MPhys Advanced Cosmology 2011–2012

John Peacock

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.

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)

Syllabus

(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:

$$-ds^2 = c^2 d\tau^2 = c^2 dt'^2 - dx'^2 - dy'^2 - dz'^2 = g_{\mu\nu} dx^\mu dx^\nu. \tag{1}$$

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 $\mathrm{diag}(+1, -1, -1, -1)$. 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 analogue of this surface: something that is almost a 4D (hyper)sphere in Euclidean 5D space:

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

where the metric is

$$ds^2 = dx^2 + dy^2 + dz^2 + dw^2 - dv^2. \tag{3}$$

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

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 choose. Suppose we re-express things using the analogues of polar coordinates:

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

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

$$ds^2 = -\mathcal{R}^2 d\alpha^2 + \mathcal{R}^2 \cosh^2 \alpha \left(d\beta^2 + \sin^2(\beta)[d\gamma^2 + \sin^2 \gamma d\delta^2] \right), \tag{5}$$

which by an obvious change of notation becomes

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

Now we have a completely different interpretation of the metric:

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

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 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.

Historically, de Sitter space was extremely important in cosmology, although it was not 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 this was interpreted as measuring the constant radius of curvature of spacetime, $\mathcal{R}$. By this time, Slipher had already established that most galaxies were redshifted. Hubble's 1929 'discovery' of the expanding universe was explicitly motivated by the possibility of finding the 'de Sitter effect' (although we now know that his sample was too shallow to be able to detect it reliably).

In short, it takes more than just the appearance of $R(t)$ in a metric to prove that something is expanding. That this is the correct way to think about things only becomes apparent when we take a local (and thus Newtonian, thanks to the equivalence principle) look at particle dynamics. Then it becomes clear that a static distribution of test particles is impossible in general, so that it makes more sense to use an expanding coordinate system defined by the locations of such a set of particles.

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 t_0:

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

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

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

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

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

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

$$t_{\rm H} \equiv H^{-1} = 9.78\,\mathrm{Gyr} \ \times \ (H/100\,\mathrm{km\,s^{-1}Mpc^{-1}})^{-1}. \tag{11}$$

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 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.

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:

$$c^2 d\tau^2 = c^2 dt^2 - R^2(t)\left[dr^2 + S_k^2(r)\, d\psi^2\right]. \tag{12}$$

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:

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

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.

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

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

so that $a = 1$ at the present.

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

$$r = \int c\, dt / R(t). \tag{15}$$

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

$$\frac{\nu_{\rm emit}}{\nu_{\rm obs}} \equiv 1 + z = \frac{R(t_{\rm obs})}{R(t_{\rm emit})}. \tag{16}$$

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.

1.2 Cosmological dynamics

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:

$$\dot{R}^2 - \frac{8\pi G}{3}\rho R^2 = -kc^2. \tag{17}$$

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.

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

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

To deduce this, we need to know $\dot{\rho}$, which comes from conservation of energy:

$$d[\rho c^2 R^3] = -p d[R^3]. \tag{19}$$

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

$$\nabla^2 \Phi = 4\pi G(\rho + 3p/c^2). \tag{20}$$

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.

But the greatest astonishment in the Friedmann equation is the term on the rhs. This is related to the curvature of spacetime, and $k = 0, \pm 1$ is the same integer that is found in the RW metric. This cannot be completely justified without the Field Equations, but the **flat** $k = 0$ case is readily understood. Write the energy-conservation equation with an arbitrary rhs, but divide through by R^2:

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

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

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

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

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

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

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:

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

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

$$h \equiv \frac{H_0}{100\,\mathrm{km\,s^{-1}Mpc^{-1}}}, \tag{25}$$

in terms of which the current density of the universe is

$$\begin{aligned} \rho_0 &= 1.878 \times 10^{-26}\, \Omega h^2 \ \mathrm{kg\,m^{-3}} \\ &= 2.775 \times 10^{11}\, \Omega h^2 \ M_\odot \,\mathrm{Mpc^{-3}}. \end{aligned} \tag{26}$$

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

$$w \equiv p/\rho c^2, \tag{27}$$

then conservation of energy says

$$d(\rho c^2 V) = -p\, dV \Rightarrow d(\rho c^2 V) = -w \rho c^2\, dV \Rightarrow d \ln \rho / d \ln a = -3(w+1), \tag{28}$$

so

$$\rho \propto a^{-3(w+1)} \tag{29}$$

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

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).

