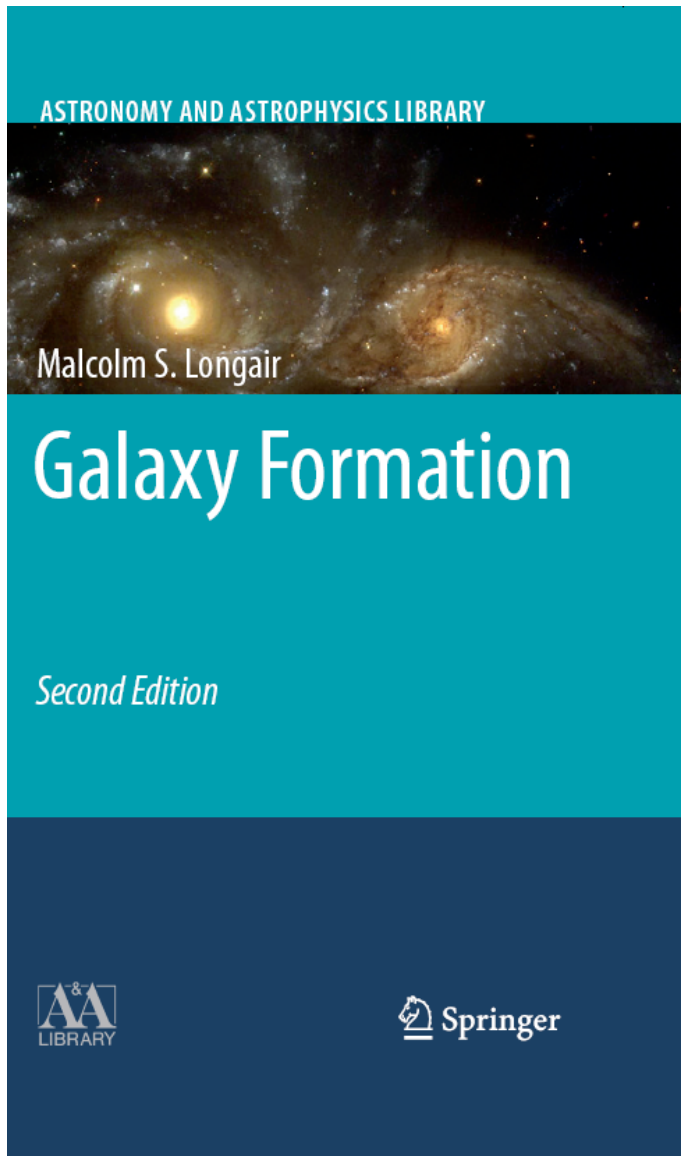


Fundamentals of Cosmology

(1) Basic Physics of Cosmological Models

- Basic observations
- Isotropic curve spaces
- Robertson-Walker metric
- Observations in cosmology
- Radiation-dominated Universes

The Book of the Lectures

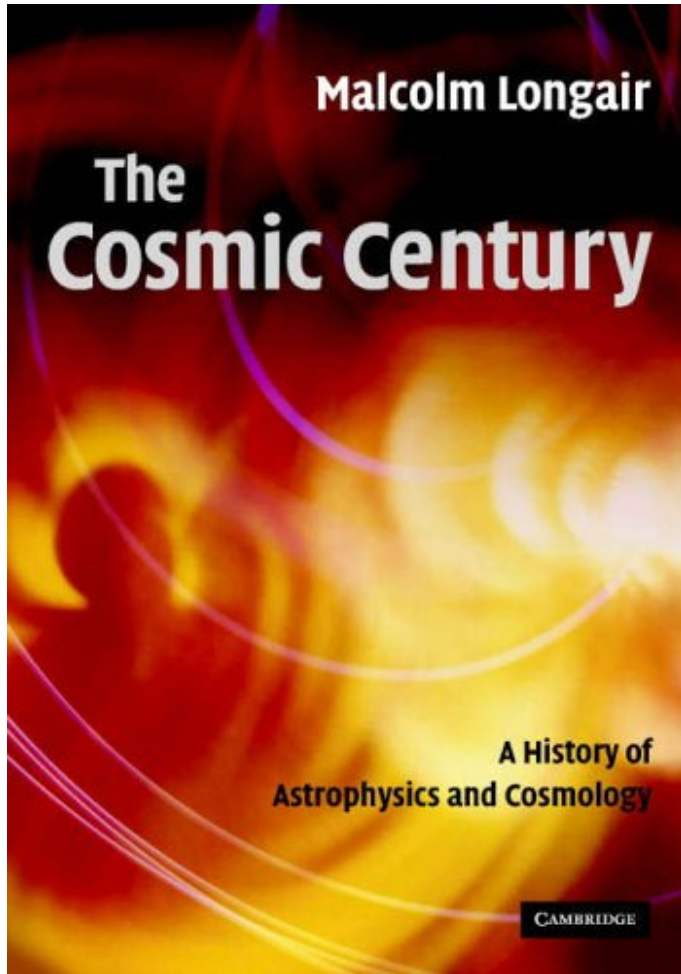


The second edition of *Galaxy Formation* was published by Springer Verlag in 2008. Many more details of the observations and calculations can be found there.

