eaa.iop.org
DOI: 10.1888/0333750888/1632

# Cosmology: Standard Model

John Peacock

From

## Encyclopedia of Astronomy & Astrophysics

P. Murdin



ISBN: 0333750888

Institute of Physics Publishing
Bristol and Philadelphia

# Cosmology Standard Model

COSMOLOGY in the modern sense of quantitative study of the large-scale properties of the universe is a surprisingly recent phenomenon. The first galaxy RADIAL VELOCITY (a blueshift, as it turned out) was only measured in 1912, by Slipher. It was not until 1924 that Hubble was able to prove that the 'nebulae' were indeed large systems of stars at vast distances, by which time it was clear that almost all galaxies had spectral lines displaced to longer wavelengths. Subsequent observations increasingly verified Hubble's (1929) linear relation between distance $d$ and the recessional velocity inferred if redshift was interpreted as a Doppler shift:

$$v = Hd.$$

The theoretical groundwork for describing the universe via GENERAL RELATIVITY was already in place by the mid-1920s, so that it was not long before the basic observational fact of an expanding universe could be given a relatively standard interpretation. The main observational and theoretical uncertainties in this interpretation concern the matter and energy content of the universe. Different possibilities for this content generate very different COSMOLOGICAL MODELS. The purpose of this article is to outline the key concepts and practical formulae of importance in understanding these models, and to show how to apply them to astronomical observations.

### Isotropic spacetime

Modern observational cosmology has demonstrated that the real universe is highly symmetric in its large-scale properties, but it would in any case make sense to start by considering the simplest possible mass distribution: one whose properties are homogeneous (uniform density) and isotropic (the same in all directions). The next step is to solve the gravitational field equations to find the corresponding metric. Many of the features of the metric can be deduced from symmetry alone—and indeed will apply even if Einstein's equations are replaced by something more complicated. These general arguments were put forward independently by H P Robertson and A G Walker in 1936.

Consider a set of 'fundamental observers', in different locations, all of whom are at rest with respect to the matter in their vicinity. We can envisage them as each sitting on a different galaxy, and so receding from each other with the general expansion (although real galaxies have in addition random velocities of order 100 km $s^{-1}$ and so are not strictly fundamental observers). A global time coordinate $t$ is supplied by the time measured with 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. The coordinate is useful globally rather than locally because the clocks can be synchronized by the exchange of light signals between observers, who agree to set their clocks to a standard time when for example the universal homogeneous density reaches some given value. Using this time coordinate plus isotropy, we already have enough information to conclude that the metric must take the following form:

$$c^2\,\mathrm{d}\tau^2 = c^2\,\mathrm{d}t^2 - R^2(t)[f^2(r)\,\mathrm{d}r^2 + g^2(r)\,\mathrm{d}\psi^2].$$

Here, we have used the equivalence principle to say that the proper time interval $\mathrm{d}\tau$ between two distant events would look locally like special relativity to a fundamental observer on the spot: for them, $c^2\,\mathrm{d}\tau^2 = c^2\,\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2$. Since we use the same time coordinate as they do, our only difficulty is in the spatial part of the metric: relating their $\mathrm{d}x$ etc to spatial coordinates centered on us.

Distances have been decomposed into the product of a time-dependent scale factor $R(t)$ and a time-independent comoving coordinate $r$. It is clear that this metric nicely incorporates the idea of a uniformly expanding model with no center. For small separations, where space is Euclidean, we have a simple scaling of vector separations: $\boldsymbol{x}(t) : R(t)\ \boldsymbol{x}(t_0)$. The same law applies irrespective of the origin we choose: $\boldsymbol{x}_1(t) - \boldsymbol{x}_2(t) : R(t)\ [\boldsymbol{x}_1(t_0) - \boldsymbol{x}_2(t_0)]$, and so every observer deduces $\boldsymbol{v} = H\,\boldsymbol{r}$

Because of spherical symmetry, the spatial part of the metric can be decomposed into a radial and a transverse part (in spherical polars, the angle on the sky between two events is $\mathrm{d}\psi^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2$). The functions $f$ and $g$ are arbitrary; however, we can choose our radial coordinate such that either $f = 1$ or $g = r^2$, to make things look as much like Euclidean space as possible. Furthermore, the remaining function is determined by symmetry arguments.

Consider first the simple case of the metric on the surface of a sphere. A balloon being inflated is a common popular analogy for the expanding universe, and it will serve as a two-dimensional example of a space of constant curvature. If we call the polar angle in spherical polars $r$ instead of the more usual $\theta$, then the element of length, $\mathrm{d}\sigma$, on the surface of a sphere of radius $R$ is

$$\mathrm{d}\sigma^2 = R^2\left(\mathrm{d}r^2 + \sin^2 r\,\mathrm{d}\phi^2\right).$$

It is possible to convert this to the metric for a 2-space of constant negative curvature by the device of considering an imaginary radius of curvature, $R \to_{\mathrm{i}} R$. If we simultaneously let $r \to \mathrm{i}r$, we obtain

$$\mathrm{d}\sigma^2 = R^2\left(\mathrm{d}r^2 + \sinh^2 r\,\mathrm{d}\phi^2\right).$$

These two forms can be combined by defining a new radial coordinate that makes the transverse part of the metric look Euclidean:

$$\mathrm{d}\sigma^2 = R^2\left(\frac{\mathrm{d}r^2}{1 - kr^2} + r^2\,\mathrm{d}\phi^2\right)$$

where $k = +1$ for positive curvature and $k = -1$ for negative curvature.

This is in fact the general form of the spatial part of the Robertson–Walker metric. To prove this in 3D, consider a 3-sphere embedded in four-dimensional Euclidean space, which is defined via the coordinate relation $x^2 + y^2 + z^2 + w^2 = R^2$. Now define the equivalent of spherical polars and write $w = R \cos \alpha$, $z = R \sin \alpha \cos \beta$, $y = R \sin \alpha \sin \beta \cos \gamma$, $x = R \sin \alpha \sin \beta \sin \gamma$, where $\alpha$, $\beta$ and $\gamma$ are three arbitrary angles. Differentiating with respect to the angles gives a four-dimensional vector ($\mathrm{d}x$,$\mathrm{d}y$,$\mathrm{d}z$,$\mathrm{d}w$), and it is a straightforward exercise to show that the squared length of this vector

$$|(\mathrm{d}x, \mathrm{d}y, \mathrm{d}z, \mathrm{d}w)|^2 = R^2[\mathrm{d}\alpha^2 + \sin^2\alpha(\mathrm{d}\beta^2 + \sin^2\beta\,\mathrm{d}\gamma^2)]$$

which is the Robertson–Walker metric for the case of positive spatial curvature.

