$

It is interesting to work out the value of this critical length. Consider a universe with a coupled photon-baryon fluid and ignore dark matter (which we can do at high redshifts, near matter-radiation equality). The sound speed, $c_s^2 = \partial p/\partial \rho$, may be found by thinking about the response of matter and radiation to small adiabatic compressions:

$$\delta p = (4/9)c^2(\delta V/V), \quad \delta \rho = [dm + (4/3)\rho](\delta V/V), \quad (193)$$

implying

$$c_s^2 = c^2 \left(\frac{4}{9} \frac{\rho}{\rho_m} + \frac{4}{9} \right) = c^2 \left[\frac{4}{9} + \frac{4}{9} \left(\frac{1+z}{1+z_{\text{rad}}} \right) \right]^{-1}. \quad (194)$$

71

Here, z_{rad} is the redshift of equality between matter and photons; $1+z_{\text{rad}} = 1.68(1+z_{\text{eq}})$ because of the neutrino contribution. At $z \gg z_{\text{rad}}$, we therefore have $c_s \propto \sqrt{1+z}$. Since $\rho = (1+z)^3 \Omega_b H_0^2 / (8\pi G)$, the *comoving* Jeans length is constant at

$$\lambda_J = \frac{c}{32\pi^2} \left(\frac{27\Omega_b(1+z_{\text{rad}})}{32\pi^2} \right)^{1/2} = 50(\Omega_b h^2)^{-1/2} \text{ Mpc}. \quad (195)$$

This is of order the horizon size at matter-radiation equality. Smaller-scale fluctuations in the photon-baryon fluid will not have undergone steady gravitational growth, but will have oscillated with time as standing waves. We will see later that the imprint of these oscillations is visible in the microwave background.

6.3 The shape of the matter power spectrum

Figure 14 shows a schematic of how a density fluctuation grows in the early Universe. For scales greater than the horizon, perturbations in matter and radiation can grow together, so fluctuations at early times grow at the same rate, independent of wavenumber. But this growth ceases once the perturbations enter the horizon – i.e. when the horizon grows sufficiently to exceed the perturbation wavelength. At this point, growth ceases. For fluids (baryons) it is the radiation pressure that prevents the perturbations from collapsing further. For collisionless matter the rapid radiation driven expansion prevents the perturbation from growing again until matter radiation equality.

This effect (called the Mészáros effect) is critical in shaping the late-time power spectrum (as will show) as the universe preserves a ‘snapshot’ of the amplitude of the mode at horizon crossing. Before this process operates, inflation predicts a scale invariant initial Zelovich spectrum (where $P^i(k) \propto k$). How does the Mészáros effect modify the shape of this initial power spectrum?

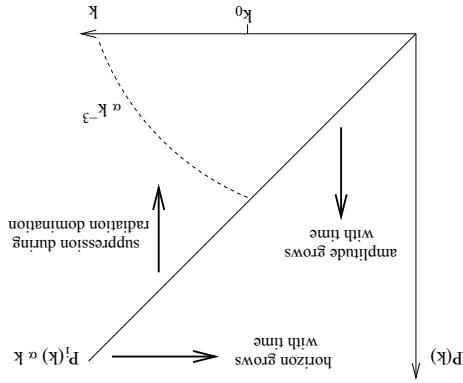
Figure 15 shows that the smallest physical scales (largest k scales) will be affected first and experience the strongest suppression to their amplitude. The largest physical scale fluctuations (smallest k scales) will be unaffected as they will enter the horizon after matter-radiation equality. We can therefore see that there will be a turnover in the power spectrum at a characteristic scale given by the horizon size at matter-radiation equality.

From Figure 14 we can see that when a fluctuation enters the horizon before matter-radiation equality its growth is suppressed by $f = (a^{\text{enter}}/a^{\text{eq}})^2$. A fluctuation k enters the horizon when

We have discussed the main effects that contribute to the form of the transfer function, but a full calculation is a technical challenge. In detail, we have a mixture of matter (both collisionless dark particles and baryonic plasma) and relativistic particles (collisionless neutrinos and collisionless photons), which does not behave as a simple fluid. Particular problems are caused by the change in the photon component from being a fluid tightly coupled to the baryons by Thomson scattering, to being collisionless after recombination. Accurate results require a solution of the Boltzmann equation to follow the evolution of the full phase-space distribution. This was first computed accurately by Bond & Szalay (1983), and is today routinely available via public-domain codes such as CMBFAST.

In principle, there is a transfer function for each constituent of the universe, and these evolve with time. As we have discussed, however, the different matter ingredients tend to come together at late times, and the overall transfer function tends to something that is the same for all matter components, and which does not change with time for low redshifts. This late-time transfer function is therefore an important tool for cosmologists who want to predict observed properties of density fields in the current universe.

Figure 15. Schematic of the how the Meszaros effect modifies the initial power spectrum. Note log scale.



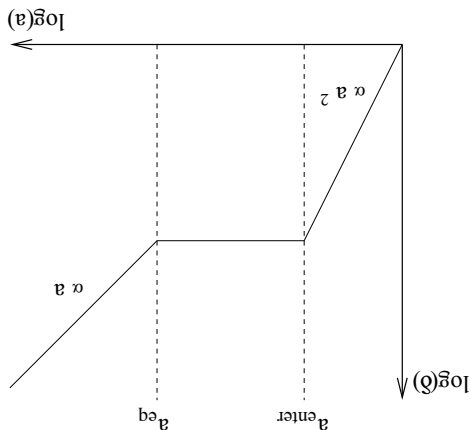
$$\delta(a) \propto g(a) T_k. \quad (196)$$

The above discussion can be summed up in the form of the linear **transfer function** for density perturbations, where we factor out the long-wavelength growth from a term that expresses how growth is modulated as a function of wavenumber:

6.4 Transfer functions and characteristic scales

$D_H \approx 1/k$. As $D_H = c/aH(a)$ and $H(a) \propto a^{-2}$ during radiation domination we see that the fluctuations are suppressed by a factor $f \propto k^{-2}$ and that the power spectrum on large k scales follows a k^{-3} power law.

Figure 14. A schematic of the suppression of fluctuation growth during the radiation dominated phase when the density perturbation enters the horizon at $a_{\text{enter}} > a_{\text{eq}}$.



A more interesting (and probably practically relevant) case is when the dark matter is a mixture of hot and cold components. The free-streaming length for the hot component can therefore mass must exceed about 1 keV.

$$\Omega_c h^2 = m/94.1 \text{ eV} \quad (199)$$

(in detail, the appropriate figure for neutrinos will be smaller by $(4/11)^{1/3}$ since they have a smaller temperature than the photons). A light neutrino-like relic that decouples while it is relativistic satisfies

$$L_{\text{free-stream}} = 112 (m/\text{eV})^{-1} \text{ Mpc} \quad (198)$$

translates to a comoving horizon scale $(2ct/a)$ during the radiation era) at $kT = mc^2$ of relativistic, so that perturbations are erased up to proper lengthscales of $\approx ct(kT = mc^2)$. This has entered the horizon will be damped. This process switches off when the particles become non-disperse. At early times ($kT > mc^2$), the particles will travel at c , and so any perturbation that perturbations can be erased simply by free streaming: random particle velocities cause blobs to more interesting case arises when thermal relics have lower masses. For collisionless dark matter, and structure formation proceeds via hierarchical process in which nonlinear structures grow via mergers. Examples of CDM would be thermal relic WIMPs with masses of order 100 GeV, but a CDM universe thus contains fluctuations in the dark matter on all scales, in which random fluctuations are negligible. A CDM universe thus contains fluctuations in the dark matter on all scales, and structure formation proceeds via hierarchical process in which nonlinear structures grow via mergers. Examples of CDM would be thermal relic WIMPs with masses of order 100 GeV, but a more interesting case arises when thermal relics have lower masses. For collisionless dark matter, perturbations can be 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, so that perturbations are erased up to proper lengthscales of $\approx ct(kT = mc^2)$. This translates to a comoving horizon scale $(2ct/a)$ during the radiation era) at $kT = mc^2$ of

fluctuations is all that applies to a universe dominated by Cold Dark Matter, in which random fluctuations are negligible. A CDM universe thus contains fluctuations in the dark matter on all scales, and structure formation proceeds via hierarchical process in which nonlinear structures grow via mergers. Examples of CDM would be thermal relic WIMPs with masses of order 100 GeV, but a more interesting case arises when thermal relics have lower masses. For collisionless dark matter, perturbations can be 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, so that perturbations are erased up to proper lengthscales of $\approx ct(kT = mc^2)$. This translates to a comoving horizon scale $(2ct/a)$ during the radiation era) at $kT = mc^2$ of

$$T_k \approx \begin{cases} 1 & k D_H(z_{\text{eq}})^{-2} \\ k D_H(z_{\text{eq}}) \gg 1 & k D_H(z_{\text{eq}}) \gg 1 \end{cases} \quad (197)$$

This process continues until the universe becomes matter dominated. We therefore expect a characteristic 'break' in the fluctuation spectrum around the comoving horizon length at this time, which we have seen is $D_H(z_{\text{eq}}) = 16 (\Omega_m h^2)^{-1} \text{ Mpc}$. Since distances in cosmology always scale as h^{-1} , this means that $\Omega_m h$ should be observable.

(2) Free-streaming length.

This relatively gentle filtering away of the initial fluctuations is all that applies to a universe dominated by Cold Dark Matter, in which random fluctuations are negligible. A CDM universe thus contains fluctuations in the dark matter on all scales, and structure formation proceeds via hierarchical process in which nonlinear structures grow via mergers. Examples of CDM would be thermal relic WIMPs with masses of order 100 GeV, but a more interesting case arises when thermal relics have lower masses. For collisionless dark matter, perturbations can be 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, so that perturbations are erased up to proper lengthscales of $\approx ct(kT = mc^2)$. This translates to a comoving horizon scale $(2ct/a)$ during the radiation era) at $kT = mc^2$ of

fluctuations is all that applies to a universe dominated by Cold Dark Matter, in which random fluctuations are negligible. A CDM universe thus contains fluctuations in the dark matter on all scales, and structure formation proceeds via hierarchical process in which nonlinear structures grow via mergers. Examples of CDM would be thermal relic WIMPs with masses of order 100 GeV, but a more interesting case arises when thermal relics have lower masses. For collisionless dark matter, perturbations can be 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, so that perturbations are erased up to proper lengthscales of $\approx ct(kT = mc^2)$. This translates to a comoving horizon scale $(2ct/a)$ during the radiation era) at $kT = mc^2$ of

fluctuations is all that applies to a universe dominated by Cold Dark Matter, in which random fluctuations are negligible. A CDM universe thus contains fluctuations in the dark matter on all scales, and structure formation proceeds via hierarchical process in which nonlinear structures grow via mergers. Examples of CDM would be thermal relic WIMPs with masses of order 100 GeV, but a more interesting case arises when thermal relics have lower masses. For collisionless dark matter, perturbations can be 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, so that perturbations are erased up to proper lengthscales of $\approx ct(kT = mc^2)$. This translates to a comoving horizon scale $(2ct/a)$ during the radiation era) at $kT = mc^2$ of

fluctuations is all that applies to a universe dominated by Cold Dark Matter, in which random fluctuations are negligible. A CDM universe thus contains fluctuations in the dark matter on all scales, and structure formation proceeds via hierarchical process in which nonlinear structures grow via mergers. Examples of CDM would be thermal relic WIMPs with masses of order 100 GeV, but a more interesting case arises when thermal relics have lower masses. For collisionless dark matter, perturbations can be 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, so that perturbations are erased up to proper lengthscales of $\approx ct(kT = mc^2)$. This translates to a comoving horizon scale $(2ct/a)$ during the radiation era) at $kT = mc^2$ of

fluctuations is all that applies to a universe dominated by Cold Dark Matter, in which random fluctuations are negligible. A CDM universe thus contains fluctuations in the dark matter on all scales, and structure formation proceeds via hierarchical process in which nonlinear structures grow via mergers. Examples of CDM would be thermal relic WIMPs with masses of order 100 GeV, but a more interesting case arises when thermal relics have lower masses. For collisionless dark matter, perturbations can be 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, so that perturbations are erased up to proper lengthscales of $\approx ct(kT = mc^2)$. This translates to a comoving horizon scale $(2ct/a)$ during the radiation era) at $kT = mc^2$ of

fluctuations is all that applies to a universe dominated by Cold Dark Matter, in which random fluctuations are negligible. A CDM universe thus contains fluctuations in the dark matter on all scales, and structure formation proceeds via hierarchical process in which nonlinear structures grow via mergers. Examples of CDM would be thermal relic WIMPs with masses of order 100 GeV, but a more interesting case arises when thermal relics have lower masses. For collisionless dark matter, perturbations can be 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, so that perturbations are erased up to proper lengthscales of $\approx ct(kT = mc^2)$. This translates to a comoving horizon scale $(2ct/a)$ during the radiation era) at $kT = mc^2$ of