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

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

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

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

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

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

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

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:

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

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

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

$$a = \Big(\sinh(\sqrt{2}\tau)\Big)^{1/2}. \tag{34}$$

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.

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

$$a = \Big(\sinh(3\tau/2\sqrt{2})\Big)^{2/3}. \tag{35}$$

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 $z \gtrsim 10^4$).

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 $\Omega_m - \Omega_v$ plane, the diagonal line $\Omega_m + \Omega_v = 1$ 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.

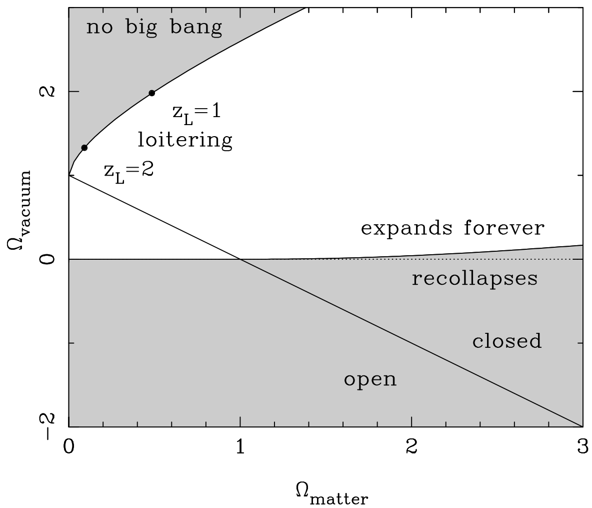

Figure 1. This plot shows the different possibilities for the cosmological expansion as a function of matter density and vacuum energy. Models with total $\Omega > 1$ are always spatially closed (open for $\Omega < 1$), although closed models can still expand to infinity if $\Omega_v \neq 0$. If the 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.

1.3 Observational cosmology

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

$$\frac{dz}{dt} = -\frac{R_0}{R^2}\,\frac{dR}{dt} = -(1+z)H(z), \tag{36}$$

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

$$H^2(a) = H_0^2\left[\Omega_v + \Omega_m a^{-3} + \Omega_r a^{-4} - (\Omega - 1)a^{-2}\right]. \tag{37}$$

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

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

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

$$t \simeq (2/3)H^{-1} \simeq (2/3)H_0^{-1}\Omega_m^{-1/2}(1+z)^{-3/2}. \tag{39}$$

For a flat universe, the current age is $H_0 t_0 \simeq (2/3)\Omega_m^{-0.3}$. 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.

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

$$R_0 r = \int \frac{c}{H(z)}\,dz. \tag{40}$$

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.

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

$$R_0 r_{\mathrm{H}} \simeq \frac{2c}{H_0}\Omega_m^{-0.4}. \tag{41}$$

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

ANGULAR DIAMETERS Recall the RW metric:

$$c^2 d\tau^2 = c^2 dt^2 - R^2(t)\left[dr^2 + S_k^2(r)\, d\psi^2\right]. \tag{42}$$

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:

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

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

LUMINOSITY AND FLUX DENSITY Imagine a source at the centre of a sphere, on which we sit. The photons from the source pass though 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 $[1+z] \times \nu_0$. Overall, the flux density is the luminosity at frequency $[1+z]\nu_0$, divided by the total area, divided by $1+z$:

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

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

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:

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

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.

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

$$\begin{aligned} \textbf{angular}-\textbf{diameter distance}: \quad & D_{\rm A} = (1+z)^{-1} R_0 S_k(r) \\ \textbf{luminosity distance}: \quad & D_{\rm L} = (1+z)\; R_0 S_k(r). \end{aligned} \tag{46}$$

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.

2 The hot big bang

Topics to be covered:

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

2.1 Thermal history

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 staring 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

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

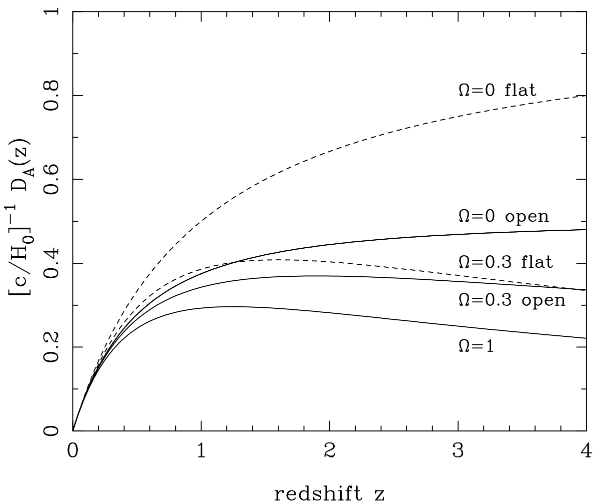

Figure 2. A plot of dimensionless angular-diameter distance versus redshift for various cosmologies. Solid lines show models with zero vacuum energy; dashed lines show flat models with $\Omega_m + \Omega_v = 1$. In both cases, results for $\Omega_m = 1, 0.3, 0$ are shown; higher density results in lower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

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

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

($+$ 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, 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

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

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.