This $k = +1$ metric describes a closed universe, in which a traveler who sets off along a trajectory of fixed $\beta$ and $\gamma$ will eventually return to their starting point (when $\alpha = 2\pi$). In this respect, the positively curved 3D universe is identical to the case of the surface of a sphere: it is finite, but unbounded. By contrast, the $k = -1$ metric describes an open universe of infinite extent; as before, changing to negative spatial curvature replaces $\sin \alpha$ with $\sinh \alpha$, and $\alpha$ can be made as large as we please without returning to the starting point. The $k = 0$ model describes a flat universe, which is also infinite in extent. This can be thought of as a limit of either of the $k = \pm 1$ cases, where the curvature scale $R$ tends to infinity.

The Robertson–Walker metric may be written in a number of different ways. The most compact forms are those where the comoving coordinates are dimensionless. In terms of the function

$$S_k(r) = \begin{cases} \sin r & (k=1) \\ \sinh r & (k=-1) \\ r & (k=0) \end{cases}$$

the metric can be written as

$$c^2\,\mathrm{d}\tau^2 = c^2\,\mathrm{d}t^2 - R^2(t)[\mathrm{d}r^2 + S_k^2(r)\,\mathrm{d}\psi^2].$$

The most common alternative is to use a different definition of comoving distance, $S_k(r) \rightarrow r$, so that the metric becomes

$$c^2\,\mathrm{d}\tau^2 = c^2\,\mathrm{d}t^2 - R^2(t)\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\,\mathrm{d}\psi^2\right).$$

There should of course be two different symbols for the different comoving radii, but each is often called $r$ in the literature. Finally, a common alternative form of the scale factor is where its present value is set to t)\equiv

$$a(t) \equiv R(t)/R_0.$$

## The redshift

How does this discussion relate to HUBBLE'S LAW: $v = Hr$? Comoving coordinates are time independent, so the proper separation of two fundamental observers is just $R(t)\,\mathrm{d}r$, and differentiation gives Hubble's law, $v = H(R\,\mathrm{d}r)$, with

$$H = \frac{\dot{R}}{R}.$$

At small separations, the recessional velocity gives the Doppler shift (see DOPPLER EFFECT):

$$\frac{\nu_{\mathrm{emit}}}{\nu_{\mathrm{obs}}} \equiv 1 + z \simeq 1 + \frac{v}{c}.$$

This defines the REDSHIFT $z$ in terms of the shift of spectral lines. What is the equivalent of this relation at larger distances? Since photons travel on null geodesics of zero proper time, we see directly from the metric that

$$r = \int \frac{c\,\mathrm{d}t}{R(t)}.$$

The comoving distance is constant, whereas the domain of integration in time extends from $t_{\mathrm{emit}}$ to $t_{\mathrm{obs}}$; these are the times of emission and reception of a photon. Photons that are emitted at later times will be received at later times, but these changes in $t_{\mathrm{emit}}$ and $t_{\mathrm{obs}}$ cannot alter the integral, since $r$ is a comoving quantity. This requires the condition $\mathrm{d}t_{\mathrm{emit}} / \mathrm{d}t_{obs} = R(t_{emit}) / R(t_{obs})$, which means that events on distant galaxies time dilate according to how much the universe has expanded since the photons we see now were emitted. Clearly (think of events separated by one period), this dilation also applies to frequency, and we therefore obtain

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

In terms of the normalized scale factor $a(t)$, $a(t) = (1 + z)^{-1}$. Photon wavelengths therefore stretch with the universe, as is intuitively reasonable.

## The meaning of the redshift

For small redshifts, the interpretation of the redshift as a Doppler shift ($z = v/c$) is quite clear. What is not so clear is what to do when the redshift becomes large. A common but incorrect approach is to use the special-relativistic Doppler formula and write

$$1 + z = \left(\frac{1 + v/c}{1 - v/c}\right)^{1/2}.$$

This is wrong in general, but it is all too common to read of the latest high-redshift QUASAR as 'receding at 95% of the speed of light'. The reason the redshift cannot be interpreted in this way is because a non-zero mass density must cause gravitational redshifts.

However, although the redshift cannot be thought of as a global Doppler shift, it is correct to think of the effect as an accumulation of the infinitesimal Doppler shifts caused by photons passing between fundamental observers separated by a small distance:

$$\frac{\delta z}{1+z} = H(z)\,\delta\ell(z)$$

(where $\delta\ldots$ is a radial increment of proper distance). This expression may be verified by substitution of the standard expressions for $H(z)$ and $\mathrm{d}\ldots/\mathrm{d}z$. The nice thing about this way of looking at the result is that it emphasizes that it is momentum that is redshifted; particle de Broglie wavelengths thus scale with the expansion, a result that is independent of whether their rest mass is non-zero.

An inability to see that the expansion is locally just kinematical also lies at the root of perhaps the worst misconception about the BIG BANG THEORY. Many semipopular accounts of cosmology contain statements to the effect that 'space itself is swelling up' in causing the galaxies to separate. In fact, objects separate now only because they have done so in the past; a pair of massless objects set up at rest with respect to each other in a uniform model will show no tendency to separate (in fact, the gravitational force of the mass lying between them will cause an inward relative acceleration). In the common elementary demonstration of the expansion by means of inflating a balloon, galaxies should be represented by glued-on coins, not ink drawings (which will spuriously expand with the universe).

## Dynamics of the expansion

The equation of motion for the scale factor can be obtained in a quasi-Newtonian fashion. Consider a sphere about some arbitrary point, and let the radius be $R(t$r$r$, where $r$ is arbitrary. The motion of a point at the edge of the sphere will, in Newtonian gravity, be influenced only by the interior mass. We can therefore apparently write down immediately a differential equation (Friedmann's equation) that expresses conservation of energy: $(\dot{R}r)^2/2 – GM/Rr = constant$. In fact, this equation really requires general relativity: the gravitation from mass shells at large distances is not Newtonian, because space is curved, and so we cannot employ the usual argument about their effect being zero. Nevertheless, the result that the gravitational field inside a uniform shell is zero does hold in general relativity, and is known as Birkhoff's theorem. General relativity becomes even more vital in giving the constant of integration in Friedmann's equation:

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

Note that this equation covers all contributions to $\rho$, i.e. those from matter, radiation and vacuum; it is independent of the equation of state.