The emphasis is upon the basic physics involved in astrophysical cosmology, trying to keep it as simple, but rigorous, as possible.

In the four lectures, I will also summarise some of the results of the recent explosion of information on observational cosmology and their interpretation.

How It Came About



If you are interested in understanding the history of the development of cosmology, you may find my book, *The Cosmic Century* (2006) useful. It covers the history of cosmology up to 2005.

Part of the value of this approach is that it gives some understanding of the problems which faced the pioneers of cosmology and the numerous wrong turns which were taken.

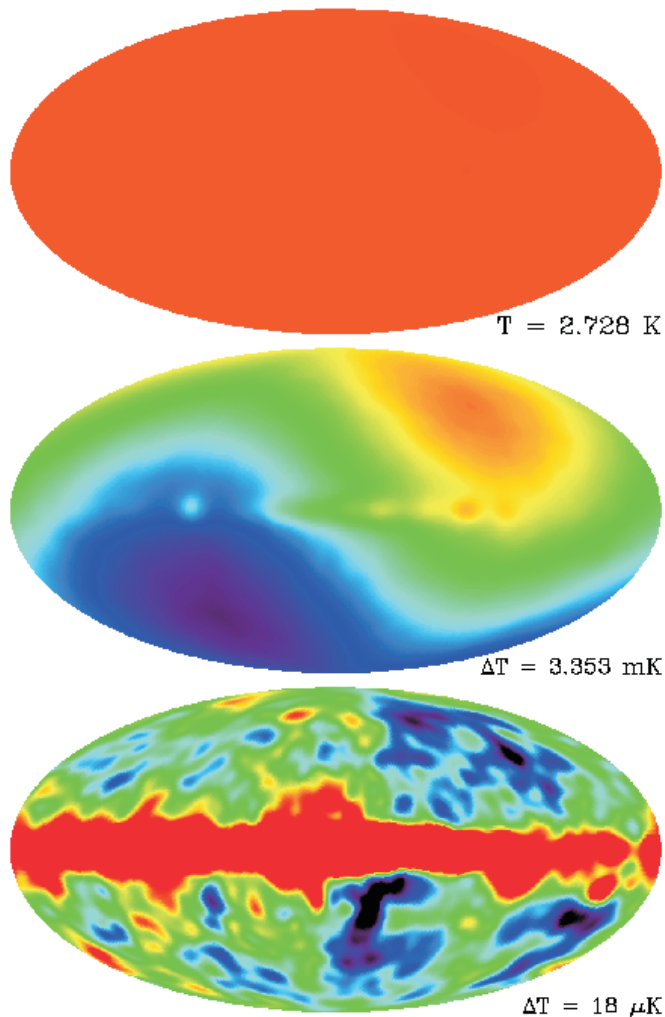
The Basic Structure of Cosmological Models

First, we examine the fundamentals of the cosmological models used in modern cosmology. The models turn out to be remarkably successful, but we need to ask how secure these foundations are. We need to examine:

- Basic observations on which the models are based.
- Basic assumptions made in the construction of cosmological models.
- Observations in Cosmology - how the models really work.

When we come to study specific examples of the models, I will assume the answer for illustrative purposes: $\Omega_0 = 0.3$, $\Omega_\Lambda = 0.7$, $h = 0.7$. We will add in other cosmological parameters later.