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

$$n = \left(\frac{kT}{c}\right)^3 \frac{4\pi g}{(2\pi\hbar)^3} \int_0^\infty \frac{y^2 dy}{e^y \pm 1}. \tag{50}$$

(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.

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

$$u = \rho \, c^2 = g \, \frac{1}{(2\pi\hbar)^3} \int_0^\infty \frac{4\pi \, p^2 \, dp}{e^{\epsilon(p)/kT} \pm 1} \, \epsilon(p). \tag{51}$$

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

$$u = \frac{\pi^2}{30(\hbar c)^3} \, g \, (kT)^4 \quad \text{(bosons)}. \tag{52}$$

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 density scale as $n \propto T^3$ and $u \propto T^4$, we find $n_{\rm F} = (3/4) \, n_{\rm B}$; $u_{\rm F} = (7/8) \, u_{\rm B}$.

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,

$$dE = TdS - PdV \quad \Rightarrow \quad \left(\frac{E}{V}dV + \frac{\partial E}{\partial T}dT\right) = \left(T\frac{S}{V}dV + T\frac{\partial S}{\partial T}dT\right) - PdV. \tag{53}$$

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

$$S = \frac{E + PV}{T} \tag{54}$$

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

$$s = (4/3)u/T = \frac{2\pi^2 k}{45(\hbar c)^3}\, g\,(kT)^3 \quad \text{(bosons)}, \tag{55}$$

and 7/8 of this for fermions. 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 number of particles. 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 called the **entropy per baryon**.

DEGREES OF FREEDOM Overall, the equilibrium relativistic density is

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

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

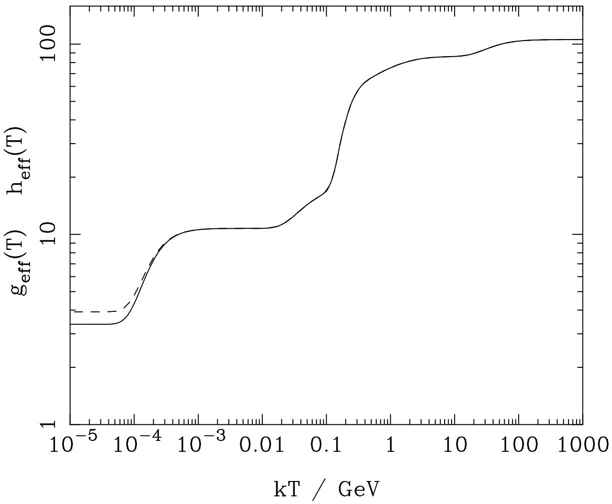

Figure 3. The number of relativistic degrees of freedom as a function of photon temperature. $g_{\rm eff}$ measures the energy density; $h_{\rm eff}$ the entropy (dashed line). The two depart significantly at low temperatures, when the neutrinos are cooler than the photons. For a universe consisting 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.

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–dominated universe, $t = (32\pi G\rho/3)^{-1/2}$, we can deduce the time–temperature relation:

$$t/\text{seconds} = g_{\rm eff}^{-1/2} \left(T/10^{10.26}\,\text{K}\right)^{-2}. \tag{57}$$

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

$$T = 2.725 \pm 0.002\,\text{K}. \tag{58}$$

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

$$T(z) = 2.725(1+z), \tag{59}$$

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 $h_{\rm eff}$ is constant, this requires $T \propto 1/a$. But clearly this does *not* apply near a threshold. At these points, $h_{\rm eff}$ changes rapidly and the universe will expand at nearly constant temperature for a period.

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

$$1 + z_{\rm eq} = 24\,074\,\Omega h^2\,(T/2.725\,{\rm K})^{-4}. \tag{60}$$

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.

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 $kT = 1$ eV for $T = 10^{4.06}$ 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.

Event	T	kT	$g_{\rm eff}$	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

2.2 Freezeout and relics

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.