It is sometimes convenient to work with the time derivative of the Friedmann equation, for the same reason that acceleration arguments in dynamics are sometimes more transparent than energy ones. Differentiating with respect to time requires a knowledge of $\dot{\rho}$, but this can be eliminated by means of conservation of energy: $\mathrm{d}(\rho c^2 R^3) = –p\,\mathrm{d}(R^3)$. We then obtain

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

Both this equation and the Friedmann equation in fact arise as independent equations from different components of Einstein's equations for the Robertson–Walker metric.

The Friedmann equation is so named because FRIEDMANN was the first to appreciate, in 1922, that Einstein's equations admitted cosmological solutions containing matter only (although it was Lemaître who in 1927 both obtained the solution and appreciated that it led to a linear distance–redshift relation). The term Friedmann model is therefore often used to indicate a matter-only cosmology, even though his equation includes contributions from all equations of state. A common shorthand for relativistic cosmological models, which are described by the Robertson–Walker metric and which obey the Friedmann equation, is to speak of FRW models.

## Density parameters etc

According to the Friedmann equation, the 'flat' universe with $k = 0$ arises for a particular critical density. We are therefore led to define a density parameter as the ratio of density to critical density:

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

Since $\rho$ and $H$ change with time, this defines an epoch-dependent density parameter. The current value of the parameter should strictly be denoted by $\Omega_0$. Because this is such a common symbol, it is normal to keep the formulae uncluttered by normally dropping the subscript; the density parameter at other epochs will be denoted by $\Omega(z)$. If we now also define a dimensionless (current) Hubble parameter as

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

then the current density of the universe may be expressed as

$$\begin{aligned}\rho_0 &= 1.88 \times 10^{-26}\Omega h^2\ \mathrm{kg\ m^{-3}}\\ &= 2.78 \times 10^{11}\Omega h^2 M_\odot\ \mathrm{Mpc^{-3}}.\end{aligned}$$

A powerful approximate model for the energy content of the universe is to divide it into pressureless matter ($\rho : R^{-3}$), radiation ($\rho : R^{-4}$) and vacuum energy ($\rho$ independent of time—i.e. there is a non-zero cosmological constant). 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; the third relation applies for Einstein's cosmological constant. In terms of observables, this means that the density is written as (where the normalized scale factor is $a = R/R_0$).

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

In terms of the deceleration parameter,

$$q \equiv -\frac{\ddot{R}R}{\dot{R}^2}$$

the $\ddot{R}$ form of the Friedmann equation says that

$$q = \Omega_m/2 + \Omega_r - \Omega_v.$$

Lastly, it is often necessary to know the present value of the scale factor, which may be read directly from the Friedmann equation:

$$R_0 = \frac{c}{H_0}[(\Omega - 1)/k]^{-1/2}.$$

The HUBBLE CONSTANT thus sets the curvature length, which becomes infinitely large as Ω approaches unity from either direction. Only in the limit of zero density does this length become equal to the other common measure of the size of the universe—the Hubble length, $c/H_0$.

### Solutions to the Friedmann equation

The Friedmann equation may be solved most simply in 'parametric' form, by recasting it in terms of the conformal time $d\eta = c\,dt/R$ (denoting derivatives with respect to $\eta$ by primes):

$$R'^2 = \frac{8\pi G}{3c^2}\rho R^4 - kR^2.$$

Because $H_0^2R_0^2 = kc^2/(\Omega-1)$, the Friedmann equation becomes

$$a'^2 = \frac{k}{\Omega - 1}[\Omega_r + \Omega_m a - (\Omega - 1)a^2 + \Omega_v a^4]$$

which is straightforward to integrate provided that $\Omega_v = 0$. Solving the Friedmann equation for $R(t)$ in this way is important for determining global quantities such as the present age of the universe, and explicit solutions for particular cases are considered below. However, from the point of view of observations, and in particular the distance–redshift relation, it is not necessary to proceed by the direct route of determining $R(t)$.

To the observer, the evolution of the scale factor is most directly characterized by the change with redshift of the Hubble parameter and the density parameter; the evolution of $H(z)$ and $\Omega(z)$ is given immediately by the Friedmann equation in the form $H^2 = 8\pi G\rho/3 - kc^2/R^2$. Inserting the model dependence of $\rho$ on $a$ gives

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

This is a crucial equation, which can be used to obtain the relation between redshift and comoving distance. The radial equation of motion for a photon is $R\,dr = c\,dt = c\,dR/\dot{R} = c\,dR/RH$. With $R = R_0/(1+z)$, this gives

$$\begin{aligned} R_0\,dr &= \frac{c}{H(z)}\,dz \\ &= \frac{c}{H_0}[\Omega_v + \Omega_m a^{-3} + \Omega_r a^{-4} - (\Omega - 1)a^{-2}]^{-1/2}\,dz. \end{aligned}$$

This relation is arguably the single most important equation in cosmology, since it shows how to relate comoving distance to the observables of redshift, Hubble constant and density parameters. The comoving distance determines the apparent brightness of distant objects, and the comoving volume element determines the numbers of objects that are observed. These aspects of observational cosmology are discussed in more detail below.

Lastly, using the expression for $H(z)$ with $\Omega(a)-1 = kc^2/H^2R^2$ gives the redshift dependence of the total density parameter:

$$\Omega(z) - 1 = \frac{\Omega - 1}{1 - \Omega + \Omega_v a^2 + \Omega_m a^{-1} + \Omega_r a^{-2}}.$$

This last equation is very important. It tells us that, at high redshift, all model universes apart from those with only vacuum energy will tend to look like the $\Omega = 1$ model. This is not surprising given the form of the Friedmann equation: provided that $\rho R^2 \to \infty$ as $R \to 0$, the $-kc^2$ curvature term will become negligible at early times. If $\Omega \neq 1$, then in the distant past $\Omega(z)$ must have differed from unity by a tiny amount: the density and rate of expansion needed to have been finely balanced for the universe to expand to the present. This tuning of the initial conditions is called the flatness problem and is one of the motivations for the applications of quantum theory to the early universe.