fluctuations is all that applies to a universe dominated by Cold Dark Matter, in which random fluctuations are negligible. A CDM universe thus contains fluctuations in the dark matter on all scales, and structure formation proceeds via hierarchical process in which nonlinear structures grow via mergers. Examples of CDM would be thermal relic WIMPs with masses of order 100 GeV, but a more interesting case arises when thermal relics have lower masses. For collisionless dark matter, perturbations can be 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, so that perturbations are erased up to proper lengthscales of $\approx ct(kT = mc^2)$. This translates to a comoving horizon scale $(2ct/a)$ during the radiation era) at $kT = mc^2$ of

fluctuations is all that applies to a universe dominated by Cold Dark Matter, in which random fluctuations are negligible. A CDM universe thus contains fluctuations in the dark matter on all scales, and structure formation proceeds via hierarchical process in which nonlinear structures grow via mergers. Examples of CDM would be thermal relic WIMPs with masses of order 100 GeV, but a more interesting case arises when thermal relics have lower masses. For collisionless dark matter, perturbations can be 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, so that perturbations are erased up to proper lengthscales of $\approx ct(kT = mc^2)$. This translates to a comoving horizon scale $(2ct/a)$ during the radiation era) at $kT = mc^2$ of

fluctuations is all that applies to a universe dominated by Cold Dark Matter, in which random fluctuations are negligible. A CDM universe thus contains fluctuations in the dark matter on all scales, and structure formation proceeds via hierarchical process in which nonlinear structures grow via mergers. Examples of CDM would be thermal relic WIMPs with masses of order 100 GeV, but a more interesting case arises when thermal relics have lower masses. For collisionless dark matter, perturbations can be 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, so that perturbations are erased up to proper lengthscales of $\approx ct(kT = mc^2)$. This translates to a comoving horizon scale $(2ct/a)$ during the radiation era) at $kT = mc^2$ of

fluctuations is all that applies to a universe dominated by Cold Dark Matter, in which random fluctuations are negligible. A CDM universe thus contains fluctuations in the dark matter on all scales, and structure formation proceeds via hierarchical process in which nonlinear structures grow via mergers. Examples of CDM would be thermal relic WIMPs with masses of order 100 GeV, but a more interesting case arises when thermal relics have lower masses. For collisionless dark matter, perturbations can be 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, so that perturbations are erased up to proper lengthscales of $\approx ct(kT = mc^2)$. This translates to a comoving horizon scale $(2ct/a)$ during the radiation era) at $kT = mc^2$ of

Some illustrative results are shown in figure 16. Leaving aside the isocurvature models, all adiabatic cases have $T \rightarrow 1$ on large scales - i.e. there is growth at the universal rate (which is such that the amplitude of potential perturbations is constant until the vacuum starts to be important at $z \lesssim 1$). The different shapes of the functions can be understood intuitively in terms of a few special length scales, as follows:

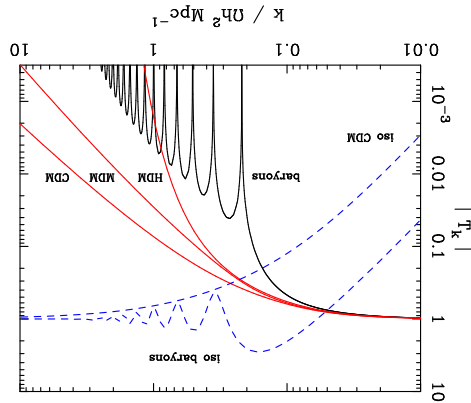


Figure 16. A plot of transfer functions for various adiabatic models, in which $T_k \rightarrow 1$ at small k . A number of possible matter contents are illustrated: pure barions; pure CDM; pure HDM. For dark-matter models, the characteristic wavenumber scales proportional to $\Omega_m h^2$, marking the break scale corresponding to the horizon length at matter-radiation equality. The scaling for barionic models does not obey this exactly; the plotted case corresponds to $\Omega_m = 1$, $h = 0.5$.

In outline, this is straightforward to solve, given some initial positions and velocities. Defining some timestep dt , particles are moved according to $d\mathbf{x} = \mathbf{u} dt$, and their velocities updated according to $d\mathbf{u} = \mathbf{a} dt$, with \mathbf{u} given by the equation of motion (in practice, more sophisticated time integration schemes are used). The hard part is finding the gravitational force, since this involves summation over $(N - 1)$ other particles each time we need a force for one particle. All the craft in the field involves finding clever ways in which all the forces can be evaluated in less than the raw $O(N^2)$ computations per timestep. We will have to omit the details of this, unfortunately, but one obvious way of proceeding is to solve Poisson's equation on a mesh using a Fast Fourier Transform. This can convert the $O(N^2)$ time scaling to $O(N \ln N)$, which is a qualitative difference given that N can be as large as 10^{10} .

where Φ is the Newtonian gravitational potential due to density perturbations. The time derivative is already in the required form of the convective time derivative observed by a particle, rather than the partial $\partial/\partial t$.

$$(203) \quad \rho \frac{d\mathbf{u}}{dt} = -\nabla \Phi,$$

The exact evolution of the density field is usually performed by means of an **N-body simulation**, in which the density field is represented by the sum of a set of fictitious discrete particles. We need to solve the equations of motion for each particle, as it moves in the gravitational field due to all the other particles. Using comoving units for length and velocity ($\mathbf{v} = a\mathbf{u}$), we have previously seen the equation of motion

The equations of motion are nonlinear, and we have only solved them in the limit of linear perturbations. We now discuss evolution beyond the linear regime, first considering the full numerical solution of the equations of motion, and then a key analytic approximation by which the exact results can be understood.

7 Structure formation – II

Unlike the power spectrum, $\sigma(R)$ is monotonic, and the value at any scale is sufficient to fix the normalization. The traditional choice is to specify σ_8 , corresponding to $R = 8 h^{-1} \text{Mpc}$. As a final field, the best current estimate is $\sigma_8 \approx 0.8$, so clearly nonlinear corrections matter in interpreting this number. The virtue of this convention is that it is then easy to calculate the spectrum normalization at any early time.

77

$$(202) \quad \sigma^2(R) = \int \Delta^2(k) |W_k|^2 d \ln k; \quad W_k = \frac{1}{3} (kR)^3 (\sin kR - kR \cos kR).$$

We now have a full recipe for specifying the matter power spectrum: historically, this is done in a slightly awkward way. First suppose we wanted to consider smoothing the density field by convolution with some **window**. One simple case is to imagine averaging within a sphere of radius R . For the effect on the power spectrum, we need the Fourier transform of this filter:

This becomes close to the horizon length by the time of last scattering, $1 + z \approx 1100$. The resulting damping effect can be seen in figure 16 at $k \sim 10 k_h$.

$$(201) \quad \lambda_s = 16.3 (1 + z)^{-5/4} (\Omega_b^2 \Omega_m^2 h^6)^{-1/4} \text{Gpc}.$$

the geometric mean of the horizon size and the mean free path. Since $\lambda = 1/(n\sigma_T) = 44.3(1 + z)^{-3} (\Omega_b h^2)^{-1}$ proper Gpc, we obtain a comoving damping length of

$$(200) \quad \lambda_s \approx \sqrt{\lambda D_H},$$

typical distance of a random walk in terms of the diffusion coefficient D is $x \approx \sqrt{Dt}$, which gives a damping length of

(4) Silk damping length. Acoustic oscillations are also damped on small scales, where the 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, converting the plasma with it. The

(3) Acoustic horizon length. The horizon at matter-radiation equality also enters in the properties of the baryon component. Since the sound speed is of order c , the largest scales that can undergo a single acoustic oscillation are of order the horizon. The transfer function for a pure baryon universe shows large modulations, reflecting the number of oscillations that have been completed before the universe becomes matter dominated and the pressure support drops. The lack of such large modulations in real data is one of the most generic reasons for believing in collisionless dark matter. Acoustic oscillations persist even when baryons are subdominant, however, and can be detectable as lower-level modulations in the transfer function. We will say more about this later.

be very large, but within range of observations. The dispersal of HDM fluctuations reduces the CDM growth rate on all scales below $L_{\text{free-stream}}$ – or, relative to small scales, there is an enhancement in large-scale power.

Computing lives by the 'garbage out' rule, so how are the initial conditions in the simulation set? This can be understood by thinking of density fluctuations in **Lagrangian** terms (also known as the **Zeldovich approximation**). The proper coordinate of a given particle can be written as

$$\mathbf{x}(t) = a(t) (\mathbf{q} + \mathbf{f}(\mathbf{q}, t)), \quad (204)$$

where \mathbf{q} is the usual comoving position, and the **displacement field** $\mathbf{f}(\mathbf{q}, t)$ tends to zero at $t = 0$. The comoving peculiar velocity is just the time derivative of this displacement:

$$\mathbf{u} = \frac{\partial \mathbf{f}}{\partial t} \quad (205)$$

(partial time derivative because each particle is labelled by an unchanging value of q – this is what is meant by a Lagrangian coordinate).

By conservation of particles, the density at a given time is just the Jacobian determinant between q and x :

$$d/d = \left| \frac{\partial \mathbf{x}}{\partial \mathbf{q}} \right|. \quad (206)$$

When the displacement is small, this is just

$$d/d = 1 - \nabla \cdot \mathbf{f}(\mathbf{q}, t), \quad (207)$$

so the linear density perturbation δ is just (minus) the divergence of the displacement field. All this can be handled quite simply if we define a **displacement potential**:

$$\mathbf{f} = -\nabla \psi(\mathbf{q}), \quad (208)$$

from which we have $\delta = \nabla^2 \psi$ in the linear regime. The displacement potential ψ is therefore proportional to the gravitational potential, Φ . These equations are easily manipulated in Fourier space: given the amplitudes of the Fourier modes, δ_k , we can obtain the potential

$$\psi_k = -\delta_k/k^2, \quad (209)$$

79

and hence the displacement and velocity

$$\begin{aligned} \mathbf{f}_k &= i\mathbf{k}\psi_k \\ \mathbf{u}_k &= i\mathbf{k}\dot{\psi}_k. \end{aligned} \quad (210)$$

Thus, given the density power spectrum to specify $|\delta_k|$ and the assumption of random phases, we can set up a field of consistent small displacements and consistent velocities. These are applied to a uniform particle 'load', and then integrated forward into the nonlinear regime.

These non-linear effects boost the amplitude of the power spectrum at small physical scales (large k scales) as can be seen in Figure 17. For cosmological observations we need to understand these non-linear effects to high precision. This is one of the issues facing modern day cosmology and non-linear effects can only be calculated through large scale suites of HPC N-body simulations.

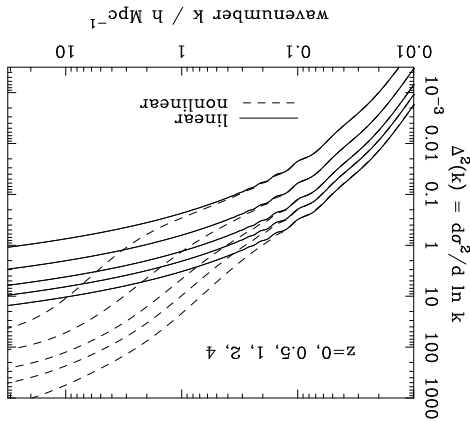
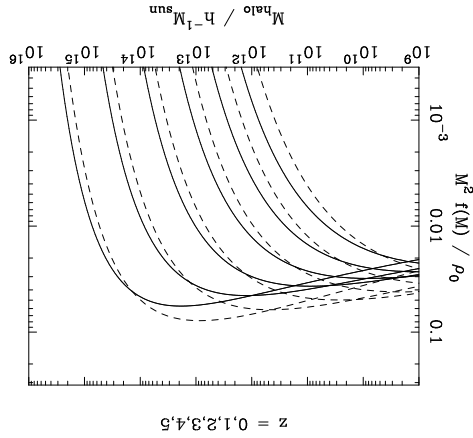


Figure 17. Λ CDM power spectrum normalised by $\sigma_8 = 0.9$. The linear power spectrum is shown solid and the non-linear power spectrum is shown dashed using the fitting formula from Smith et al. (2003).

Figure 18. The mass function in the form of the **multiplicity function**: fraction of mass in the universe found in virialized haloes per unit range in M . The solid lines show a fitting formula to N -body data and the dashed lines contrast the original Press-Schechter formula.



where $(a, b, c) = (1.529, 0.704, 0.412)$. Empirically, one can use $\delta_c = 1.686$ independent of the density parameter (see Section 15.8 in Peacock 1999 for the spherical model argument for the value of δ_c). A plot of the mass function according to this prescription is given in figure 18, assuming what we believe to be the best values for the cosmological parameters. This shows that the Press-Schechter formula captures the main features of the evolution, even though it is inaccurate in detail. We see that the richest clusters of galaxies, with $M \approx 10^{15} h^{-1} M_\odot$, are just coming into existence now, whereas at $z = 5$ even a halo with the mass of the Milky Way, $M \approx 10^{12} h^{-1} M_\odot$ was similarly rare. It can be seen that the abundance of low-mass haloes declines with redshift, reflecting their destruction in the merging processes that build up the large haloes.