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**:

$$\dot{n} + 3Hn = -\langle \sigma v \rangle n^2 + S, \tag{61}$$

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:

$$\begin{aligned} \textbf{expansion timescale} &= H(z)^{-1} \\ \textbf{interaction timescale} &= (\langle \sigma v \rangle n)^{-1} \end{aligned} \tag{62}$$

Both these times increase as the universe expands, but the interaction time usually changes fastest. 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 'snapshot' of the properties of the universe at the time the particle was last in thermal equilibrium. This phenomenon of freezeout is essential to the understanding of the present-day nature of the universe. It allows for a whole set of **relics** to exist from different stages of the hot big bang.

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

$$S = \langle \sigma v \rangle n_T^2. \tag{63}$$

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

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

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

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, $N \simeq N_{\scriptscriptstyle T}$. If the population under study is relativistic, $N_{\scriptscriptstyle T}$ does not change with time, because $n_{\scriptscriptstyle T} \propto T^3$ and $T \propto a^{-1}$. This means that it is possible to keep $N = N_{\scriptscriptstyle T}$ exactly, whatever Γ/H. It would however be grossly incorrect to conclude from this that the population stays in thermal equilibrium: if $\Gamma/H \ll 1$, 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 $z \simeq 1100$.

Now consider the opposite case, where the thermal solution would be nonrelativistic, with $N_{\scriptscriptstyle T} \propto T^{-3/2} \exp(-mc^2/kT)$. 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 $(N_{\scriptscriptstyle T}/N)^2$ term on the rhs can still be close to unity. However, if $\Gamma/H \ll 1$, there must be a deviation from equilibrium. When $N_{\scriptscriptstyle T}$ changes sufficiently fast with a, the actual abundance cannot keep up, so that the $(N_{\scriptscriptstyle T}/N)^2$ term on the rhs becomes negligible and $d \ln N / d \ln a \simeq -\Gamma/H$, which is $\ll 1$. There 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,

$$N(a \to \infty) = N_{\scriptscriptstyle T}(\Gamma/H = 1). \tag{65}$$

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

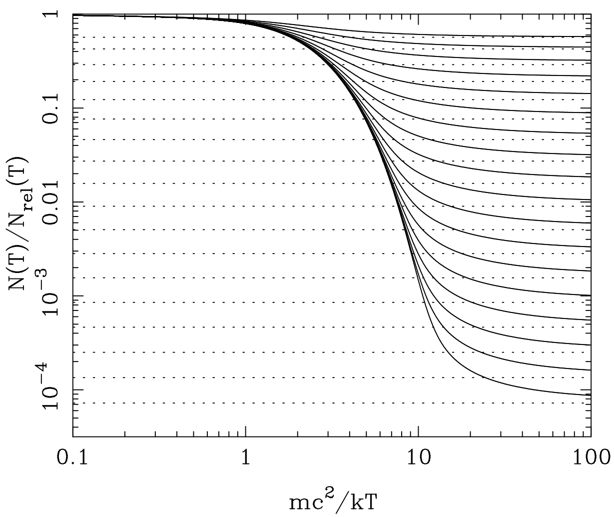

Figure 4. Solution of the Boltzmann equation for freezeout of a single massive fermion. We set $\Gamma/H = \epsilon(kT/mc^2)N/N_{\rm rel}$, as appropriate for a radiation-dominated universe in which $\langle\sigma v\rangle$ is assumed to be independent of temperature. The solid lines show the case $\epsilon = 1$ 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 $\Gamma/H = 1$.

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:

$$\Omega_{\rm relic} h^2 \simeq 0.03\,(\sigma/{\rm pb})^{-1}, \tag{66}$$

where the 'picobarn' is $1\,{\rm pb} = 10^{-40}$ 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 $n_f = H_f/\langle\sigma v\rangle$; (2) $H = (8\pi G\rho/3)^{1/2}$, where $\rho c^2 = (\pi^2/30\hbar^3 c^3) g_{\rm eff}(kT)^4$; (3) $\Omega_{\rm relic} = 8\pi G m n_0/3H_0^2$. The only missing ingredient here is how 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:

$$n_f/n_0 = (h_{\rm eff}^f T_f^3)/(h_{\rm eff}^0 T_0^3). \tag{67}$$

Today, $h_{\rm eff}^0 = 43/11$, and $h_{\rm eff}^f = g_{\rm eff}$ at high redshift. This allows us to deduce the relic density, given the mass, cross-section and temperature of freezeout:

$$\Omega_{\rm relic} h^2 \simeq \frac{10^{-33.0}\ {\rm m}^2}{\langle \sigma v \rangle} \left(\frac{mc^2}{kT_f} \right) g_{\rm eff}^{-1/2}. \tag{68}$$

We see from figure 4 that $mc^2/kT_f \sim 10$ 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.