### Matter-dominated universe

From the observed temperature of the microwave background (2.73 K) and the assumption of three species of neutrino at a slightly lower temperature, we deduce that the total relativistic density parameter is $\Omega_r h^2 \simeq 4.2\text{–}10^{-5}$, so at present it should be a good approximation to ignore radiation. However, the different redshift dependences of matter and radiation densities mean that this assumption fails at early times: $\rho_m/\rho_r : (1+z)^{-1}$. One of the critical epochs in cosmology is therefore the point at which these contributions were equal: the redshift of matter–radiation equality

$$1 + z_{eq} \simeq 23\,900\,\Omega h^2.$$

At redshifts higher than this, the universal dynamics was dominated by the relativistic-particle content. By an interesting coincidence, this epoch is close to another important event in cosmological history: recombination. Once the temperature falls below $\approx 10^4$ K, ionized material can form neutral hydrogen. Observational astronomy is only possible from this point on, since Thomson scattering from electrons in ionized material prevents photon propagation. In practice, this limits the maximum redshift of observational interest to about 1100; unless $\Omega$ is very low or vacuum energy is important, a matter-dominated model is therefore a good approximation to reality.

By conserving matter, we can introduce a characteristic mass $M_*$, and from this a characteristic radius $R_*$:

$$\frac{4\pi G}{3c^2}\rho R^3 = \frac{c}{H_0}\frac{\Omega[k(\Omega-1)]^{3/2}}{2} \equiv \frac{GM_*}{c^2} \equiv R_*$$

where we have used the expression for $R_0$ in the first step. When only matter is present, the conformal-time version of the Friedmann equation is simple to integrate for $R(\eta)$, and integration of $dt = d\eta/R$ gives $t(\eta)$:

$$R = kR_*[1 - C_k(\eta)]$$

$$ct = kR_*[\eta - S_k(\eta)].$$

The evolution of $R(t)$ in this solution is plotted in figure 1. A particular point to note is that the behavior at early times is always the same: potential and kinetic energies greatly exceed total energy and we always have the $k = 0$ form $R:t^{2/3}$.

## Radiation-dominated universe

At high enough redshifts, the $R:t^{2/3}$ law will fail, because radiation pressure will become important. At these redshifts, it is an excellent approximation to ignore the effects of spatial curvature, so that the Friedmann equation for a matter–radiation mix is

$$\dot{a}^2 = H_0^2(\Omega_m a^{-1} + \Omega_r a^{-2}).$$

This may be integrated to give the time as a function of scale factor:

$$H_0 t = \frac{2}{3\Omega_m^2}[(\Omega_r + \Omega_m a)^{1/2}(\Omega_m a - 2\Omega_r) + 2\Omega_r^{3/2}]$$

which goes to $\frac{2}{3}a^{3/2}$ for a matter-only model and to $a^2/2$ for radiation only. At early times, the scale factor thus grows as $R:t^{1/2}$.

One further way of presenting the model's dependence on time is via the density. Following the above, it is easy to show that

$$t = \begin{cases} \left(\dfrac{1}{6\pi G\rho}\right)^{1/2} & \text{(matter domination)} \\ \left(\dfrac{3}{32\pi G\rho}\right)^{1/2} & \text{(radiation domination).} \end{cases}$$

## Models with vacuum energy

The solution of the Friedmann equation becomes more complicated if we allow a significant contribution from vacuum energy—i.e. a non-zero cosmological constant. The Friedmann equation itself is independent of the equation of state, and just says $H^2R^2 = kc^2/(\Omega-1)$, whatever the form of the contributions to $\Omega$. In terms of the cosmological constant itself, we have

$$\Omega_{vac} = \frac{8\pi G\rho_{vac}}{3H^2} = \frac{\Lambda c^2}{3H^2}.$$

The reason that the cosmological constant was first introduced by Einstein was not simply because there was no general reason to expect empty space to be of zero density, but because it allows a non-expanding cosmology to be constructed. This is perhaps not so obvious from some forms of the Friedmann equation, since now $H = 0$ and $\Omega = \infty$; if we cast the equation in its original form without defining these parameters, then zero expansion implies

$$\rho = \frac{3kc^2}{8\pi GR^2}.$$

Since $\Lambda$ can have either sign, this appears not to constrain $k$. However, we also want to have zero acceleration for this model, and so need the time derivative of the Friedmann equation: $\ddot{R} = -4\pi GR(\rho + 3p)/3$. A further condition for a static model is therefore that

$$\rho = -3p.$$

Since $\rho = -p$ for vacuum energy, and this is the only source of pressure if we ignore radiation, this tells us that $\rho = 3\rho_{vac}$ and hence that the mass density is twice the vacuum density. The total density is hence positive and $k = 1$; we have a closed model.

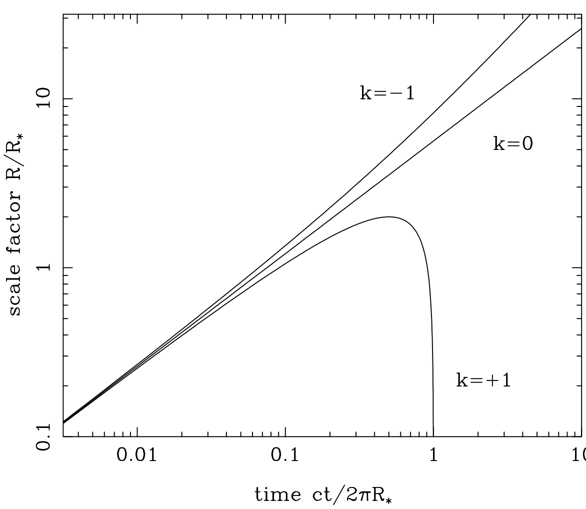


**Figure 1.** The time dependence of the scale factor for open, closed and critical matter-dominated cosmological models. The logarithmic scale is designed to bring out the early-time behavior, although it obscures the fact that the closed model is a symmetric cycloid on a linear plot of $R$ against $t$. Reproduced from *Cosmological Physics* (Cambridge University Press).

Notice that what this says is that a positive vacuum energy acts in a repulsive way, balancing the attraction of normal matter. This shows that the static model cannot be stable: if we perturb the scale factor by a small positive amount, the vacuum repulsion is unchanged whereas the 'normal' gravitational attraction is reduced, so that the model will tend to expand further (or contract, if the initial perturbation was negative).