83

$$F(>\nu) = (1 + a\nu^b)^{-1} \exp(-c\nu^2), \quad (219)$$

Remarkably, given the dubious assumptions, this expression matches very well to what is found in direct N -body calculations, when these are analysed in order to pick out candidate haloes: connected groups of particles with density about 200 times the mean. The PS form is imperfect in detail, but the idea of a mass function that is universal in terms of ν seems to hold, and a good approximation is

which is the fraction of the mass carried by objects in a unit range of $\ln M$.

$$M^2 f(M) / \rho_0 = \frac{dF}{dM} = \left| \frac{d \ln M}{d \ln \sigma} \right| \sqrt{\frac{\pi}{2}} \nu \exp\left(-\frac{\nu^2}{2}\right). \quad (218)$$

where ρ_0 is the total comoving density. We can write this result in terms of the **multiplicity function**, $M^2 f(M) / \rho_0$,

$$M f(M) / \rho_0 = |dF/dM|, \quad (217)$$

This is the linear-theory view, before the object has collapsed. We define the mass function $f(M)$ where $f(M) dM$ is the comoving number density of objects in the range dM . The probability of a point in space forming as mass between M and $M + dM$ is dF/dM , therefore;

$$M = \frac{4}{\pi} \rho R^3. \quad (216)$$

where $\nu_c = \delta_c / \sigma(M)$ is the threshold in units of the rms density fluctuation and M is the mass contained in a sphere of comoving radius R in a homogeneous universe

$$F(>M) = \sqrt{\frac{\pi}{2}} \int_{\nu_c}^{\infty} \exp(-\nu^2/2) d\nu, \quad (215)$$

written in the universal form
 except it for now. The fraction of the universe condensed into objects with mass $> M$ can then be multiplying the probability by a factor 2. This fudge can be given some justification, but we just The problem with this argument is that half the mass remains unaccounted for: PS therefore simply have $\delta = \delta_c$ when filtered on some larger scale and will be counted as an object of the larger scale. This is really assuming that the only objects that exist at a given epoch are those that scale $> R$. If a point has $\delta > \delta_c$ for a given R , then it will collapse threshold; if a point has $\delta > \delta_c$ for a given R , then it will

Gravitational lensing is the phenomenon whereby a ray of light experiences a curvature of its path, when passing through a gravitational field from nearby mass concentrations. This can be rigorously described from Einstein's theory of relativity, whereby the light propagates along null geodesics, as described by the perturbed space time Robertson-Walker metric. In most astrophysical situations however a more simple approximate description, called gravitational lens theory, is permitted. In this chapter we will derive some of the basics of gravitational lens theory, which we will then build upon in the weak lensing regime.

8.1 The Lens equation

Figure 19 sketches a typical gravitational lensing system where the thin lens is at distance D_d from the observer, perturbs the path of a light ray from a luminous source at distance D_s from the observer, where the distance between the lens plane and source plane is D_{ds} , and all distances are angular diameter distances. In the absence of the lens, the observer would see the source at position β . Instead, the lens deflection by angle α , causes the observer to see the source image at position θ , where all angles, in typical lensing situations, are very small. From figure 19 we see

$$(220) \quad \theta D_s = \beta D_s + \alpha D_{ds}.$$

Defining the reduced deflection angle $\alpha = \alpha D_{ds}/D_s$, the lens equation is given by

$$(221) \quad \beta = \theta - \alpha.$$

THE DEFLECTION ANGLE α AND LENSING POTENTIAL ψ . The deflection angle can be determined by considering the line integral of the gravitational acceleration perpendicular to the light path a_{\perp} .

$$(222) \quad \alpha = \frac{c^2}{2} \int a_{\perp} dl,$$

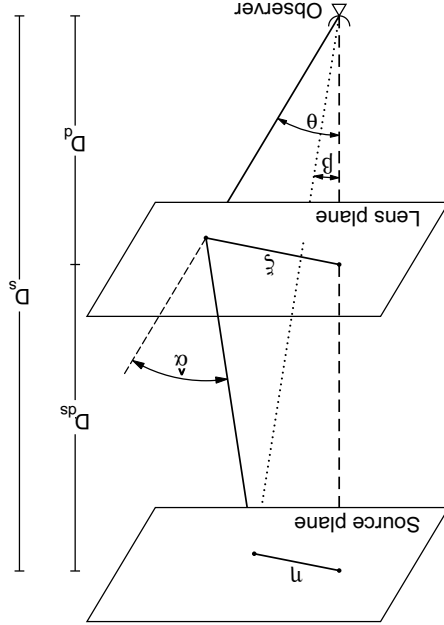


Figure 19. Sketch of a gravitational lensing system, taken from Bartelmann & Schneider 2001.

where acceleration is caused by the gravitational potential of the lens Φ such that $a_{\perp} = \nabla_{\perp}^2 \Phi$, and hence

$$(223) \quad \alpha = \frac{c^2}{2} \int \nabla_{\perp}^2 \Phi dl.$$

A mass distribution which has $\kappa \geq 1$ at some θ , will produce multiple images for some source positions, as we observe in cases of strong gravitational lensing. κ therefore distinguishes between the strong lensing regime ($\kappa \geq 1$) and weak lensing regime ($\kappa \ll 1$).

$$(233) \quad \kappa(\theta) = \frac{\Sigma(\mathbb{D}^a \theta)}{\Sigma^{\text{cr}}(\theta)}.$$

GENERAL CASE For a more general case we now define the dimensionless surface mass density, or convergence κ ,

For $\Sigma = \Sigma^{\text{cr}}$, $\alpha = \theta$, and we see an Einstein ring.

$$(232) \quad \Sigma^{\text{cr}} = \frac{4\pi G}{c^2} \frac{\mathbb{D}^a \mathbb{D}^s}{\mathbb{D}^s}.$$

where Σ^{cr} is the critical surface mass density of the lens defined to be

$$(231) \quad \alpha(\theta) = \frac{\Sigma^{\text{cr}}}{\Sigma}(\theta),$$

For a constant surface mass density, and rewriting in terms of the reduced deflection angle and the lens parameters, $\xi = \mathbb{D}^a \theta$

$$(230) \quad \alpha(\xi) = \frac{|\xi|}{\xi} \int \int d\xi' d\xi'' 2\pi \frac{|\xi|}{4G} \frac{c^2}{\Sigma(\xi')} \Sigma(\xi'').$$

THE AXIALLY SYMMETRIC LENS In the special case of an axially symmetric lens characterized by $\Sigma(\xi) = \Sigma(|\xi|)$, we can choose the origin as the centre of symmetry. The deflection angle is then collinear to ξ and one obtains

$$(229) \quad \alpha(\xi) = \frac{c^2}{4G} \int \int d\xi' d\xi'' \Sigma(\xi') \Sigma(\xi'') \frac{|\xi - \xi'|}{|\xi - \xi''|^2}.$$

we find the two dimensional vector of the deflection angle

87

$$(228) \quad \Sigma(\xi) \equiv \int dl \rho(\xi_1, \xi_2, l),$$

Defining the surface mass density of the lens plane

$$(227) \quad \alpha(\xi) = \frac{c^2}{4G} \int \int d\xi' d\xi'' \rho(\xi_1', \xi_2', l') \frac{|\xi - \xi'|}{|\xi - \xi''|^2}.$$

The total deflection angle is the sum of each small deflection

for impact parameters $\xi \gg R_s \equiv 2GMc^{-2}$. For a mass distribution $\rho(\mathbf{r})$, if the gravitational field is weak, then we can approximate the deflection angle produced by the total mass distribution, as the sum of deflection angles produced by a series of point masses. We can divide our mass distribution in cells of volume dV , with each cell acting as a point mass lens, with mass $dm = \rho(\mathbf{r})dV$. A light ray propagating along the line of sight (l) with position (ξ, l) where ξ is a two dimensional vector in the lens plane, is allowed to pass through the mass distribution. At the mass element dm with position (ξ', l') , the light ray has impact parameter $\xi - \xi'$, if we assume that the deflected light ray corresponds to the Born approximation in atomic and nuclear physics and is valid as long as the deviation of the actual light ray from a straight line within the mass distribution is small compared to the scale on which the mass distribution changes significantly.

$$(226) \quad \alpha = \frac{c^2}{4GM} \xi,$$

A POINT MASS LENS AND THE BORN APPROXIMATION From Einstein's theory of General Relativity, it can be shown that a light ray passing within distance ξ of a point lens of mass M , is deflected by an angle α given by

$$(225) \quad \psi = \frac{\mathbb{D}^a \mathbb{D}^s}{2} \Phi(\xi, l) dl.$$

Nothing that $\xi = \mathbb{D}^a \theta$ the lensing potential can be related to the gravitational potential;

$$(224) \quad \theta - \beta = \Delta \theta \psi(\theta)$$

We can define a lensing potential such that

This wonderfully simple relationship means that if all sources were circular ($e_s = 0$) a measure of the lensed galaxy ellipticity recovers the gravitational shear g and hence the underlying gravitational potential. In practice galaxies have a intrinsic shape, but averaged over many galaxies $\langle e_s \rangle = 0$. We can therefore determine g by measuring the average ellipticity of a large sample of galaxies.

$$e_s = e - g. \tag{245}$$

How is the ellipticity of the galaxy that we observe related to its intrinsic ellipticity before it was lensed? For this we use the jacobian to transform the image quadrupole moments ,

$$Q_s^{ij} = A^i A^j Q_{lm} A^l A^m. \tag{244}$$

and calculate the intrinsic ellipticity of the source e_s in terms of the observed ellipticity e and the reduced shear g . In the weak gravitational lensing limit ($\kappa \ll 1$) Schneider & Seitz (1995) show

For a perfect ellipse we have written the ellipticity in terms of Q_{ij} where $N = Q_{11} + Q_{22} + 2(Q_{11}Q_{22} - Q_{12}^2)^{1/2}$. Figure 21 shows these ellipticity parameters for a series of ellipses.

$$\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} \cos 2\phi \\ \sin 2\phi \end{pmatrix} \frac{1}{1-\beta} = \frac{N}{1} \begin{pmatrix} Q_{11} \\ 2Q_{12} \\ Q_{22} \end{pmatrix} \tag{243}$$

and an ellipticity from its axial ratio β and orientation ϕ ,

$$Q_{ij} = \frac{\int d^2\theta I(\theta) \theta_i \theta_j}{\int d^2\theta I(\theta)}. \tag{242}$$

Consider an isolated galaxy with surface brightness $I(\theta)$. We can define its shape through the quadrupole moment of the light distribution,

8.4 Observing weak lensing

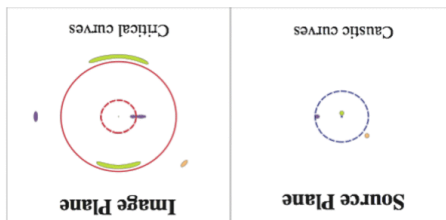
What about lens mass at radii $\theta > \theta_E$? Consider an annulus lens and a point within the annulus. The deflection angle from opposite sides of the annulus are equal and opposite. Hence a lensed source is only effected by lens mass within the impact radius. Strong lensing can therefore only tell us about the mass enclosed in the densest regions of the lens that exhibit strong lensing features. We therefore need resort to weak lensing to measure the total mass of a lens.

This implies that if you observe a tangential arc you can immediately determine the mass of the enclosed lens.

$$\theta_E = \left(4M(<\theta_E) \frac{D_s D_d}{D_s^2 D_d^2} \right)^{1/2} \approx 0.9'' \left(\frac{M(<\theta_E)}{10^{12} M_\odot} \right)^{1/2} \left(\frac{D_s D_d}{D_s^2 D_d^2} \right)^{1/2} \tag{241}$$

TANGENTIAL CURVES We can determine critical curves in the lens plane where the magnification $\mu \rightarrow \infty$ i.e. where $\det(A) \rightarrow 0$. A source galaxy positioned on a tangential critical curve will become distorted tangentially into a giant arc characterised by the condition $\bar{m} = 1$. The radius of this arc is called the Einstein radius θ_E

Figure 20. The images formed by three sources on the source plane due to gravitational lensing. We see that the number of images formed by a source depends on its position relative to the caustic curves (blue). The orange source lies outside the caustic and so produces only one image. The blue source lies inside the radial caustic (dashed blue line), producing three images, two of which are distorted in the radial direction. Due to the radial symmetry of the lens, we see that the tangential critical curve (solid red) maps back to a single degenerate point. The line green source close to the tangential caustic point forms three images, two of which are stretched tangentially and one is a central demagnified image. Taken from www.icosmo.org.



COSMOLOGICAL PARAMETERS Weak lensing gives us an unbiased measurement of the matter distribution and hence the underlying matter power spectrum. We can therefore use it to constrain cosmological parameters and it is particularly sensitive to a combination of the matter density Ω_m and the normalisation of the matter power spectrum σ_8 . This is because lensing is sensitive to both

Up to this point we have focused on extended but discrete lens systems: the strong lensing cluster, the weak lensing galaxy halo. One of the great promises of weak lensing as a tool for cosmology is the ability map out the extended large scale structure of the Universe. Lensing is unique in this respect as it is sensitive to all matter irrespective of its state or nature. Whilst the derivation of the lens theory for large scale structure differs somewhat from our derivations thus far (we can no longer assume that the deflection of light rays is small compared to the scale on which the lens mass distribution changes), the fundamental results remain unchanged and we will use them. For a full derivation see Bartelmann & Schneider (2001).