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 $T_e = 10^{9.7}$ K), the pairs will annihilate. Electrons can interact via either the electromagnetic or the weak interaction, so in principle the annihilations might yield pairs of photons or neutrinos. However, in practice the weak reactions freeze out earlier, at $T \simeq 10^{10}$ K.

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

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

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

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

for $T_\gamma = 2.725$ 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 density in relativistic particles by a factor 1.68 (there are three different kinds of neutrinos, just as there are three **leptons**: the μ and τ particles are heavy analogues of the electron).

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 $n(\nu + \overline{\nu}) = (3/4)n(\gamma;\, T = 1.945\,\mathrm{K})$. 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 $kT < m_\nu c^2$, 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

$$\Omega_\nu h^2 = \frac{\sum m_i}{94.1\,\mathrm{eV}}. \tag{71}$$

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

The current direct laboratory limits to the neutrino masses are

$$\nu_e \lesssim 2.2\,\mathrm{eV} \quad \nu_\mu \lesssim 0.17\,\mathrm{MeV} \quad \nu_\tau \lesssim 15\,\mathrm{MeV}. \tag{72}$$

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

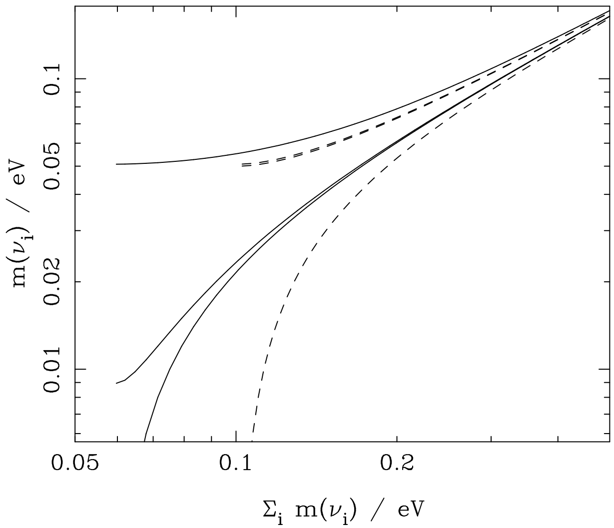

Figure 5. The masses of the individual neutrino mass eigenstates, plotted against the total neutrino mass for a normal hierarchy (solid lines) and an inverted hierarchy (dashed lines). Current cosmological data set an upper limit on the total mass of light neutrinos of around 0.5 eV.

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 $E \simeq m + m^2/2p$). These mixings are known from wonderfully precise experiments detecting neutrinos generated in the sun and the Earth's atmosphere:

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

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 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.

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.

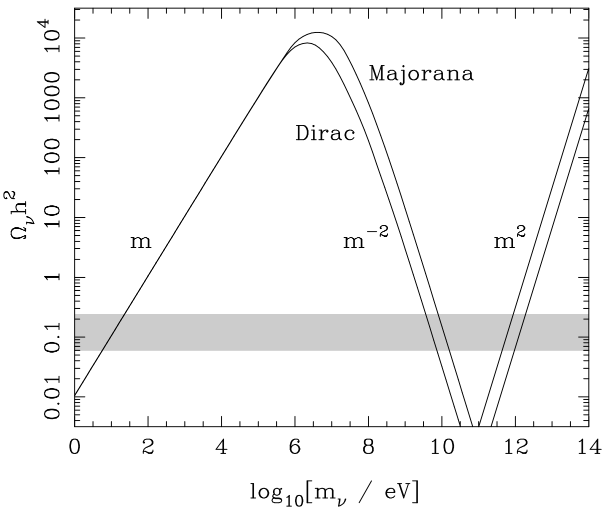

Figure 6. The contribution to the density parameter produced by relic neutrinos (or neutrino-like particles) as a function of their rest mass. The shaded band shows a factor of 2 either side of the observed CDM density. At low masses, the neutrinos are highly relativistic when they decouple: their abundance 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.