### de Sitter space

The endpoint of an outwards perturbation of Einstein's static model was first studied by DE SITTER. This universe is completely dominated by vacuum energy and is clearly the limit of the unstable expansion, since the density of matter redshifts to zero while the vacuum energy remains constant. Consider again the Friedmann equation in its general form $\dot{R}^2 - 8\pi G\rho R^2/3 = -kc^2$: since the density is constant and $R$ will increase without limit, the two terms on the lhs must eventually become almost exactly equal and the curvature term on the rhs will be negligible. Thus, even if $k \neq 0$, the universe will have a density that differs only infinitesimally from the critical, so that we can solve the equation by setting $k = 0$, in which case

$$R \propto \exp(Ht) \qquad H = \left(\frac{8\pi G\rho_{\text{vac}}}{3}\right)^{1/2} = \left(\frac{\Lambda c^2}{3}\right)^{1/2}.$$

An interesting interpretation of this behavior was promoted in the early days of cosmology by Eddington: the cosmological constant is what caused the expansion. In models without $\Lambda$, the expansion is merely an initial condition: anyone who asks why the universe expands at a given epoch is given the unsatisfactory reply that it does so because it was expanding at some earlier time. It would be more satisfying to have some mechanism that set the expansion into motion, and this is what is provided by vacuum repulsion. This tendency of models with positive $\Lambda$ to end up undergoing an exponential phase of expansion (and moreover one with $\Omega = 1$) is exactly what is used in inflationary cosmology to generate the initial conditions for the big bang.

### The steady-state model

The behavior of de Sitter space is in some ways reminiscent of the steady-state universe, which was popular in the 1960s. This STEADY-STATE THEORY drew its motivation from the philosophical problems of big-bang models—which begin in a singularity at $t = 0$, and for which earlier times have no meaning. Instead, Hoyle, Bondi and Gold suggested the perfect cosmological principle in which the universe is homogeneous not only in space but also in time: apart from local fluctuations, the universe appears the same to all observers at all times. This tells us that the Hubble constant really is constant, and so the model necessarily has exponential expansion, $R : \exp(Ht)$, exactly as for de Sitter space. Indeed, de Sitter space is a steady-state universe: it contains a constant vacuum energy density and has an infinite age, lacking any big-bang singularity. However, de Sitter space is a rather uninteresting model because it contains no matter. Introducing matter into a steady-state universe violates energy conservation, since matter does not have the $p = -\rho c^2$ equation of state that allows the density to remain constant. This is the most radical aspect of steady-state models: they require continuous creation of matter. The energy to accomplish this has to come from somewhere, and Einstein's equations are modified by adding some 'creation' or '$C$-field' term to the energy–momentum tensor:

$$T'^{\mu\nu} = T^{\mu\nu} + C^{\mu\nu} \qquad T'^{\mu\nu}_{;\nu} = 0.$$

The effect of this extra term must be to cancel the matter density and pressure, leaving just the overall effective form of the vacuum tensor, which is required to produce de Sitter space and the exponential expansion. This *ad hoc* field and the lack of any physical motivation for it beyond the cosmological problem it was designed to solve was always the most unsatisfactory feature of the steady-state model, and may account for the strong reactions generated by the theory.

### Bouncing and loitering models

Returning to the general case of models with a mixture of energy in the vacuum and normal components, we have to distinguish three cases. For models that start from a big bang (in which case radiation dominates completely at the earliest times), the universe will either recollapse or expand forever. The latter outcome becomes more likely for low densities of matter and radiation, but high vacuum density. It is, however, also possible to have models in which there is no big bang: the universe was collapsing in the distant past, but was slowed by the repulsion of a positive $\Lambda$ term and underwent a 'bounce' to reach its present state of expansion. Working out the conditions for these different events is a matter of integrating the Friedmann equation. For the addition of $\Lambda$, this can only in general be done numerically. However, we can find the conditions for the different behaviors described above analytically, at least if we simplify things by ignoring radiation. The equation in the form of the time-dependent Hubble parameter looks like

$$\frac{H^2}{H_0^2} = \Omega_v(1 - a^{-2}) + \Omega_m(a^{-3} - a^{-2}) + a^{-2}$$

and we are interested in the conditions under which the lhs vanishes, defining a turning point in the expansion. Setting the rhs to zero yields a cubic equation, and it is possible to give the conditions under which this has a solution, which are as follows.

(1) First, negative $\Lambda$ always implies recollapse, which is intuitively reasonable (either the mass causes recollapse before $\Lambda$ dominates, or the density is low enough that $\Lambda$ comes to dominate, which cannot lead to infinite expansion unless $\Lambda$ is positive).
(2) If $\Lambda$ is positive and $\Omega_m<1$, the model always expands to infinity.
(3) If $\Omega_m>1$, recollapse is only avoided if $\Omega_v$ exceeds a critical value

$$\Omega_v > 4\Omega_m\{\cos[\tfrac{1}{3}\cos^{-1}(\Omega_m^{-1}-1)+\tfrac{4}{3}\pi]\}^3.$$

(4) If $\Lambda$ is large enough, the stationary point of the expansion is at $a<1$ and we have a bounce cosmology. This critical value is

$$\Omega_v > 4\Omega_m\{f[\tfrac{1}{3}f^{-1}(\Omega_m^{-1}-1)]\}^3$$

where the function $f$ is cosh if $\Omega_m<0.5$, otherwise cos. If the universe lies exactly on the critical line, the bounce is at infinitely early times and we have a solution that is the result of a perturbation of the Einstein static model.

In fact, bounce models can be ruled out quite strongly. The same cubic equations that define the critical conditions for a bounce also give an inequality for the maximum redshift possible (that of the bounce):

$$1+z_B \leqslant 2f[\tfrac{1}{3}f^{-1}(\Omega_m^{-1}-1)].$$

A reasonable lower limit for $\Omega_m$ of 0.1 then rules out a bounce once objects are seen at $z>2$.

The main results of this section are summed up in figure 2. Since the radiation density is very small today, the main task of relativistic cosmology is to work out where on the $\Omega_{matter}$–$\Omega_{vacuum}$ plane the real universe lies. The existence of high-redshift objects rules out the bounce models, so that the idea of a hot big bang cannot be evaded.

At this point, we have reproduced one of the great conclusions of relativistic cosmology: the universe is of finite age, and had its origin in a mathematical singularity at which the scale factor went to zero, leading to a divergent spacetime curvature. Since zero scale factor also implies infinite density (and temperature), the inferred picture of the early universe is one of unimaginable violence. The term big bang was coined by FRED HOYLE to describe this beginning, although it was intended somewhat critically. The problem with the singularity is that it marks the breakdown of the laws of physics; we cannot extrapolate the solution for $R(t)$ to $t<0$, and so the origin of the expansion becomes an unexplained boundary condition. It was only after about 1980 that a consistent set of ideas became available for ways of avoiding this barrier, in the form of inflationary cosmology.