8.6 Weak lensing by Large Scale Structure

For a typical spiral galaxy halo the lensing shear is very weak $\gamma \sim 0.005$. Compare this to the intrinsic galaxy ellipticity which has a distribution $<e^2> \sim 0.3$. This weakly induced distortion is therefore very difficult to measure and can only be measured statistically by stacking the lensing signal around many thousands of halos.

$$(249) \quad \gamma = \kappa - \kappa = \frac{2\pi\sigma_x^2 D_s^2}{2\pi\sigma_y^2 D_s^2} \theta$$

and the shear

$$(248) \quad \kappa = \frac{\Sigma_{cr}}{\Sigma} = \frac{\theta^2 \pi D_s^2}{2\sigma^2 \pi D_s^2}$$

The convergence is then

$$(247) \quad \Sigma(\xi) = \int_{-\infty}^{\infty} dl \rho(\sqrt{\xi^2 + l^2}) = \frac{2G\xi}{\sigma^2}$$

First we calculate the projected surface mass density, setting our co-ordinate origin at the centre of the SIS halo.

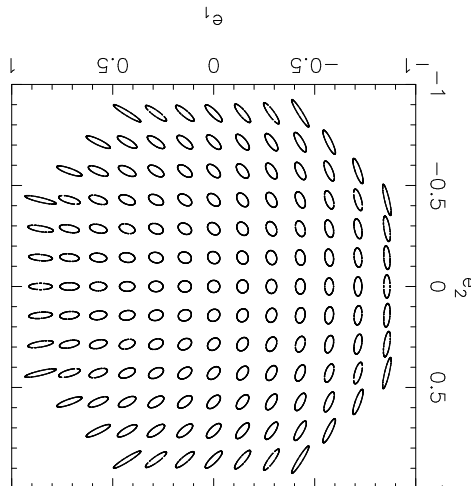
where σ_v is the velocity dispersion of the halo. This profile produces flat rotation curves but is singular (as $r \rightarrow \infty, p \rightarrow \infty$). It is therefore usual to also include a truncation radius where $p(>r_T) = 0$. What lensing effects to we expect to observe around a SIS?

$$(246) \quad p(r) = \frac{\sigma_v^2}{2\pi G r^2}$$

A model that is often used to describe the density of dark matter haloes is the simple isothermal sphere (SIS):

8.5 The Simple Isothermal Sphere Model

Figure 21. Ellipticity parameters for a series of ellipses



SYSTEMATICS A short and final word on systematics. The theory of weak lensing is very elegant and provides a direct route from an observable (galaxy shape) to the underlying matter power spectrum. It is therefore touted as the most promising probe of the ‘‘Dark Universe’’. The observational measurement however is very non-trivial. It requires exquisite knowledge of the telescope, optics and atmospheric conditions and novel, fast computational techniques to extract shape information, at high accuracy, for thousands upon thousands of galaxies. Further, in all our

DARK ENERGY AND MODIFIED GRAVITY We have shown that weak lensing can probe dark matter but what about dark energy? Dark energy acts to oppose the clustering of dark matter over time, suppressing the growth of structure. In addition dark energy changes the distance-redshift relation. Lensing is sensitive to both these effects and in the future will be able to set very tight constraints on the properties of dark energy. The other main probes of dark energy are supernovae or baryon acoustic oscillations. These probes are only sensitive to the distance-redshift relation and cannot distinguish between the cosmological constant or a modification to our theories of gravity. If we find that the constraints from lensing (based on GR) are in disagreement with those from distance-redshift probes this would be evidence for a new beyond-Einstein model of gravity.

A measurement of the correlation between galaxy ellipticities can therefore be directly related to the underlying matter power spectrum! Figure 23 shows the cosmological constraints from this data set compared to constraints from the CMB.

P_{δ} is the 3D matter power spectrum, $f^{\mathcal{K}}(w)$ is the comoving angular diameter distance out to a radial distance w , and $g(w)$ is a weighting function that depends on the redshift distribution of the survey (Bartelmann & Schneider 2001).

$$(252) \quad P_{\delta}(l) = \int_{w_{min}}^0 \frac{9H_0^4 \Omega_m^2}{4c} dw \frac{g^2(w)}{\bar{g}^2(w)} P_{\delta}^{\mathcal{K}}(w),$$

where J_0 is the zeroth order Bessel function and P_{δ} is the power spectrum of the convergence,

$$(251) \quad \xi(\theta) = \frac{2\pi}{\theta} \int dk k P_{\delta}(k) J_0(k\theta),$$

It can be shown (but we won't ask you to show it) that the 2pt shear correlation function is related to the matter power spectrum

Figure 22. The 2pt correlation function measured from the CFHTLS survey. In red, the measured signal. In black (open), an estimate of the systematic errors.

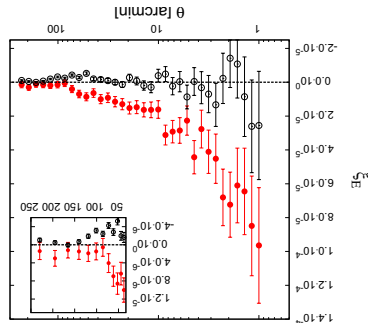


Figure 22 shows the most recent measurement of this statistic from the Canada-France-Hawaii Telescope Legacy Survey. On small angular scales we see galaxy shapes are very correlated (their light has been distorted by the same intervening matter). On large angular scales the correlation weakens as the galaxies are lensed by different structures that are only weakly correlated.

$$(250) \quad E[\xi] = \sum_2^{\alpha=1} \frac{N_{pairs}^{\alpha}}{N_{pairs}}.$$

Stronger clustering results in a higher fraction of regions with mass (Ω_m) and its distribution (σ_8). To constrain cosmological parameters we typically use 2pt statistics. We'll focus on the 2pt shear correlation function ξ which can be estimated from the data using

$$\mathcal{I}_2(\ell) = \ell(\ell+1)C_\ell/2\pi \tag{255}$$

Once again, it is common to define a 'power per octave' measure for the temperature fluctuations: therefore one can associate a 'wavelength' $2\pi/\ell$ with each mode.

The spherical harmonics are familiar as the eigenfunctions of the angular part of ∇^2 , and there are $2\ell+1$ modes of given ℓ , hence the notation for the angular power spectrum, C_ℓ . For $\ell \gg 1$, the spherical harmonics become equivalent to Fourier modes, in which the angular wavenumber is ℓ ;

$$\left\langle \left(\frac{\mathcal{I}}{\delta T} \right)_z \right\rangle = \frac{1}{4\pi} \sum_{\ell,m} |a_{\ell m}|^2 = \frac{1}{4\pi} \sum_{\ell} (2\ell+1) C_\ell \tag{254}$$

where \mathbf{q} is a unit vector that specifies direction on the sky. The spherical harmonics satisfy the orthogonality relation $\int Y_{\ell m} Y_{\ell' m'}^* d^2q = \delta_{\ell\ell'} \delta_{mm'}$, so the variance in temperature averaged over the sky is

$$\frac{\delta T}{\mathcal{I}}(\mathbf{q}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\mathbf{q}), \tag{253}$$

Fluctuations in the 2D temperature perturbation field are treated similarly to density fluctuations, except that the field is expanded in spherical harmonics, so modes of different scales are labelled by multipole number, ℓ ;

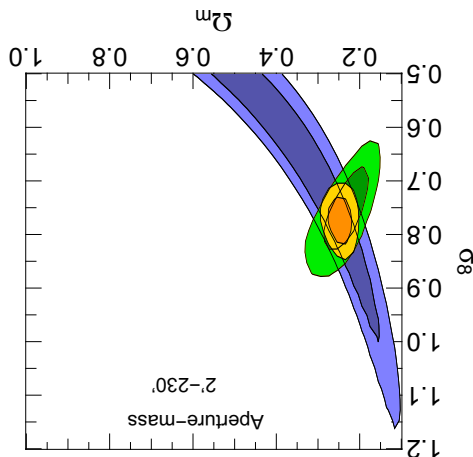
9.1 Anisotropy mechanisms

So far, we have concentrated on describing perturbations in the matter density, and will go on to discuss ways in which these may be observed. But first, we should put in place the corresponding machinery for the fluctuations in the radiation density. These can be observed directly in terms of fluctuations in the temperature of the CMB, which relate to the density fluctuation field at $z \approx 1100$. We therefore have the chance to observe both current cosmic structure and its early seeds. By putting the two together and requiring consistency, the cosmological model can be pinned down with amazing precision.

9 CMB anisotropies – I

derivations we have assumed that galaxies have random orientation - i.e any alignment we see comes exclusively from the lensing distortion. This has been shown not to be the case as neighbouring galaxies have a weak tendency to align. Current research focuses on ensuring these sources of systematics can be minimised and accounted for in preparation for the next generation of lensing telescopes that are being built to "observe" the Dark Universe.

Figure 23. Constraints on the matter density parameter Ω_m and the normalisation of the matter power spectrum σ_8 from weak lensing (CFHTLS survey, blue), and the CMB (WMAP, green), with joint constraints (yellow). Figure taken from Fu et al (2008).



In the usual $\Psi = \Phi$ limit, this is twice as large as one might have expected from Newtonian intuition. This factor 2 thus has an origin that is similar to the factor 2 for relativistic light deflection (where

$$(259) \quad \frac{\delta T}{T} = \frac{1}{c^2} \int \Psi + \Phi \, dt.$$

There are in addition effects generated along the line of sight. One important effect is the integrated Sachs-Wolfe effect (ISW effect), which arises when the potential perturbations evolve:

To the above list should be added 'tensor modes': anisotropies due to a background of primordial gravitational waves, potentially generated during an inflationary era (see below).

$$(258) \quad \frac{\delta T}{T} = \frac{J}{\delta \mathbf{v} \cdot \hat{\mathbf{r}}}.$$

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

$$(257) \quad \frac{\delta T}{T} = \frac{J}{\delta(z_{rs})},$$

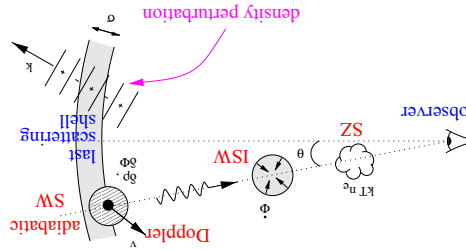
(2) Intrinsic (adiabatic) perturbations. In high-density regions, the coupling of matter and radiation can compress the radiation also, giving a higher temperature:

The factor $1/3$ is a surprise, which arises because Φ has two effects: (i) it redshifts the photons we see, so that an overdensity *cools* the background as the photons climb out, $\delta T/T = \Phi/c^2$; (ii) it causes time dilation at the last-scattering surface, so that we seem to be looking at a younger (and hence *hotter*) universe where there is an overdensity. The time dilation is $\delta t/t = \Phi/c^2$; since the time dependence of the scale factor is $a \propto t^{2/3}$ and $T \propto 1/a$, this produces the counterterm $-\delta T/T = -2/3 \Phi/c^2$.

$$(256) \quad \frac{\delta T}{T} = \frac{1}{3} \Phi/c^2.$$

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

Figure 24. 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 (Rees-Sciama effect) and by electron scattering from hot gas (Sunyaev-Zeldovich effect). The density field at last scattering can be Fourier analysed into modes of wavevector \mathbf{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 (i) because more of them fit inside the scattering shell, and (ii) because their wavevectors point more nearly radially for a given projected wavelength.



We now list the mechanisms that cause **primary anisotropies** in the CMB (as opposed to **secondary anisotropies**, which are generated by scattering along the line of sight). There are three basic primary effects, illustrated in figure 24, which are important on respectively large, intermediate and small angular scales:

(although shouldn't $\ell(\ell + 1)$ be $\ell(\ell + 1/2)$? - see later). Note that $T^2(\ell)$ is a power per $\ln \ell$; the modern trend is often to plot CMB fluctuations with a linear scale for ℓ - in which case one should really use $T^2(\ell)/\ell$.

$$\Delta_{2\text{D}}^{\text{D}}(K) = \left(\frac{2\pi}{L}\right)^2 2\pi K^2 P_{2\text{D}}(K) = \int_0^\infty \Delta_{3\text{D}}^{\text{D}}(k) dk^z / k^3, \quad (262)$$

In terms of dimensionless power, this is

$$P_{2\text{D}}(k_x, k_y) = \sum_{k_z} P_{3\text{D}}(k_x, k_y, k_z) = \int_{-\infty}^\infty \frac{2\pi}{L} P_{3\text{D}}(k) dk^z. \quad (261)$$

The projection is easily performed in the **flat-sky approximation**, where we ignore the curvature of the celestial sphere. The angular wavenumber is then just $\ell = K D_{\text{H}}$, where D_{H} is the distance to the last-scattering surface and K is a 2D transverse physical wavenumber ($K^2 = k_x^2 + k_y^2$). The relation between 3D and 2D power spectra is easily derived: we just add up the power along the unused axis, k_z :

An alternative approximate method is to imagine that the temperature anisotropies exist as a 3D spatial field. The last-scattering surface can be envisaged as a slice through this field, so the angular properties are really just a question of understanding the projection that is involved. This works reasonably well in the **tight coupling** limit where photons and baryons are a single fluid – but this is of course breaking down at last scattering, where the photon mean free path is becoming large.