COBE Observations of the Cosmic Microwave Background Radiation (1990s)

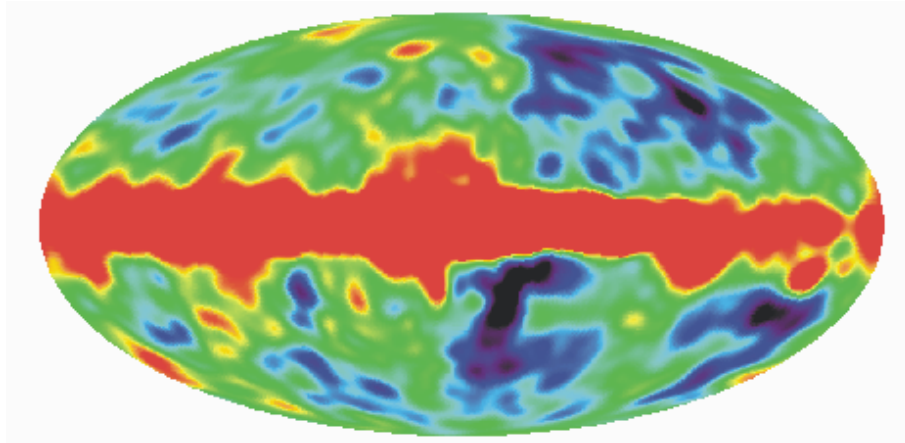


Whole sky in Hammer-Aitoff projection

The starting points for cosmological studies are the observations of the Cosmic Microwave Background Radiation by the COBE satellite.

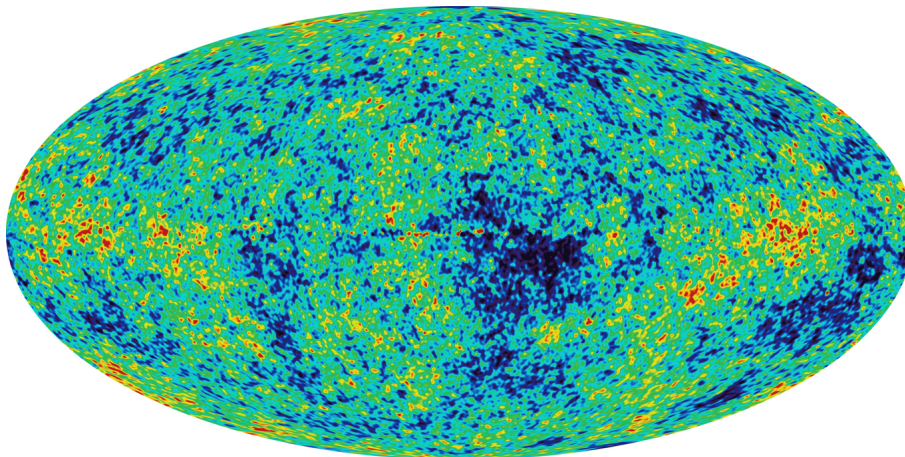
- The spectrum is very precisely that of a perfect black-body at $T = 2.726 \text{ K}$.
- A perfect dipole component is detected, corresponding to the motion of the Earth through the frame in which the radiation would be perfectly isotropic.
- Away from the Galactic plane, the radiation is isotropic to better than one part in 10^5 . Significant temperature fluctuations $\Delta T/T \approx 10^{-5}$ were detected on scales $\theta \geq 10^\circ$.

WMAP Observations of the Cosmic Microwave Background Radiation (2003)



Cosmic Background Explorer (COBE): $\theta = 7^\circ$.

The same features are present in the WMAP image of the sky. The WMAP experiment had much higher angular resolution than COBE.