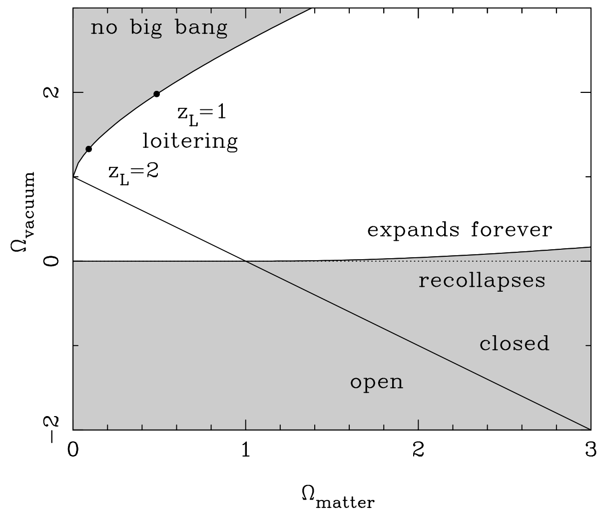


**Figure 2.** This plot shows the different possibilities for the cosmological expansion as a function of matter density and vacuum energy. Models with total $\Omega>1$ are always spatially closed (open for $\Omega<1$), although closed models can still expand to infinity if $\Omega_v\neq0$. If the cosmological constant is negative, recollapse always occurs; recollapse is also possible with a positive $\Omega_v$ if $\Omega_m\gg\Omega_v$. If $\Omega_v>1$ and $\Omega_m$ 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. Reproduced from *Cosmological Physics* (Cambridge University Press).

## Flat universe

The most important model in cosmological research is that with $k=0\Rightarrow\Omega_{total}=1$; when dominated by matter, this is often termed the Einstein–de Sitter model. Paradoxically, this importance arises because it is an unstable state: as we have seen earlier, the universe will evolve away from $\Omega=1$, given a slight perturbation. For the universe to have expanded by so many e-foldings (factors of e expansion) and yet still have Ωd1 implies that it was very close to being spatially flat at early times. Many workers have therefore conjectured that it would be contrived if this flatness was other than perfect.

An alternative $k=0$ model of greater observational interest has a significant cosmological constant, so that $\Omega_m+\Omega v=1$ (radiation being neglected for simplicity). This may seem contrived, but once $k=0$ has been established, it cannot change: individual contributions to $\Omega$ must adjust to keep in balance. The advantage of this model is that it is the only way of retaining the theoretical attractiveness of $k=0$ while changing the age of the universe from the relation $H_0t_0=2/3$, which characterizes the Einstein–de Sitter model. Since much observational evidence indicates that $H_0t_0\simeq1$, this model has received a good deal of interest in recent years. For this case, the Friedmann equation is

$$\dot{a}^2 = H_0^2[\Omega_m a^{-1}+(1-\Omega_m)a^2]$$

and the $t(a)$ relation is

$$H_0 t(a) = \int_0^a \frac{x\,\mathrm{d}x}{[\Omega_m x + (1-\Omega_m)x^4]^{1/2}}$$

which integrates to

$$H_0 t(a) = \frac{2}{3}\frac{S_k^{-1}[(a^3|\Omega_m - 1|/\Omega_m)^{1/2}]}{(|\Omega_m - 1|)^{1/2}}.$$

Here, $k$ in $S_k$ is used to mean sin if $\Omega_m>1$, otherwise sinh; these are still $k = 0$ models. This $t(a)$ relation is compared with models without vacuum energy in figure 3. Since there is nothing special about the current era, we can clearly also rewrite this expression as

$$\begin{aligned} H(a)t(a) &= \frac{2}{3}\frac{S_k^{-1}\{[|\Omega_m(a) - 1|/\Omega_m(a)]^{1/2}\}}{[|\Omega_m(a) - 1|]^{1/2}} \\ &\simeq \frac{2}{3}\Omega_m(a)^{-0.3} \end{aligned}$$

where we include a simple approximation that is accurate to a few % over the region of interest ($\Omega_m \geqq 0.1$). In the general case of significant $\Lambda$ but $k \simeq 0$, this expression still gives a very good approximation to the exact result, provided that $\Omega_m$ is replaced by $0.7\Omega_m - 0.3\Omega_v + 0.3$ (Carroll *et al* 1992).

## Horizons

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

$$\Delta r = \int_{t_0}^{t_1} \frac{c\,\mathrm{d}t}{R(t)} = \Delta\eta$$

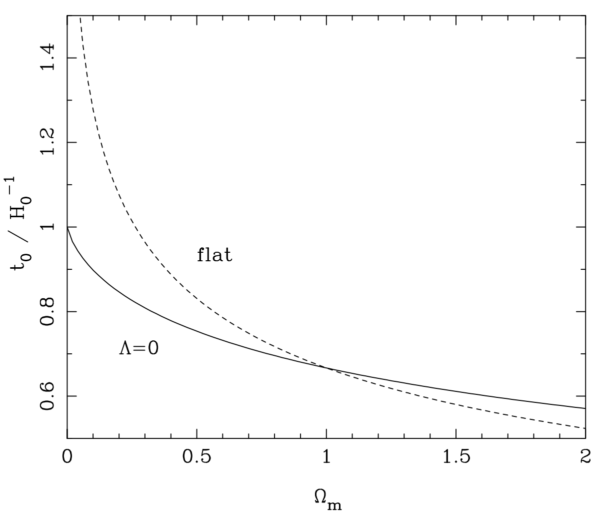


**Figure 3** Age of the universe versus matter density parameter. The solid curve shows open models; the dotted curve shows flat models with $\Omega_m + \Omega_v = 1$. The accelerating effects of vacuum energy allow $H_0 t_0 >\approx 1$ for $\Omega_m >\approx 0.25$ in the flat case, which is much older than the open model of the same density. Reproduced from *Cosmological Physics* (Cambridge University Press).

i.e. just the interval of conformal time. What happens as $t_0 \to 0$ in this expression? We can replace $\mathrm{d}t$ by $\mathrm{d}R/\dot{R}$, which the Friedmann equation says is $:\mathrm{d}R/(\rho R^2)^{1/2}$ at early times. Thus, this integral converges if $\rho R^2 \to \infty$ as $t_0 \to 0$, otherwise it diverges. Provided that the equation of state is such that $\rho$ changes faster than $R^{-2}$, light signals can only propagate a finite distance between the big bang and the present; there is then said to be a particle horizon. Such a horizon therefore exists in conventional big-bang models, which are dominated by radiation at early times.