In practice, rather than trying to solve numerically for the photon distribution function (normally denoted by Θ), we can carry out a multipole transform to work with Θ_ℓ . The integro-differential equation then becomes a set of equations that couple different ℓ values. These have to be solved as a large set of equations (we will see that the CMB power spectrum contains signal at least to $\ell \lesssim 1000$), and when this is done we still have to integrate over k space. It took many years to solve this numerical challenge, and even then the computations were very slow. But a key event in cosmology was the 1996 release of CMBFAST, a public Boltzmann code that allowed computation of the CMB angular power spectrum sufficiently rapidly that a large range of models could be investigated by non-specialists.

Mathematically, we have an integro-differential equation. In fact, it is clear that the force must be proportional to the dipole moment of the distribution function, and this is obviously a problem: it couples the evolution of the number of photons travelling at a given direction with a knowledge of the whole distribution.

101

The natural approach is to start in Fourier space and consider a density fluctuation of given wavevector \mathbf{k} : if we can work out how this appears as an induced temperature fluctuation on the CMB sky, then the problem can be solved by superposition. The wavevector \mathbf{k} sets a natural polar axis, and the temperature anisotropy corresponds to knowing the photon phase-space distribution at our location in space (i.e. the distribution of the photons in energy and as a function of angle with respect to \mathbf{k}). Evolving this function is hard principally because of the coupling between radiation and matter, which is by Thomson scattering. Scattering a beam of photons that come from a given direction will tend to push the electron in the opposite direction, so a net force requires an

We now need to see how the angular power spectrum of the CMB arises from the implementation of these effects. The physical separation we have made is useful for insight, although it is not exactly how things are calculated in practice. We have not been able to spend time going into the detailed formalism used on CMB anisotropies, and the details will have to be omitted here – although the actual equations to be integrated are not enormously complicated. For the present purpose, we will make a few comments about why the exact approach is complicated, and then retreat to a simpler approximate treatment.

9.2 Power spectrum

where the optical depth must exceed $\tau \approx 0.04$, based on the highest-redshift quasars and the BBN baryon density. As we will see later, CMB polarization data have detected a signature consistent with $\tau = 0.1 \pm 0.03$, implying reionization at $z \approx 10$.

$$\frac{\delta T}{\delta T} \rightarrow \frac{\delta T}{\delta T} \exp -\tau, \quad (260)$$

Other foreground effects are to do with the development of nonlinear structure, and are mainly on small scales (principally the Sunyaev–Zeldovich effect from IGM Comptonization). The exception is the effect of reionization; to a good approximation, this merely damps the fluctuations on all scales:

the one-line argument is that the gravitational potential modifies both the time and space parts of the metric, and each contribute equally to the effective change in the coordinate speed of light). But the ISW effect is a little more subtle, and we shall just accept the result as intuitively plausible. As long as $\Omega_m \approx 1$, so the source term for the ISW effect vanishes for much of the universe's history. The ISW effect then becomes only important quite near to the last scattering redshift (because radiation is still important) and at low z (because of Λ).

In other words, perturbations with wavelengths above the horizon size at last scattering generate $\delta T/T$ via gravitational redshift, but on smaller scales it is adiabatic perturbations that matter.

$$(271) \quad k_{\text{crit}}^{-1} \sim (ct)^{-1}.$$

Poisson's equation says $\Delta^2 \Phi = -k^2 \Phi = 4\pi G \rho(\delta\rho/\rho)$, so there is a critical (proper) wavenumber where these two effects are equal: $k_{\text{crit}}^2 \sim G\rho/c^2$. The age of the universe is always $t \sim (G\rho)^{-1/2}$, so this says that

$$(270) \quad \frac{\delta T}{T} \approx \frac{1}{3} \frac{c^2}{\Phi} \quad (\text{gravity}); \quad \frac{\delta T}{T} \approx \frac{1}{3} \frac{\rho}{\delta\rho} \quad (\text{adiabatic}).$$

The main point to appreciate is that the gravitational effects are the ones that dominate on large angular scales. This is easily seen by contrasting the temperature perturbations from the gravitational and adiabatic perturbations:

CHARACTERISTIC SCALES We now want to look at the smaller-scale features of the CMB. The current data are contrasted with some Λ CDM models in figure 25. The key feature that is picked out is the dominant peak at $\ell \approx 220$, together with harmonics of this scale at higher ℓ . How can these features be understood?

Finally, a word about units. The temperature fluctuation $\Delta T/T$ is dimensionless, but anisotropy experiments generally measure ΔT directly, independent of the mean temperature. It is therefore common practice to quote T^2 in units of $(\mu\text{K})^2$.

$$(269) \quad T^2(\ell) = \frac{2\pi}{\ell(\ell+1)} C_\ell.$$

which is why the broad-band measure of the 'power per log ℓ ' is defined as

$$(268) \quad C_\ell = \frac{\ell(\ell+1)}{6} C_2,$$

Finally, it is also possible with some effort to calculate the full spherical-harmonic spectrum from the 3D spatial spectrum. For a scale-invariant spectrum, the result is

103

so the potential well of the richest clusters with velocity dispersion $\sim 1000 \text{ km s}^{-1}$ is of order 10^{-5} deep. It is therefore no surprise to see this level of fluctuation on the CMB sky.

$$(267) \quad v^2 \sim \frac{r}{GM} \Leftrightarrow \Phi \sim \frac{r}{c^2},$$

to say that we can estimate the depth of potential wells today: deduce from CMB anisotropies. As an aside, a more informal argument in the opposite direction is modified on small scales by the transfer function, but it shows how mass fluctuations today can be (where $f(\Omega_m) \approx \Omega_m^{0.23}$ for a flat universe, is the growth suppression factor). This expression is

$$(266) \quad \Delta^2 = (4/9)\delta_H^2 \left(\frac{c\ell}{H_0} \right)^4 \Omega_m^{-2} f(\Omega_m)_2$$

Converting to the present, $\Delta^2 = a^{-2} \Delta^2(a) f(\Omega_m)_2$, and we get

$$(265) \quad \Delta_\Phi^2 = (4\pi G \rho_m a^2 / k^2)_2 \Delta^2(a) \equiv \delta_H^2$$

This immediately determines the large-scale matter power spectrum in the universe today. We know from Poisson's equation that the relation between potential and density power spectra at scale factor a is

$$(264) \quad \delta_H \approx 3 \times 10^{-5}.$$

This is the signature that formed the first detection of CMB anisotropies – by COBE in 1992; we will see below that this corresponds to

The important application of this is to the Sachs-Wolfe effect, where the 3D dimensionless spectrum of interest is that of the potential, $\Delta_\Phi^2 = \delta_H^2$. This shows that the angular spectrum of the CMB should have a flat portion at low ℓ that measures directly the metric fluctuations.

$$(263) \quad \Delta_\Phi^2(K) = \Delta_{\text{3D}}^2.$$

where $k^2 = K^2 + k_z^2$: The 2D spectrum is thus a smeared version of the 3D one, but the relation is pleasingly simple for a scale-invariant spectrum in which $\Delta_{\text{3D}}^2(k)$ is a constant:

$$(276) \quad n_e = \Omega_b \frac{3H_0^2}{8\pi G} \frac{1}{(1+z)^3},$$

(for a flat model). If we re-express the electron number density in terms of the baryon density parameter as

$$(275) \quad \tau = \int \sigma_T n_e dl_{\text{prop}} = \int \frac{c}{dz} \frac{H_0}{c} (1+z) \sqrt{1 - \Omega_m - \Omega_b} (1+z)^3 dz$$

depend on the Thomson-scattering optical depth: REIONIZATION As mentioned previously, it is plausible that energy output from young stars and AGN at high redshift can reionize the intergalactic medium. Certainly, we know empirically from the lack of Gunn–Peterson neutral hydrogen absorption in quasars that such **reheating** did occur, and at a redshift in excess of 6. The consequences for the microwave background of this reionization

at z_{eq} . This is $16(\Omega_b h^2)^{-1} \text{Mpc}$ and subtends a similar angle to θ_r . Finally, a characteristic scale in many density power spectra is set by the horizon for flat models. This is partly responsible for the fall in power at high ℓ (Silk damping also

$$(274) \quad \theta_r \simeq 4 \text{ arcmin},$$

Mpc. This subtends an angle a width in redshift of about $\delta z = 80$. This imposes a radial smearing over scales $\sigma_r = 7(\Omega_m h^2)^{-1/2}$ The second dominant scale is imposed by the fact that the last-scattering surface is fuzzy – with

CMB well. universe – although it turns out that large degrees of spatial curvature *and* Λ can also match the parameters. These expression lie behind the common statement that the CMB data require a flat $D_H \simeq 6000 \Omega_m^{-0.4} h^{-1} \text{Mpc}$, so the peak is predicted at $\ell \simeq 2\pi / (184/6000) \simeq 200$ almost independent limits at $\ell \simeq 1500$ (Bond & Efstathiou 1984). In contrast, a flat vacuum-dominated universe has for this reason long before any useful data existed near the peak, simply because of strict upper than the observed $\ell \simeq 220$, and these can be ruled out. Indeed, open models were disfavoured Figure 25 shows that heavily open universes thus yield a main CMB peak at scales much smaller

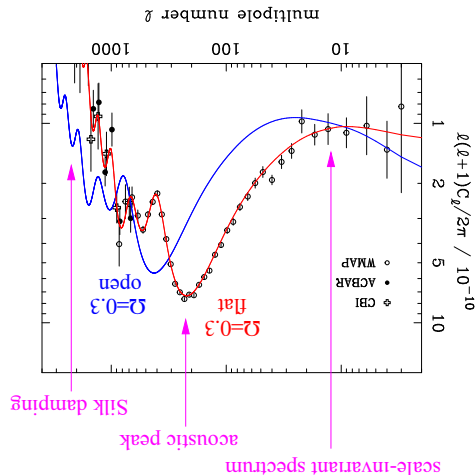
$$(273) \quad D_H = 6000 \Omega_m^{-1} h^{-1} \text{Mpc} \quad \Leftrightarrow \quad \theta_H = D_H^H / D_H = 1.8 \Omega_m^{0.5} \text{ degrees.}$$

The angle this subtends is given by dividing by the current size of the horizon (strictly, the comoving angular-diameter distance to z_{LS}). Again, for a matter-dominated model with $\Lambda = 0$, this is

$$(272) \quad D_H^H = 184 (\Omega_m h^2)^{-1/2} \text{Mpc.}$$

The significance of the main **acoustic peak** is therefore that it picks out the (sound) horizon at last scattering. The redshift of last scattering is almost independent of cosmological parameters at $z_{\text{LS}} \simeq 1100$, as we have seen. If we assume that the universe is matter dominated at last scattering,

right, as discussed in the text. **Figure 25.** Angular power spectra $\mathcal{T}_2(\ell) = \ell(\ell+1)C_\ell/2\pi$ for the CMB, plotted against angular wavenumber ℓ in radians $^{-1}$. For references to the experimental data, see Spergel et al. (2006). The two lines show model predictions for adiabatic scale-invariant CDM fluctuations, calculated using the `CMBFAST` package (Seljak & Zaldarriaga 1996). These have $(n, \Omega_m, \Omega_b, h) = (1, 0.3, 0.05, 0.65)$ and have respectively $\Omega_\nu = 1 - \Omega_m$ (flat), and $\Omega_\nu = 0$ (open). The main effect is that open models shift the peaks to the



It is now clear how LSS data combines with the CMB: $\Omega_m h$ is the main combination probed by the matter power spectrum so this approximate degeneracy is strongly broken using the combined data.

$$(282) \quad \Omega_m h^{3.3} = \text{constant.}$$

Thus for moderate variations from a 'fiducial' model, the CMB peak multipole number scales approximately as $\ell_{\text{peak}} \propto \Omega_{-0.15}^m h^{-0.5}$, i.e. the condition for constant CMB peak location is well approximated as

$$(281) \quad \left. \frac{\partial \ln \Omega_m}{\partial \ln \theta_H} \right|_{\omega_m} = -0.1; \quad \left. \frac{\partial \ln \omega_m}{\partial \ln \theta_H} \right|_{\Omega_m} = \frac{1}{2} \left(1 + \frac{a_{\text{eq}}}{a_{\text{rs}}} \right)^{-1/2} = +0.25,$$

where $z_{\text{rs}} = 1100$ and $a_{\text{eq}} = (23900 \omega_m)^{-1}$. This remarkably simple result captures most of the parameter dependence of CMB peak locations within flat Λ CDM models. Differentiating this equation near a fiducial $\omega_m = 0.13$ gives

$$(280) \quad \theta_H \approx \frac{\Omega_{-0.1}^m}{\sqrt{1 + \frac{a_{\text{rs}}}{a_{\text{eq}}}}} \left[\sqrt{1 + \frac{a_{\text{rs}}}{a_{\text{eq}}}} - \sqrt{\frac{a_{\text{rs}}}{a_{\text{eq}}}} \right],$$

by the light horizon as effect is that a_{eq} depends on Ω_m . Dividing by $D_H(z=0)$ therefore gives the angle subtended today where vacuum energy is neglected at these high redshifts: the expansion factor $a \equiv (1+z)^{-1}$ and $a_{\text{rs}}, a_{\text{eq}} \approx 1100$ independent of the matter and baryon densities, and c_s is fixed by Ω_b . Thus the main

$$(279) \quad D_s(z_{\text{rs}}) \equiv \frac{1}{c_s} \int_{a_{\text{rs}}}^0 \frac{H_0 \Omega_{-0.1}^m}{a + a_{\text{eq}}^{1/2}} da$$

However, this argument is incomplete in detail because the earlier expression for $D_H(z_{\text{rs}})$ assumes that the universe is completely matter dominated at last scattering, and this is not perfectly true. The comoving sound horizon size at last scattering is defined by

107

so there is very little dependence of peak location on cosmological parameters. This contrast between little dependence on density for flat models and a large density dependence for models with cosmological constant is often used to argue that the CMB proves flatness; but this ignores the case where both curvature and Λ are important, and independent constraints on the density are needed before this possibility can be ruled out.

$$(278) \quad D_H = 6000 \Omega_{-0.4}^m h^{-1} \text{Mpc} \Leftrightarrow \theta_H = D_{\text{rs}}^H / D_H^H \Omega_{-0.1}^m,$$

The information we gain from the CMB is dominated by the main acoustic peak at $\ell = 220$, and it is interesting to ask what this tells us. We have argued that the location of this feature marks the angle subtended by the acoustic horizon at last scattering, which has been given as $D_{\text{rs}}^H = 184 (\Omega_m h^2)^{-1/2} \text{Mpc}$. Using the current size of the horizon, the angle subtended in a flat model is

Having given an outline of the physical mechanisms that contribute to the CMB anisotropies, we now examine how the CMB is used in conjunction with other probes to pin down the cosmological model.