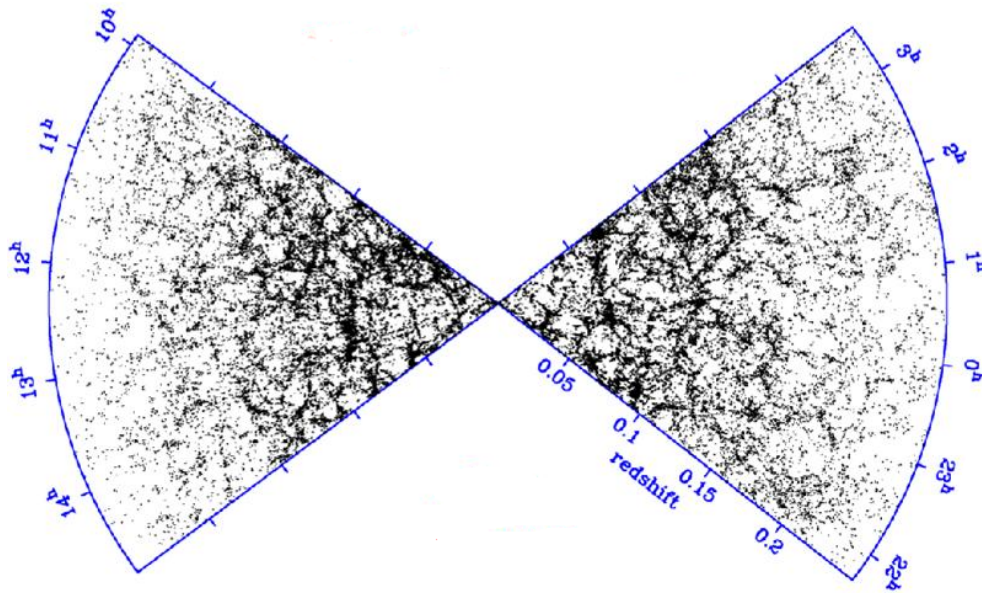
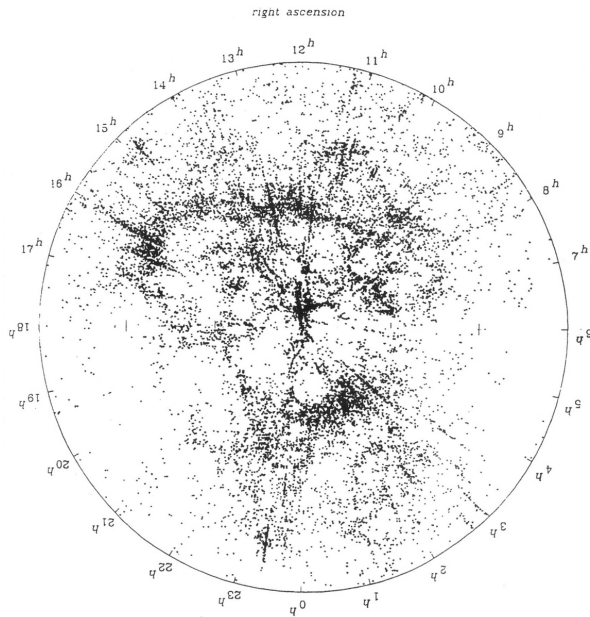


Wilkinson Microwave Anisotropy Probe (WMAP) $\theta = 0.3^\circ$.

Galactic foreground emission has been subtracted.

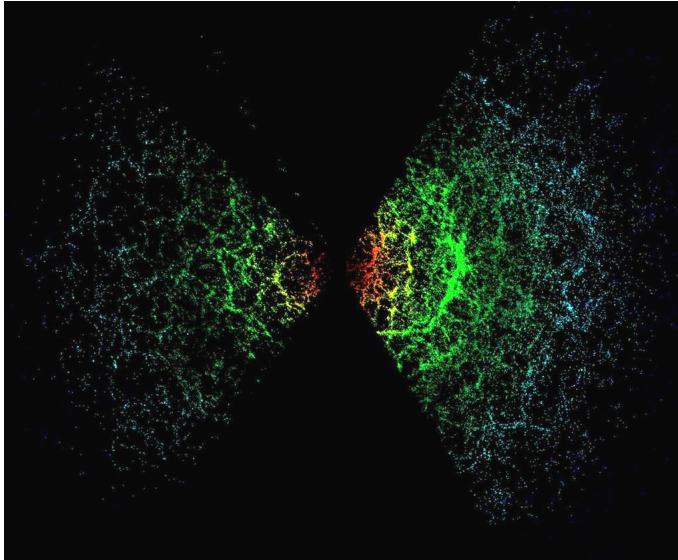
The radiation originated on the last scattering surface at a redshift $z \sim 1000$, or scale factor $a \sim 10^{-3}$.

The Homogeneity of the Universe



The COBE and WMAP observations have established the isotropy of the Universe, but we also need to know about its [homogeneity](#). This has been established by large surveys of galaxies, starting with the local distribution determined by Geller and Huchra (top) and proceeding to the largest scales accessible at the present epoch by the 2dF (bottom) and SDSS surveys which each contain over 200,000 galaxies.