A particle horizon is not at all the same thing as an event horizon: for the latter, we ask whether $\Delta r$ diverges as $t \to \infty$. If it does, then seeing a given event is just a question of waiting long enough. Clearly, an event horizon requires $R(t)$ to increase more quickly than $t$, so that distant parts of the universe recede faster than light. This does not occur unless the universe is dominated by vacuum energy at late times. Despite this distinction, cosmologists usually say 'the horizon' when they mean 'the particle horizon'.

## Observations in cosmology

The main task of observational cosmology is to use observed quantities such as fluxes and angular sizes in order to deduce intrinsic quantities such as luminosities and physical sizes for a distant object. This conversion requires a knowledge of distance, which is not directly measurable; all that we know about a distant object is its redshift. Observers therefore place heavy reliance on formulae for expressing distance in terms of redshift. For high-redshift objects such as quasars, this has led to a history of controversy over whether a component of the redshift could be of non-cosmological origin:

$$1 + z_{obs} = (1 + z_{cos})(1 + z_{extra}).$$

For now, we assume that the cosmological contribution to the redshift can always be identified.

We now assemble some essential formulae for interpreting cosmological observations. Our observables are redshift, $z$, and angular difference between two points on the sky, $\mathrm{d}\psi$. We write the metric in the form

$$c^2\,\mathrm{d}\tau^2 = c^2\,\mathrm{d}t^2 - R^2(t)[\mathrm{d}r^2 + S_k^2(r)\,\mathrm{d}\psi^2]$$

so that the comoving volume element is

$$\mathrm{d}V = 4\pi[R_0 S_k(r)]^2\ R_0\,\mathrm{d}r.$$

The proper transverse size of an object seen by us is its comoving size $\mathrm{d}\psi$; $S_k(r)$ times the scale factor at the time of emission:

$$\mathrm{d}\ell = \mathrm{d}\psi\ R_0 S_k(r)/(1+z).$$

Probably the most important relation for observational cosmology is that between monochromatic flux density and luminosity. Start by assuming isotropic emission, so that the photons emitted by the source pass with a uniform flux density through any sphere surrounding the source. We can now make a shift of origin, and consider the Robertson–Walker metric as being centered on the source; however, because of homogeneity, the comoving distance between the source and the observer is the same as we would calculate when we place the origin at our location. The photons from the source are therefore passing through a sphere, on which we sit, of proper surface area $4\pi[R_0 S_k(r)]^2$. However, redshift still affects the flux density in four further ways: photon energies and arrival rates are redshifted, reducing the flux density by a factor $(1 + z)^2$; opposing this, the bandwidth $d\nu$ is reduced by a factor $1 + z$, so the energy flux per unit bandwidth goes down by one power of $1 + z$; finally, the observed photons at frequency $\nu_0$ were emitted at frequency $\nu_0(1+z)$, so the flux density is the luminosity at this frequency, divided by the total area, divided by $1 + z$:

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

A word about units: $L_\nu$ in this equation would be measured in units of W Hz$^{-1}$. Recognizing that emission is often not isotropic, it is common to consider instead the luminosity emitted into unit solid angle—in which case there would be no factor of $4\pi$, and the units of $L_\nu$ would be W Hz$^{-1}$ sr$^{-1}$.

The flux density received by a given observer can be expressed by definition as the product of the specific intensity $I_\nu$ (the flux density received from unit solid angle of the sky) and the solid angle subtended by the source: $S_\nu = I_\nu\, d\Omega$. Combining the angular size and flux–density relations thus gives the relativistic version of surface-brightness conservation. This is independent of cosmology:

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

where $B_\nu$ is surface brightness (luminosity emitted into unit solid angle per unit area of source). We can integrate over $\nu_0$ to obtain the corresponding total or bolometric formulae, which are needed for example for spectral-line emission: The form of these relations lead to

$$S_{\rm tot} = \frac{L_{\rm tot}}{4\pi R_0^2 S_k^2(r)(1+z)^2}$$

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

the following definitions for particular kinds of distances: angular-diameter distance is

$$D_{\rm A} = (1+z)^{-1} R_0 S_k(r)$$

luminosity distance is

$$D_{\rm L} = (1+z)\; R_0 S_k(r).$$

Angular-diameter distance versus redshift is illustrated in figure 4.

The last element needed for the analysis of observations is a relation between redshift and age for the object being studied. This brings in our earlier relation between time and comoving radius (consider a null geodesic traversed by a photon that arrives at the present):

$$c\,\mathrm{d}t = R_0\,\mathrm{d}r/(1+z).$$

So far, all this is completely general; to complete the toolkit, we need the crucial input of relativistic dynamics, which is to give the distance–redshift relation.

## Distance–redshift relation

The general relation between comoving distance and redshift was given earlier as

$$\begin{aligned} R_0\,\mathrm{d}r &= \frac{c}{H(z)}\,\mathrm{d}z \\ &= \frac{c}{H_0}[\Omega_{\rm v} + \Omega_{\rm m}a^{-3} + \Omega_{\rm r}a^{-4} - (\Omega-1)a^{-2}]^{-1/2}\,\mathrm{d}z. \end{aligned}$$

For a matter-dominated Friedmann model, this means that the distance of an object from which we receive photons today is

$$R_0 r = \frac{c}{H_0}\int_0^z \frac{\mathrm{d}z'}{(1+z')(1+\Omega z')^{1/2}}.$$

Integrals of this form often arise when manipulating Friedmann models; they can usually be tackled by the substitution $u^2 = k(\Omega-1)/\Omega(1+z)$. This substitution produces Mattig's formula, which is one of the single most useful equations in cosmology as far as observers are concerned:

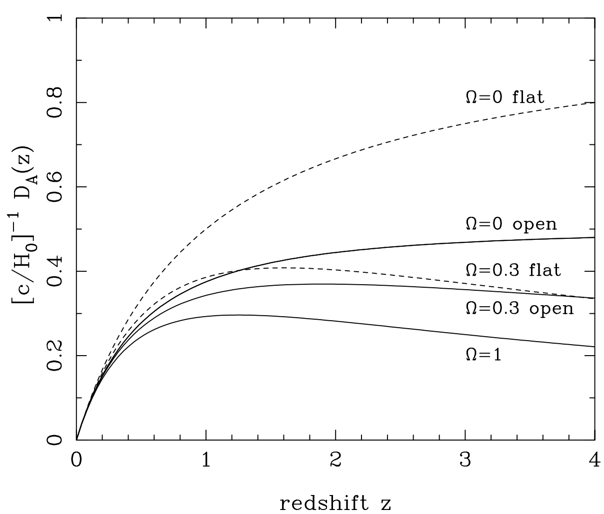


**Figure 4.** A plot of dimensionless angular-diameter distance versus redshift for various cosmologies. Solid curves show models with zero vacuum energy; dashed curves show flat models with $\Omega_{\rm m} + \Omega_v = 1$. In both cases, results for $\Omega_{\rm m} = 1, 0.3, 0$ are shown; higher density results in lower distance at high $z$, owing to gravitational focusing of light rays. Reproduced from *Cosmological Physics* (Cambridge University Press).

$$R_0 S_k(r) = \frac{2c}{H_0} \frac{\Omega z + (\Omega - 2)[(1 + \Omega z)^{1/2} - 1]}{\Omega^2(1 + z)}.$$

There is no such compact expression if one wishes to allow for vacuum energy as well. The comoving distance has to be obtained by numerical integration of the fundamental $dr/dz$, even in the $k = 0$ case. However, for all forms of contribution to the energy content of the universe, the second-order distance–redshift relation is identical, and depends only on the deceleration parameter:

$$R_0 S_k(r) \simeq \frac{c}{H_0} \left( z - \frac{1 + q_0}{2} z^2 \right).$$

The sizes and flux densities of objects at moderate redshift therefore determine the geometry of the universe only once an equation of state is assumed, so that $q_0$ and $\Omega_0$ can be related. At larger redshifts, this degeneracy is broken, and accurate measurements of the distance–redshift relation can in principle determine the parameters $\Omega_m$, $\Omega_v$ etc independently.

### Recent observations

NASA's Far Ultraviolet Spectroscopic Explorer (FUSE) satellite has given astronomers a glimpse of the ghostly cobweb of helium gas left over from the Big Bang, which underlies the universe's structure. The helium is not found in galaxies or stars but spread thinly through space. The observations help confirm theoretical models of how matter in the expanding universe condensed into a web-like structure pervading all the space between galaxies. The helium traces the architecture of the universe back to very early times. This structure arose from small gravitational instabilities seeded in the chaos just after the Big Bang.

### Outstanding issues

As outlined above, the basic isotropic models of relativistic cosmology depend on four main parameters: the present rate of expansion and the present contributions to the total density of non-relativistic matter, ultrarelativistic matter and vacuum. The appearance of distant objects depends on these numbers, so in principle it is possible to determine these parameters, and also to investigate the weakest class of non-standard cosmologies—in which there may exist additional contributions to the density, with more exotic equations of state. All such possibilities can be investigated empirically within the framework of FRW models.

A deeper issue is to ask whether the basic assumption of isotropy and homogeneity is valid, and if so, why this should be. Studies of large-scale structure and anisotropies in the microwave background suggest that deviations from the Robertson–Walker metric are limited to fractional perturbations at about the $10^{-5}$ level, so the basic metric seems a good zero-order model. However, the existence of a particle horizon that is small at early times means that it is a major surprise to find the universe to be nearly homogeneous over regions that have only recently come into causal contact. This is one of a number of peculiarities in the standard model that cry out for explanation: like the basic fact of expansion and the near-perfect flatness, these are puzzles of the initial conditions that require explanation by a more complete theory, such as INFLATION.

Lastly, there are a number of peculiar features which relate to our status as observers. Unless $\Omega = 1$, we live near a special time—at which the contributions to the Friedmann equation from spatial curvature or vacuum energy are comparable with that of non-relativistic matter. There exists a body of ideas under the heading of the 'ANTHROPIC PRINCIPLE' which attempt to quantify the selection effects imposed by the need for intelligent observers. In some cases, these arguments are relatively uncontroversial: we should not be surprised that the universe is now roughly as old as a typical star, since stars are needed to make the heavy nuclei needed for interesting chemistry. Whether such reasoning explains all the features of the observed universe is likely to remain controversial. However, at a practical level, the standard isotropic cosmological models provide a context within which this difficult debate can at least be conducted with confidence.

### Web update (31 July 2002)

A team of 27 astronomers led by Professor George Efstathiou of the University of Cambridge has published strong evidence for the existence of dark energy using the clustering pattern of 250 000 galaxies in a large volume of the universe surveyed with the Anglo–Australian Telescope at Siding Spring in New South Wales, Australia. By comparing the structure in the universe now, some 15 billion years after the Big Bang, with structure observed in the cosmic microwave background radiation, which preserved information about what the universe was like when it was only 300 000 years old, the Anglo–Australian team could apply a simple geometrical test to elucidate the composition of the universe. Their results show that the universe is full of vacuum energy, completely consistent with the earlier supernovae results.

Christopher Kochanek of the Harvard–Smithsonian Centre for Astrophysics in Cambridge, Massachusetts and Neal Dalal of the University of California, San Diego have used radio telescopes and gravitational lensing to search for cold dark matter. They have studied seven galaxies, each magnified by four nearer ones. Because each lensing galaxy is in a slightly different position, the researchers got four different images of each of the seven distant galaxies. The four images should have been identical. But each is actually slightly different. The difference was enough to have been caused by the kind of clumps of dark matter around lensing galaxies that mathematical models predict.

*Web Update references:*

Efstathiou G *et al* February 2002 *Monthly Notices of the Royal Astronomical Society* **330**, No. 2

Dalal N and Kochanek C S (in press) 2002 Direct detection of CDM substructure. *Astrophysical Journal*

*Bibliography*

Kriss G A *et al* 2001 *Science* **293** 1112–1116

Carroll S, Press W H and Turner E L 1992 *Ann. Rev. Astron. Astrophys.* **30** 499

Peacock J A 1999 *Cosmological Physics* (Cambridge: Cambridge University Press)

Peebles P J E 1993 *Principles of Physical Cosmology* (Princeton, NJ: Princeton University Press)

Weinberg S 1972 *Gravitation and Cosmology* (New York: Wiley)

*John Peacock*