10 CMB anisotropies – II

Predictions from CDM galaxy formation models tend to predict a reheating redshift between 10 and 15, thus τ between 0.1 and 0.2 for standard parameters. The main effect of this scattering is to damp the CMB fluctuations by a factor $\exp(-\tau)$, but this does not apply to the largest-scale angular fluctuations. To see this, think backwards: where could a set of photons scattered at z have come from? If they are scattered by an angle of order unity, they can be separated at the last-scattering surface by at most the distance from z to z_{rs} – which is almost exactly the horizon size at z . The critical angle is thus the angle subtended today by the horizon size at the reheating time; for a flat model, this is approximately $z^{-1/2}$ radians, so modes with $\ell > z^{1/2}$ are unaffected. This turns out to be a critical factor in changing the apparent shape of the CMB power spectrum

$$(277) \quad \tau = 0.04h \frac{\Omega_b}{\Omega_{\text{ph}}} \left[\sqrt{1 + \Omega_m z(3 + 3z + z^2)} - 1 \right] \approx 0.04h \frac{\Omega_{-0.1}^m}{z^{3/2}}.$$

where the parameter h is approximately 1.143 for a gas of 25% helium by mass, and do the integral over redshift, we get

These tensor metric distortions are observable via the large-scale CMB anisotropies, where the tensor modes produce a spectrum with the same scale dependence as the Sachs–Wolfe gravitational redshift

$$h_{\text{rms}} \sim H^{\text{inflation}}/m_{\text{P}}. \quad (283)$$

All of our discussion to date applies to models in which scalar modes dominate. But we know that gravity-wave metric perturbations in the form of a traceless symmetric tensor $h_{\mu\nu}$ are also possible, and that inflation predicts that a background of such waves is generated, with amplitude

10.2 Tensor modes

Overall, the kind of precision data now delivered by WMAP allows these effects to be measured, and the degeneracy between Ω_m , Ω_b and h broken without external data. pronounced if the baryon fraction is high (see the second panel of figure 26).

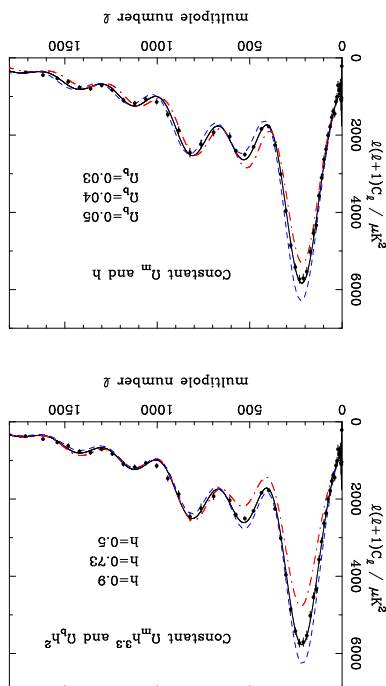
(2) **Baryon loading.** If we keep the overall matter density fixed but alter the baryon fraction, the sound speed at last scattering changes. This has the effect of making a change in the amplitude of the acoustic oscillations beyond the first peak: the drop to the second peak is more

(1) **Early ISW.** We have seen that the transition from radiation domination to matter domination occurs only just before last scattering. Although we have proved that potential fluctuations Φ stay constant during the radiation and matter eras (while vacuum and curvature are negligible), this is not true at the junction, and there is a small change in Φ during the radiation–matter transition (by a factor 9/10; see chapter 7 of Mukhanov’s book). This introduces an additional ISW effect, which boosts the amplitude of the peak, especially for models with low $\Omega_m h^2$, which brings z_{eq} right down to z_{rs} (see the first panel of figure 26).

Although the main horizon-scale peak in the power spectrum dominates the appearance of the CMB, giving degenerate information about cosmological parameters, the fine detail of the pattern is also important. As the quality of the CMB measurements improve, more information can be extracted, and the parameter degeneracies are increasingly broken by the CMB alone. Regarding the structure around the peak, two physical effects are important in giving this extra information:

10.1 Degeneracy breaking with detailed CMB data

Figure 26. The location of the principal peak in the CMB power spectrum is largely determined by the combination $\Omega_m h^{3.3}$, representing the scaling of the angular size of the horizon at last scattering. The two other main characteristics are the rise to the peak above the second and subsequent maxima. The former (the height of the peak above the Sachs–Wolfe plateau) is influenced by the early-time ISW effect: the change in gravitational potential associated with the transition from radiation domination to matter domination. This is illustrated in the first panel, where we fix $\Omega_b h^2$ and hence the sound speed. For fixed peak location, higher h gives lower matter density, and hence a higher peak from the early-time ISW effect (all models are normalized at $\ell = 20$). The second panel shows the influence of varying the baryon density at constant matter density, where we suggest a higher baryon fraction



We have shown that, given perfect data, the CMB anisotropy power spectrum alone is able to determine the main cosmological parameters. But current data are still some way from being ideal – and this capability weakens when we expand the model to include ingredients that are as yet undetected, but which have a reasonable theoretical motivation. However, additional information from large-scale structure cures these problems effectively.

THE GALAXY POWER SPECTRUM A key aim in observational cosmology has long been to use the expected feature at the z_{eq} horizon scale to measure the density of the universe. Data on galaxy clustering is now sufficiently good that this can be done quite accurately. The measured power spectrum from the 2dF Galaxy Redshift Survey is contrasted with CDM models (for which $|\delta_k|^2 \propto k^n T^k$) in figure 27. The curvature of the spectrum is clearly measured, leading to the constraint

$$\Omega_m h = 0.168 \pm 0.016. \quad (289)$$

For $h = 0.7 \pm 10\%$, as indicated by absolute external measurements, this gives $\Omega_m = 0.24 \pm 0.03$. The 2dFGRS results also give a detection of the expected baryon features, leading to a measurement of the baryon fraction:

$$\Omega_b/\Omega_m = 0.185 \pm 0.046 \quad (290)$$

(see Cole et al. astro-ph/0501174). Although this is not as accurate a measurement of the baryon fraction as we obtain from the CMB, it is a more direct piece of evidence that collisionless dark matter is needed; with only baryonic matter, the galaxy power spectrum would be expected to display the same order-unity oscillations that we see in the CMB power spectrum.

In order to reach these conclusions, however, it is necessary to make an assumption about the primordial spectrum, which was taken to be scale-invariant with $n = 1$. Values $n > 1$ would correspond to a larger inferred density, and LSS data cannot break this degeneracy with tilt. The best way to constrain n is to combine with data on CMB anisotropies; as we have discussed, these probe larger scales and give a robust measure of n , which indeed turns out to be very close to unity. Similarly, LSS data do not make any statement about the curvature of the universe. Again, this can be measured from the CMB given the use of LSS data to limit possible combinations of matter content and h , and so break the geometrical degeneracy.

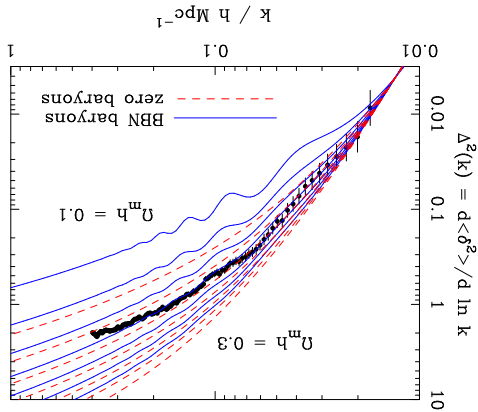


Figure 27. The galaxy power spectrum from the 2dF Galaxy Redshift Survey, shown as the contribution to the fractional density variance per $\ln k$ against wavenumber (spatial wavelength is $\lambda = 2\pi/k$). The data are contrasted with CDM models having scale-invariant primordial fluctuations ($n_s = 1$) and $\Omega_m h = 0.1, 0.15, 0.2, 0.25, 0.3$. The dotted lines show pure CDM models, whereas the solid lines show the effect of baryons at the nucleosynthesis level (assuming $\Omega_b = 0.04$ and $h = 0.7$).

COMBINED CONSTRAINTS FROM CMB+LSS Following the superb 3-year WMAP results (Spergel et al. 2006; astro-ph/0603449), the detailed TT, EB and TB power spectra are measured sufficiently precisely that many of the parameter degeneracies we have worried about are broken, at least weakly. This comes partly from the polarization measurements, and also via the ISW effect. In general, what we have to do is explore a multidimensional parameter space, which can easily be 11-dimensional, as shown in Table 1.

This is frequently reduced to 7 free parameters (ignoring tensors and the neutrino mass fraction, and assuming $w = -1$): a scalar CDM universe. In this case, the interesting parameter to focus on is the curvature. The **likelihood** of the data given the model parameters is regarded as a probability density for the parameters, and we **marginalize** by integrating this distribution

Table 1. Cosmological parameters.

Parameter	Meaning
ω_{dm}	Physical density of dark matter
ω_b	Physical density of baryons
ω_v	Physical density of vacuum
w	Equation of state of vacuum
ω_k	Curvature 'density'
n_s	Scalar spectral index
r	Tensor-to-scalar ratio
n_t	Tensor spectral index
σ_8	Spectrum normalization
τ	Optical depth from reionization
f_ν	Neutrino mass fraction

over the uninteresting parameters. This leaves a probability distribution for the curvature, which is sharply peaked about zero:

$$\Omega_k = -0.01 \pm 0.01. \quad (291)$$

This is normally taken as sufficient empirical justification (in addition to inflationary prejudice) to assume exact flatness when trying to set constraints on more exotic ingredients (tensors: $w \neq -1$). But so far these are not required, and there is a very well specified 6-parameter standard model, as shown in figure 28 and Table 2.

The impressive thing here is the specification of a relatively low optical depth due to reionization, leading to evidence in favour of $n_s > 1$; exact scale-invariance would need a larger optical depth, and thus stronger large-scale polarization than observed. The detection of tilt (a roughly 3.5 σ rejection of the $n_s = 1$ model) has to be considered an impressive success for inflation, given that such deviations from scale invariance were a clear prediction. So should we consider inflation to be proved? Perhaps not yet, but one is certainly encouraged to look more closely at the tensor signal.

LIMITS ON THE TENSOR FRACTION The possibility of a large tensor component yields additional degeneracies, as shown in figure 29. An $n_s = 1$ model with a large tensor component

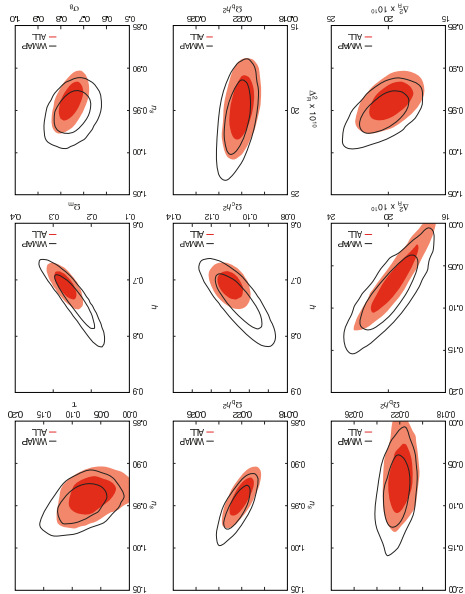


Figure 28. The basic WMAP3 confidence contours on the key cosmological parameters for flat scalar-only models (from Spergel et al. 2006).

can be made to resemble a zero-tensor model with large blue tilt ($n_s > 1$) and high baryon content. This is only weakly broken with current data, as shown in figure 30. This illustrates that we cannot be sure about the 'detection' of tilt: the data can be well matched with $n_s = 1$, but then a substantial tensor fraction is needed.