(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 $(\text{energy})^2$, so that the relic density falls as $1/m^2$. Interesting masses then lie in the $\simeq 10$ 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 $(\text{energy})^{-2}$. The relic density now rises as m^2, so that the observed dark matter density is attained at $m \simeq 1$ 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 nucleon 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.physto.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 to about 5 times the background. If they run for more years, or increase the detector size, to the point of expecting around 10 background events, these possibilities will be distinguishable; we will will then have either a detection, or will be able to reduce the current upper limits.

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, thay are too fast. Consider a single recombination; if this were to occur directly to the ground state, a photon with

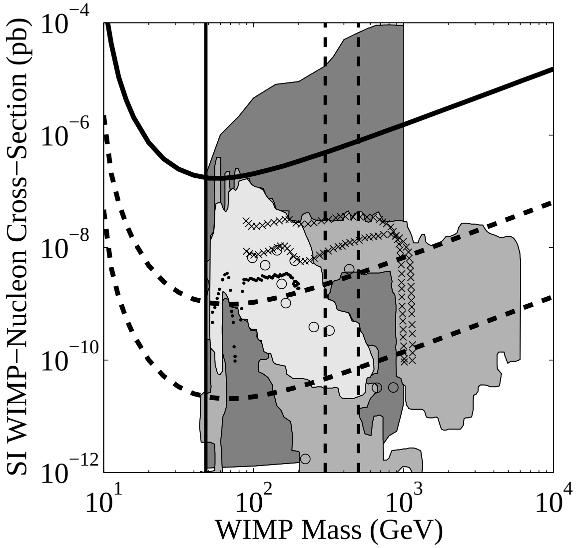

Figure 7. A plot of the dark-matter experimentalists' space: cross-section for scattering off nucleons (in wonderfully baroque units: the 'picobarn' is $10^{-40}\,\mathrm{m}^2$) against WIMP mass. The shaded areas and points indicate various supersymmetric models that match particle-physics constraints and have the correct relic density. The upper curve indicates current 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.

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

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

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

$$\frac{d(nx)}{dt} = -R\,(nx)^2\,\frac{\Lambda_{2\gamma}}{\Lambda_{2\gamma}+\Lambda_{\rm U}(T)}, \tag{74}$$

where n is the number density of protons, x is the fractional ionization, R is the recombination coefficient ($R \simeq 3{\times}10^{-17} T^{-1/2}\,\mathrm{m^3 s^{-1}}$), $\Lambda_{2\gamma}$ is the two-photon decay rate, and $\Lambda_{\rm U}(T)$ 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.

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 rhs: both R and $\Lambda_{\rm U}(T)$ 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:

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

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

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

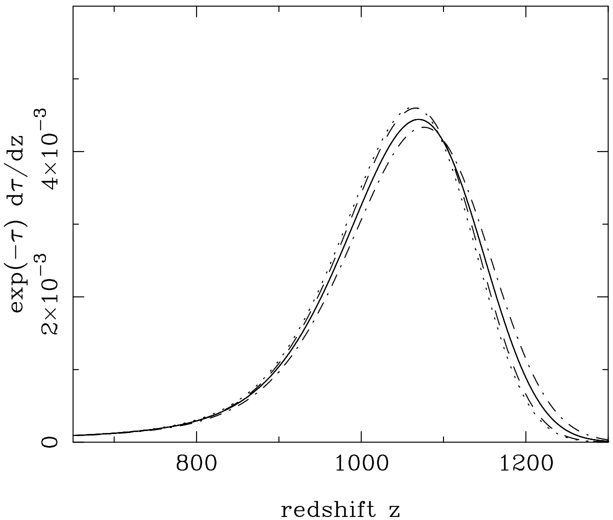

Figure 8. The 'visibility function' governing where photons in the CMB undergo their final scattering. This is very nearly independent of cosmological parameters, as illustrated by the effect of a 50% increase in Ω_b (dotted line), Ω_m (dot-dashed line) and h (dashed line), relative to the standard model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

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,

$$\tau = \int n_e^{\rm tot} x \sigma_{\rm T} d\ell_{\rm proper}; \qquad d\ell_{\rm proper} = R(z)\, dr = R_0\, dr/(1+z). \tag{77}$$

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

$$n_e^{\rm tot}(z) = 9.83\, \Omega_b h^2\, (1+z)^3 \ {\rm m}^{-3}. \tag{78}$$

Also, $d\ell_{\rm proper} = c\,dt$, 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

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

(*cf.* Jones & Wyse 1985).

3 Inflation – I

Topics to be covered:

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

3.1 Initial condition problems

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 in our description of the early stages of cosmological history.

QUANTUM GRAVITY LIMIT In principle, $T \to \infty$ as $R \to 0$, 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 $2\pi\hbar/(mc)$

to $2Gm/c^2$ yields a characteristic mass for quantum gravity known as the **Planck mass**. This mass, and the corresponding length $\hbar/(m_{\rm P}c)$ and time $\ell_{\rm P}/c$ form the system of **Planck units**:

$$\begin{aligned} m_{\rm P} &\equiv \sqrt{\frac{\hbar c}{G}} \simeq 10^{19}{\rm GeV} \\ \ell_{\rm P} &\equiv \sqrt{\frac{\hbar G}{c^3}} \simeq 10^{-35}{\rm m} \\ t_{\rm P} &\equiv \sqrt{\frac{\hbar G}{c^5}} \simeq 10^{-43}{\rm s}. \end{aligned} \tag{80}$$

The Planck time therefore sets the origin of time for the classical phase of the big bang. It is incorrect to extend the classical solution to $R = 0$ and conclude that the universe began in a singularity of infinite density. A common question about the big bang is 'what happened at $t < 0$?', but in fact it is not even possible to get to zero time without adding new physical laws.

NATURAL UNITS To simplify the appearance of equations, it is common practice in high-energy physics to adopt **natural units**, where we take

$$k = \hbar = c = \mu_0 = \epsilon_0 = 1. \tag{81}$$

This convention makes the meaning of equations clearer by reducing the algebraic clutter, and is also useful in the construction of intuitive arguments for the order of magnitude of quantities of interest. Hereafter, natural units will frequently be adopted, although it will occasionally be convenient to re-insert explicit powers of $\hbar$ *etc.*

The adoption of natural units corresponds to fixing the units of charge, mass, length and time relative to each other. This leaves one free unit, usually taken to be energy. Natural units are thus one step short of the Planck system, in which $G = 1$ also, so that all units are fixed and all physical quantities are dimensionless. In natural units, the following dimensional equalities hold:

$$\begin{aligned} [E] = [T] &= [m] \\ [L] &= [m]^{-1} \end{aligned} \tag{82}$$

Hence, the dimensions of energy density are

$$[u] = [m]^4, \tag{83}$$

with units often quoted in GeV^4. It is however often of interest to express things in Planck units: energy as a multiple of m_{P}, energy density as a multiple of m_{P}^4 *etc.* The gravitational constant itself is then

$$G = m_{\mathrm{P}}^{-2}. \tag{84}$$

FLATNESS PROBLEM   Now to quantify the first of the many puzzles concerning initial conditions. From the Friedmann equation, we can write the density parameter as a function of era:

$$\Omega(a) = \frac{8\pi G\rho(a)}{H^2(a)} = \frac{\Omega_v + \Omega_m a^{-3} + \Omega_r a^{-4}}{\Omega_v + \Omega_m a^{-3} + \Omega_r a^{-4} - (\Omega - 1)a^{-2}} \tag{85}$$

(and corresponding expressions for the $\Omega(a)$ corresponding to any one component just by picking the appropriate term on the top line). This tells us that, if the total Ω is unity today, it was always unity (a geometrical statement: if $k = 0$, it can't make a continuous transition to $k = \pm 1$). But if $\Omega \neq 1$, how does $\Omega(a)$ evolve? It should be clear that $\Omega(a) \to 1$ at very large and very small a, provided Ω_v is nonzero in the former case, and provided Ω_m or Ω_r is nonzero in the latter case (without vacuum energy, $\Omega = 1$ is unstable). In short, the $\Omega = 1$ state is an **attractor**, looking in either direction in time. It has long been clear that this presents a puzzle with regard to the initial conditions. These will be radiation dominated, so we have

$$\Omega(a_{\mathrm{init}}) \simeq 1 + \frac{(\Omega - 1)}{\Omega_r}\, a_{\mathrm{init}}^2. \tag{86}$$

If we are willing to consider a Planck-scale origin with $a_{\mathrm{init}} \sim 10^{-32}$, then clearly conditions at that time must be flat to perhaps 60 powers of 10. A more democratic initial condition might be thought to have $\Omega(a_{\mathrm{init}}) - 1$ of order unity, so some mechanism to make it very small (or zero) is clearly required. This 'how could the universe have known?' argument is a general basis for a prejudice that $\Omega = 1$ holds exactly today.