So far, the tensor contribution to the large-angle anisotropy power spectrum is limited to a fraction $r \lesssim 0.3$ from WMAP. To do much better, we need to detect the characteristic 'B-mode'

Table 2. Constraints on the basic 6-parameter model (flat; no tensors) from WMAP in combination with 2dFGRS in each case.

Parameter	WMAP + 2dFGRS
σ_8	$0.737^{+0.036}_{-0.036}$
τ	$0.083^{+0.028}_{-0.028}$
n_s	$0.948^{+0.015}_{-0.015}$
ω_b	$0.0222^{+0.0007}_{-0.0007}$
ω_m	$0.126^{+0.005}_{-0.005}$
h	$0.733^{+0.020}_{-0.021}$
$\Rightarrow \Omega_m$	$0.236^{+0.020}_{-0.020}$

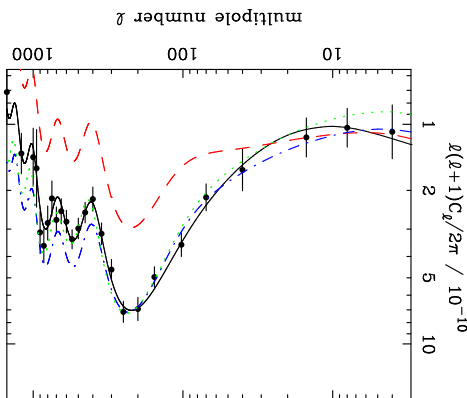
The B modes are excited only by tensors, so all future large-scale polarization experiments will be searching for this signature; it will not be easy, even if the foregrounds are gentle. Planck will only be able to detect tensors if $r \gtrsim 0.1$, although the ultimate limit from cosmic variance is more like $r \approx 10^{-5}$. This sounds like there is a lot of future scope, but it should be recalled that the energy scale of inflation scales as the tensor $C_\ell^{t/4}$. Therefore, we will need a degree of luck with the energy scale if there is to be a detection.

12 The puzzle of dark energy

12.1 Cosmological effects of the vacuum

One of the most radical conclusions of recent cosmological research has been the necessity for a non-zero vacuum density. This was detected on the assumption that Einstein's **cosmological constant**, Λ , might contribute to the energy budget of the universe. But if this ingredient is a reality, it raises many questions about the physical origin of the vacuum energy; as we will see, a variety of models may lead to something similar in effect to Λ , and the general term **dark energy** is used to describe these.

Figure 29. The tensor degeneracy. Adding a large tensor component to an $n_s = 1$ scalar model (solid line) greatly lowers the peak (dashed line), once COBE normalization is imposed. Tiltting to $n_s = 1.3$ cures this (dot-dashed line), but the 2nd and subsequent harmonics are too high. Raising the baryon density by a factor 1.5 (dotted line) leaves us approximately back where we started.



$$t_{\text{rip}} - t_0 \approx -\frac{3}{2} H_0^{-1} |1 + w|^{-1} (\Omega^m)^{-1/2}. \quad (298)$$

If w is negative at all, this leads to models that become progressively more vacuum-dominated as time goes by. When this process is complete, the scale factor should vary as a power of time. The case $w < -1$ is particularly interesting, sometimes known as **phantom dark energy**. Here the vacuum energy density will eventually diverge, which has two consequences: this singularity happens in a finite time, rather than asymptotically; as it does so, vacuum repulsion will overcome the normal electromagnetic binding force of matter, so that all objects will be torn apart in the **big rip**. Integrating the Friedmann equation forward, ignoring the current matter density, the time to this event is

allow dw/dz to be non-zero. Frequently it is sufficient to stick with constant w ; most experiments are sensitive to w at a particular redshift of order unity, and w at this redshift can be estimated with little dependence on whether we

$$w(a) = w_0 + w_a(1 - a). \quad (297)$$

Some complete dynamical model is needed to calculate $w(a)$. Given the lack of a unique model, a common empirical parameterization is

$$H^2(a) = H_0^2 \left[\Omega^v e^{\int -3(w(a)+1) da} + \Omega^m a^{-3} + \Omega^r a^{-4} + \Omega(1 - a)^{-2} \right]. \quad (296)$$

In general, we therefore need

$$\frac{8\pi G \rho}{3H_0^2} = \Omega^v \exp \left(\int -3(w(a)+1) da \right). \quad (295)$$

More generally, we can allow w to vary; in this case, we should regard $-3(w+1)$ as $d \ln \rho / d \ln a$, so that

$$\frac{8\pi G \rho}{3H_0^2} = \Omega^v a^{-3(w+1)}. \quad (294)$$

If this is constant, adiabatic expansion of the vacuum gives

119

$$w \equiv d \ln \rho / d \ln a \quad (293)$$

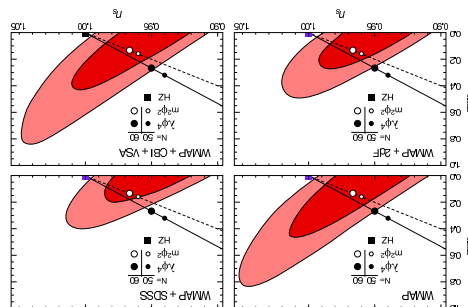
To insert vacuum energy into the Friedmann equation, we need the equation of state

radical change. Nevertheless, as a practical framework, it makes sense to stick with the Friedmann equation and models do exist, particularly in the context of extra dimensions, and these must be borne in mind. At the outset, then we should be very clear that the deduced existence of dark energy depends on the correctness of the Friedmann equation, and this is not guaranteed. Possibly we have the wrong theory of gravity, and we have to replace the Friedmann equation by something else. Alternative

$$\dot{H}^2 - \frac{3}{8\pi G} \rho R^2 = -k^2. \quad (292)$$

The properties of dark energy can be probed by the same means that we used to deduce its existence in the first place: via its effect on the expansion history of the universe. The vacuum density is included in the Friedmann equation, independent of the equation of state

Figure 30. The marginalized WMAP3 confidence contours on the inflationary $r - n$ plane (revised version of plot from Spergel et al. 2006).



OBSERVABLE EFFECTS OF THE VACUUM The comoving distance-redshift relation is one of the chief diagnostics of w . The general definition is

$$D \equiv R_0 r = \int_z^0 \frac{H(z)}{c} dz. \quad (299)$$

Perturbing this about a fiducial $\Omega_m = 0.25$ $w = -1$ model shows a **sensitivity multiplier** of about 5 – i.e. a measurement of w to 10% requires D to 2%. Also, there is a near-perfect degeneracy with Ω_m , so this parameter must be known very well before the effect of varying w becomes detectable.

The other main diagnostic of w is its effect on the growth of density perturbations. These are also sensitive to the vacuum, as may be seen from the growth equation:

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} = 4\pi G\rho_0\delta. \quad (300)$$

The vacuum energy manifests itself in the factor of H in the ‘Hubble drag’ term $2(a/a)\dot{\delta}$. For flat models with $w = -1$, we have seen that the growing mode for density perturbations is approximately $g(a) \propto a\mathcal{L}(a)^{0.23}$. If w is made more negative, this makes the growth law closer to the Einstein–de Sitter $g(a) \propto a$ (for very large negative w , the vacuum was unimportant until very recently). Therefore, increasing w (making it less negative) has an effect in the same sense as *decreasing* Ω_m . Again, the sensitivity to w is rather poor: $|d \ln g/dw| \approx 0.2$.

In the CMB, the main observable is the angle subtended by the horizon at last scattering

$$\theta_{\text{H}} = D(z_{\text{LS}})/D(z = 0). \quad (301)$$

This has the approximate scaling with cosmological parameters (for a flat universe)

$$\theta_{\text{H}} \propto (\Omega_m h_{3.3})^{0.15} \Omega_{\text{r}}^{0.4}; \quad \alpha(w) = -2w/(1 - 3.8w). \quad (302)$$

The latter term comes from a convenient approximation for the current horizon size:

$$D_0 = 2 \frac{H_0}{c} \Omega_{\text{r}}^{-\alpha(w)}. \quad (303)$$

121

At first sight, this looks bad: the single observable of the horizon angle depends on three parameters (four, if we permit curvature). Thus, even in a flat model, we can only pin down w if we know both Ω_m and h .

However, if we have more detail on the CMB than just the main peak location, then we have seen that the $\Omega_m - h$ degeneracy is weakly broken, and that this situation improves with information from large-scale structure, which yields an estimate of $\Omega_m h$. In effect, we have two constraints on the $\Omega_m - h$ plane that are consistent if $w = -1$, but this is not the case for other values of w . In this way, the current combined constraints from CMB plus alternative probes (LSS and the Supernova Hubble diagram) yield an impressive accuracy:

$$w = -0.926^{+0.054}_{-0.053}. \quad (304)$$

for a spatially flat model – see Spergel et al. (2006). The confidence contours are plotted in detail in figure 31, and it is clear that so far there is very good consistency with a simple cosmological constant. But as we will see, plenty of models exist in which some deviation is predicted. The next goal of the global cosmology community is therefore to push the errors on w down substantially – to about 1%. There is no guarantee that this will yield any signal, but certainly it will cut down the range of viable models for dark energy.

One of the future tools for improving the accuracy in w will be large-scale structure. We have seen how this helps pin down the parameter degeneracies inherent in a CMB-only analysis, but it also contains unique information from the acoustic horizon. Earlier, we approximated this without considering how the speed of sound would depend on the baryon density; a good approximation to the exact result is

$$D_a \approx 60 (\Omega_m h^2)^{-0.25} (\Omega_b h^2)^{-0.08} \text{Mpc}. \quad (305)$$

This forms a standard measuring rod, as seen in the ‘baryon wiggles’ in the galaxy power spectrum. In future galaxy surveys, the measurement of this signature as a function of redshift will be a further useful geometrical probe.

where H is the Ricci scalar. Einstein's field equations arise from requiring a stationary action, and it is now obvious how to generate a more complex theory: replace $R + \Lambda$ by some other scalar; a popular choice is $f(R)$. This substitution has to be done with care, however, since there exist stringent constraints on deviations from Einstein gravity in the Solar System. The value of R is proportional to the matter density, which is about 10^6 times larger in interplanetary space than on cosmological scales. Thus what is required empirically is $f(R) \simeq R$ when R is large, but $f(R) \rightarrow$ constant as $R \rightarrow 0$. One might legitimately ask whether it is plausible that nature should carefully make sure that modifications of gravity are locally undetectable in this way.

$$S \propto \int (R + 2\Lambda) \sqrt{-g} d^4x, \tag{307}$$

This possibility is frequently termed 'violation of general relativity', but one should be clear at the outset that this is a misnomer: general relativity means assuming the existence of a metric and writing physics equations in covariant form, most simply by using relativistic invariants. Einstein's field equations are the simplest set consistent with this requirement, but are easy to generalize. This is most easily seen by using the Lagrangian formalism and writing the Einstein-Hilbert action:

In other words, the expansion history of the universe cannot be satisfied without adding a constant to the rhs. But this could mean that the standard Friedmann equation was wrong all along and that the presence of the constant indicates the need for changes to the theory of gravity.

$$H^2(a) = H_0^2 [\Omega_m a^{-3} + \Omega_r a^{-4} - (\Omega - 1)a^{-2} + \Omega_\nu]. \tag{306}$$

An alternative point of view on dark energy, which is receiving increasing interest in the research literature is to suggest that dark energy may not be a genuine physical entity at all. All our current knowledge about it comes from the Friedmann equation:

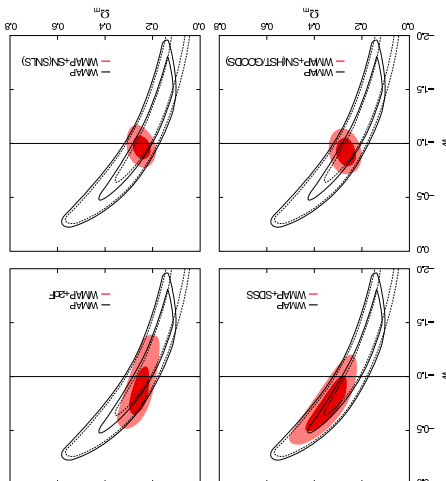
12.4 Modifying gravity

In the interests of time, we will not go into this topic here. Suffice it to say that there is a general problem with all such models: whatever $V(\phi)$ function one chooses, the dynamics of the field are identical if $V \rightarrow V + \text{const}$. One therefore recovers the classical cosmological constant problem, which is to understand why such a constant should be so small. Normally, quintessence models adopt $V(\phi) \rightarrow 0$ as $\phi \rightarrow \infty$ without any good justification.