HORIZON PROBLEM   We have already mentioned the puzzle that it has apparently been impossible to establish causal contact throughout the present observable universe. Consider the integral for the horizon length:

$$r_{\mathrm{H}} = \int \frac{c\,dt}{R(t)}. \tag{87}$$

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,

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

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

$$\rho c^2 + 3p < 0. \tag{89}$$

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**:

$$R \propto \exp Ht, \tag{90}$$

where $H = \sqrt{8\pi G \rho_{\rm vac}/3}$. 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.

This is illustrated by in figure 9, where we assume that the universe can be made to change its equation of state abruptly from vacuum dominated to radiation dominated at some time t_c. Before 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 easy to show that $t_c = 1/2H$. In principle, the question 'what happened before the big bang?' is now answered: there was no big bang. There might have still been a singularity at large negative time, but one could imagine the de Sitter phase being of indefinite duration. In a sense, then, an inflationary start to the expansion would in reality be a very slow one – as compared to the common popular description of 'an extraordinarily rapid phase of expansion'.

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

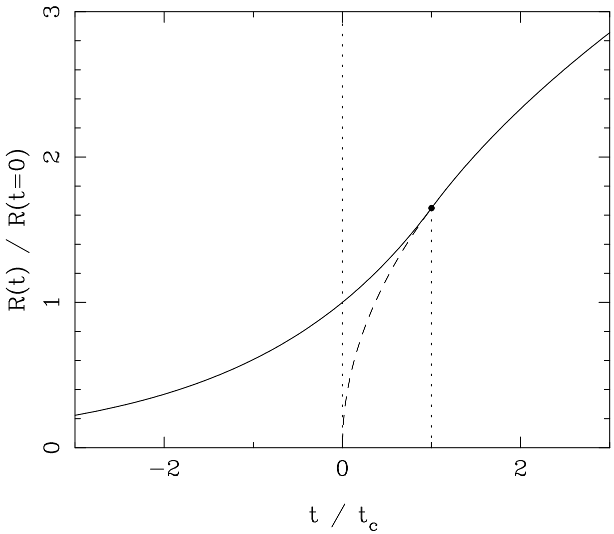

Figure 9. Illustrating the true history of the scale factor in the simplest possible inflationary model. Here, the universe stays in an exponential de Sitter phase for an indefinite time until its equation of state abruptly changes from vacuum dominated to radiation dominated at time t_c. This must occur 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.

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,

$$\dot{R}^2 = \frac{8\pi G\rho R^2}{3} - kc^2. \tag{91}$$

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

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

where $H = \sqrt{8\pi G\rho_{\rm vac}/3}$. 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

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

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

$$\ln f \gtrsim 60 \tag{94}$$

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

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, $t \simeq 10^{-43}$ s, at which point the causal scale was $ct \simeq 10^{-35}$ m. What comoving scale is this? The redshift is roughly (ignoring changes in $g_{\rm eff}$) the Planck energy (10^{19} GeV) divided by the CMB energy ($kT \simeq 10^{-3.6}$ eV), or

$$z_{\rm P} \simeq 10^{31.6}. \tag{95}$$

This expands the Planck length to 0.4 mm today. This is far short of the present horizon ($\sim 6000\,h^{-1}$ 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:

$$\Delta t_{\rm inflation} > 60\, H^{-1}_{\rm inflation}. \tag{96}$$

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 $H^{-1}_{\rm inflation}$. Therefore, the whole inflationary episode need last no longer than about 10^{-33} s.

3.2 Dynamics of scalar fields

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**:

$$\frac{1}{c^2}\ddot{\phi} - \nabla^2\phi + (m^2c^2/\hbar^2)\phi = 0, \tag{97}$$

which is just the standard wave equation if $m = 0$. This is easy to derive just by substituting the de Broglie relations $\mathbf{p} = -i\hbar\boldsymbol{\nabla}$ and $E = i\hbar\partial/\partial t$ into $E^2 = p^2c^2 + m^2c^4$. 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

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

This is just a harmonic oscillator equation, and we can see that the field will oscillate in the potential, with 'kinetic energy' $T = \dot{\phi}^2/2$. This behaviour is rather different to the familiar oscillations of the electromagnetic field: if the field is homogeneous, it does not oscillate. This is because the familiar energy density in electromagnetism ($\epsilon_0 E^2/2 + B^2/2\mu_0$) is entirely kinetic energy in this analogy (to see this, write the fields in terms of the potentials: $\mathbf{B} = \boldsymbol{\nabla}\wedge\mathbf{A}$ and $\mathbf{E} = -\boldsymbol{\nabla}\phi - \dot{\mathbf{A}}$. We don't see coherent oscillations in electromagnetism because the photon has no mass.

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.