The simplest physical model for dynamical vacuum energy is a scalar field. We know from inflationary models that this can yield something close in properties to a cosmological constant, and so we can immediately borrow the whole apparatus for modelling vacuum energy at late times. This idea of scalar fields as a dynamical substitute for Λ was first explored by Ratra & Peebles (1988). Of course, this means yet another scalar field that is introduced without much or any motivation from fundamental physics. This hypothetical field is given the fanciful name 'quintessence', implying a new addition to the ancient Greek list of elements (fire, air, earth, water).

12.3 Quintessence

Figure 31. The marginalized WMAP3 confidence contours on the plane of dark-energy equation of state (w) vs Ω_m (from Spergel et al. 2006). A flat universe is assumed, although this is not critical to the conclusions.



$$\rho_v = \frac{c}{h} \left(\frac{E_v}{hc} \right)^4, \quad (312)$$

scale E_v by writing the dimensional scaling of the critical density of the universe ($\sim 10^{-26} \text{kg m}^{-3}$). We can express things in terms of an energy scale. However, even with a cutoff at the hardy microscopic level of $\lambda \sim 1 \text{ nm}$, ρ^{vac} already exceeds what would be the case if space were a lattice, which is perhaps conceivably small which diverges horribly. Is it possible that the upper limit of the integral should be finite? This

$$\rho^{\text{vac}} = \frac{1}{h} \int \omega^3 d\omega, \quad (311)$$

To review why zero is a problematic value for the vacuum density, recall what we mean by the vacuum: $|0\rangle$, or zero occupation number for each wave mode inside a given box. But standard quantum mechanics assigns a zero-point energy of $h\omega/2$ to each mode. Integrating $h\omega/2c^2$ over k -space (with a degeneracy of 2 for polarization) gives a total density of

Whether or not one finds the 'essence' approach to dark energy compelling, there remains one big problem. All the models are constructed using Lagrangians with a particular zero level. All quintessence potentials have the field rolling down towards $V = 0$, and k -essence models lack a potential altogether. They are therefore subject to the classical dilemma of the cosmological constant: adding a pure constant to the Lagrangian has no effect on field dynamics, but mimics a cosmological constant. With so many possible contributions to this vacuum energy from the zero-point energies of different fields (if nothing else), it seems contrived to force $V(\phi)$ to asymptote to zero without a reason.

12.5 The anthropic landscape

A large number of future cosmological surveys are thus gearing up to measure these parameters and search for deviations from $(m, \lambda, n) = (-1, 0.55, 1)$. Whether or not one expects this search to succeed, is it undeniably good for science that cosmology is able to test the correctness of Einstein gravity, rather than simply assuming it.

$$n = \Phi / \Psi. \quad (310)$$

able to probe the sum of the two metric potentials, $\Psi + \Phi$: complete the set, we note that gravitational lensing adds a specific degree of freedom in that it is

125

The standard model is well fitted by $\gamma \approx 0.55$, but many of the modified models discussed above require values of γ that differ from this by of order 0.1. The parameter γ thus forms one natural target for observers, to be added to w as an empirical description of fundamental cosmology. To

$$f_g \equiv \frac{d \ln \delta}{d \ln a} \approx \Omega_m(a)^\gamma. \quad (309)$$

This may all seem pointless if the aim is simply to come up with an alternative model that gives an expansion history $a(t)$ that is just like the standard case with matter plus dark energy. But more recent work has emphasised that it is possible to tell the difference by looking at the growth of structure. Informally speaking, we are exploring the possibility that gravity may have a different strength on the 10-Gpc scale of the entire visible universe than it does on small scales. Here, 'small scales' can mean as large as the kpc scales of galaxies, since the central parts of these can be explained dynamically using standard gravity and no dark matter. We stress that the aim here is to dispose of dark energy, not dark matter: that is the subject of a more radical programme known as **MOND**, or Modified Newtonian Dynamics. There is then the possibility that the behaviour on the intermediate 10-Mpc scales of large-scale structure may be a diagnostic of modified gravity. An empirical parameterization has been developed to deal with this:

This quadratic dependence on density is startling and inconsistent with nucleosynthesis, whatever the value of the **dark radiation** term parameterized by C . More realistic brane models allow a bulk cosmological constant, so that the metric is **warped** and no longer of the form $dt^2 = g^{\mu\nu} dx^\mu dx^\nu - dw^2$. These generalized brane models are known as **Randall-Sundrum models**.

$$H^2(a) \propto \rho^2 + C/a^4. \quad (308)$$

A further popular way in which gravity might have non-standard properties is if the universe has more than the normal 3+1 spacetime dimensions. This was first introduced in the **Kaluza-Klein** picture, in which our universe is a lower-dimensional hypersurface in a higher-dimensional system. Such models were first discussed in the 1920s, and the device chosen to hide the extra dimensions was that they were **compactified** and have the topology of a very small cylinder in the hidden direction(s). A more recent development has been the **brane world** model, in which the extra dimension is not assumed to be small. There is then a larger space, termed the **bulk**, which lies away from the (m)brane on which our universe is located. If Einstein gravity applies to the joint space of bulk and brane, and matter is confined only to the brane, then it has been shown that the apparent Friedmann equation on the Brane is of the form

$$P(A | \text{Observer}) \propto P^{\text{prior}}(A)P(\text{Observer} | A), \quad (313)$$

We then take a Bayesian viewpoint to the distribution of Λ given the existence of observers:

It is far from clear that these minimal assumptions are correct. For example, in the string theory **landscape**, there is no unique form for low-energy particle physics, but instead a large number of possibilities in which numbers such as the fine-structure constant, neutrino masses etc. are different. From the point of view of understanding Λ , we need there to be at least 10^{100} possible states so that at least some have a smaller than the natural m_p^4 density by a sufficient factor. The landscape hypothesis provides this variation in Λ , but does not support the idea that particle physics is otherwise invariant. Still, it makes sense to start with the simplest forms of anthropic variation: if this can be ruled out, it might be taken as evidence in favour of the fuller landscape picture.

$$(3) \quad \text{The horizon-scale inhomogeneity } \delta_h \approx 10^{-5}.$$

$$(2) \quad \text{The entropy per particle } S = (T/2.73)^3/\Omega_m h^2$$

$$(1) \quad \text{The Baryon fraction } f_b = p_b/p_m.$$

THE DISTRIBUTION OF Λ We are now almost in a position to calculate a probability distribution for Λ . First, we have to set some ground rules: what will vary and what will be held fixed? We should try to change as little as possible, so we assume that all universes have the same values for

With this as a starting point, the question now becomes whether we can arrange for the different members of this ensemble to have different values of Λ . This is easily achieved. Let there be some quiddance field with a very flat potential, so that it is capable of simulating Λ effectively. Quantum fluctuations during inflation can also displace this field, so that each member of the **multiverse** would have a different Λ .

creating a set of universes each of which can potentially host observers. eventually random-walk close enough to the origin that the classical end of inflation can occur, thus automatically breaks up into different bubbles with their own histories. Some random subset of these means that some regions of space can inflate for an indefinite time, and a single inflating universe effect of fluctuations becomes more marked and can overwhelm the classical downhill rolling. This **stochastic eternal inflation** due to Linde. Sufficiently far from the origin, the random walk to push ϕ further out – creating further inflation in a self-sustaining process. This is the concept of

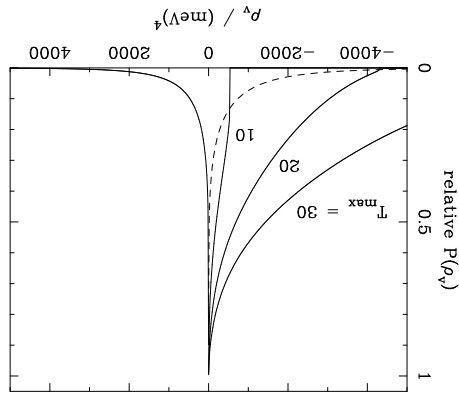
MANY UNIVERSES At first sight, this argument seems quite appealing, but it rapidly leads us into deep waters. How can we talk about changing Λ ? It has the value that it has. We are implicitly invoking an **ensemble picture** in which there are many universes with differing properties. This is a big step (although exciting, if this turns out to be the only way to explain the vacuum level we see). In fact, the idea of an ensemble emerges inevitably from the framework of inflationary cosmology, since the fluctuations in the scalar field can affect the progress of inflation itself. We have used this idea to look at the changes in when inflation ends – but fluctuations can affect the field at all stages of its evolution. They can be thought of as adding a random-walk element to the classical rolling of the scalar field down the trough defined by $V(\phi)$. In cases where ϕ is too close to the origin for inflation to persist for sufficiently long, it is possible for the quantum fluctuations

This leads us in the direction of anthropic arguments, which are able to limit Λ to some extent: if the universe had become vacuum-dominated at $z > 1000$, gravitational instability would have been impossible, Weinberg made the astonishingly precise prediction on this basis that a non-zero vacuum density would be detected at Ω_p of order unity, since there was no reason for it to be much smaller. Indeed, Weinberg made the prediction on this basis that a non-zero vacuum

observations – so we can have almost any effective value of Λ we like. the natural vacuum density of E_{pl}^4 into what is almost a continuous range from the point of view of constant can vary. It has been estimated that there are about 10^{500} distinct minima, which divides by these fields sitting at the minimum of their various potentials, then the effective cosmological additional scalar fields, each with an associated potential. If we assume that a vacuum state is defined For the present purpose, this can be regarded as requiring the introduction of a large number of been sharpened by recent developments in string theory, known under the heading of the **landscape**. supplemented by classical contributions from $V(\phi)$ of any number of scalar fields. This problem has ignorance of the field content of the universe, and because these zero-point contributions can be It should however be clear that this prediction is hard to make fixed, partly because of our

than seems reasonable – a lot fewer than 120 powers of 10 , but still enough to cause a problem. and fermionic oscillators, and the scale of supersymmetry breaking could be as low as 10 TeV. So unbroken supersymmetry, there would be an exact cancellation of the zero point energy of bosonic one should focus on E_v rather than E_{pl}^4 . Also, the solution may lurk at much smaller energies. In vacuum density 120 power of 10 larger than observed. But this is over-dramatising the problem: gravity effects must destroy the flat-space assumptions of quantum field theory. This would give a key cosmological parameters, then $E_v = 2.39 \text{ meV}$ is known to a tolerance of about 1% . What or simply $\rho_v = E_{\text{pl}}^4$ in natural units. If we adopt the values $\Omega_p = 0.75$ and $h = 0.73$ for the

Figure 32. The collapse fraction as a function of the vacuum density, which is assumed to give the relative weighting of different models. The dashed line for negative density corresponds to the expanding phase only, whereas the solid lines for positive density include the recollapse phase, up to maximum temperatures of 10 K, 20 K, 30 K.



So is the anthropic explanation the correct one? Many people find the hypothesis too radical: why postulate an infinity of universes in order to explain a detail of one of them? Certainly, if an alternative explanation for the 'why now' problem existed in the form of e.g. a naturally successful quintessence model, one might tend to prefer that. But so far, there is no such alternative. The longer this situation persists, the more we will be forced to accept that the universe we see can only be understood by making proper allowance for our role as observers.

With this proviso, figure 32 shows the posterior distribution of Λ conditional on the existence of observers in the multiverse. Provided we consider recollapse only to a maximum temperature of about 10 K, the observed figure is matched well by the anthropic prediction: with this cutoff, most observers will see a positive Λ , and something of order 10% of observers will see Λ as big as we do, or smaller.

late stages of recollapse: once the universe becomes too hot, star-formation may be affected and in any case there is little time for life to form. where f_c is the collapse fraction into galaxy-scale objects. For large values of Λ , growth ceases at high redshift, and f_c is exponentially suppressed. But things are less clear-cut if $\Lambda < 0$. Here the universe eventually recollapses, and the high density means that the collapse fraction always tends to unity. So why do we not observe $\Lambda > 0$? The answer is that we have to cut off the calculation at late stages of recollapse: once the universe becomes too hot, star-formation may be affected and in any case there is little time for life to form.

$$dP(p_v) \propto f_c dp_v, \tag{314}$$

We therefore have the simple model

The more difficult part is the prior distribution of Λ , and a common argument is to say that it has a uniform distribution – which seems reasonable enough if we are to allow it to have either sign, but know that we will be interested in a very small range near zero. This is the startling proposition of the anthropic model: the vacuum density takes large ranges, and in almost all realizations, the values are comparable in magnitude to the natural scale m_p^4 ; such models are stupendously inimical to life.

The more difficult part is the prior distribution of Λ , and a common argument is to say that it has a uniform distribution – which seems reasonable enough if we are to allow it to have either sign, but know that we will be interested in a very small range near zero. This is the startling proposition of the anthropic model: the vacuum density takes large ranges, and in almost all realizations, the values are comparable in magnitude to the natural scale m_p^4 ; such models are stupendously inimical to life.

where we need both the prior distribution of Λ between different members of the ensemble and how the chance of getting an observer is modified by Λ . The latter factor should be proportional to the number of stars, which is generally taken to be proportional to the fraction of the baryons that are incorporated into nonlinear structures. We can estimate this using the Press-Schechter apparatus to get the collapse fraction into systems of a galaxy-scale mass. The exact definition of this is not very important, since the CDM power spectrum is very flat on small scales: any mass at all close to $10^{12} M_\odot$ gives similar answers.