The Homogeneity of the Universe

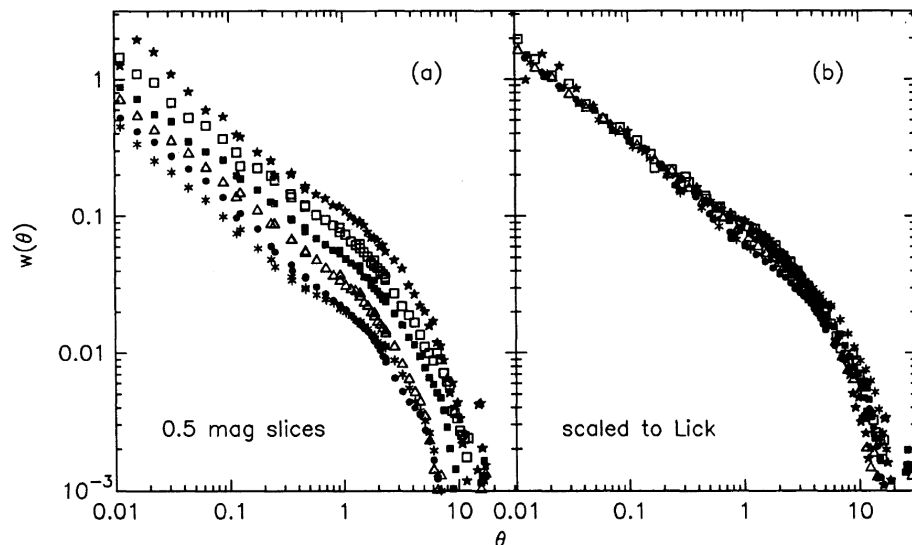


The Sloan Digital Sky Survey

The large scale distribution of galaxies is irregular with giant walls and holes on scales much greater than those of clusters of galaxies.

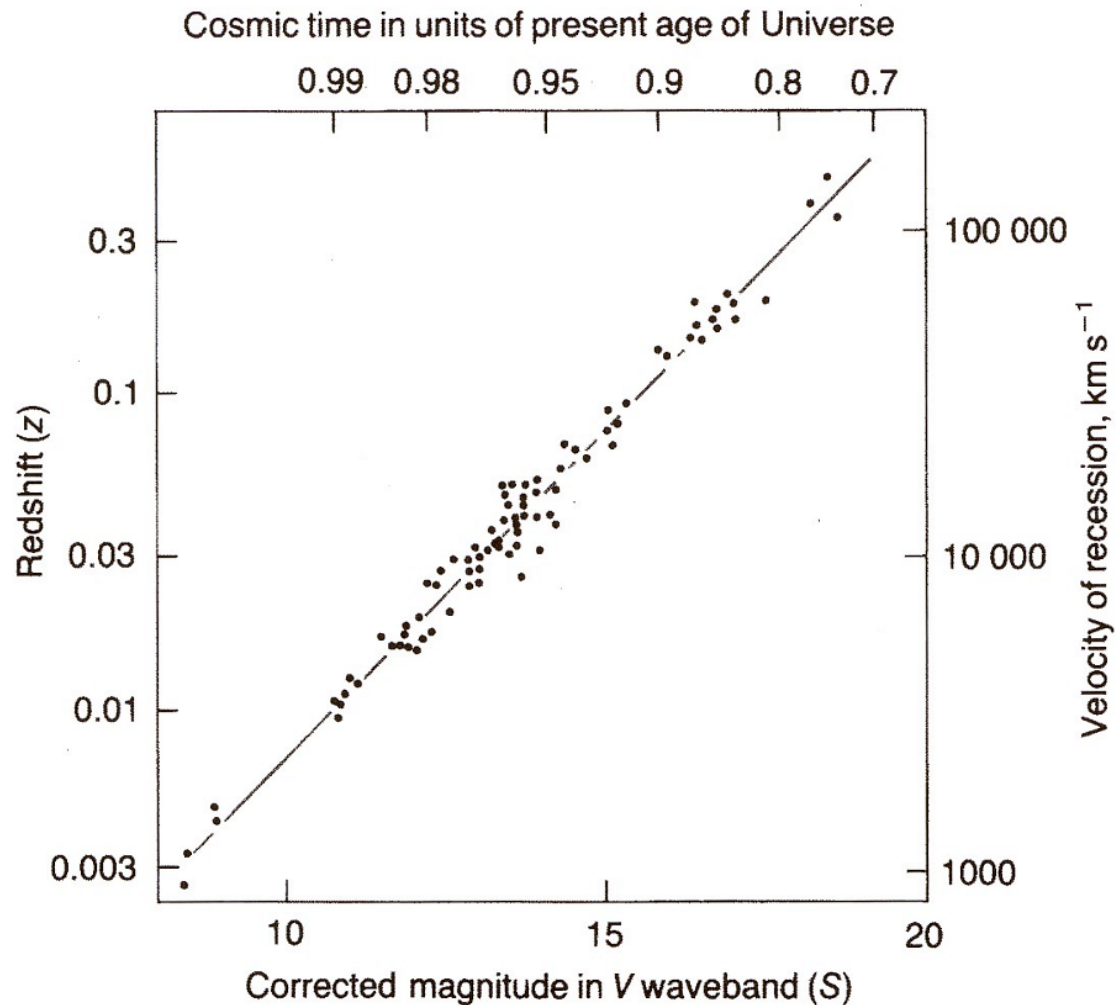
The distributions display, however, the same degree of inhomogeneity as we observe to larger distances in the Universe. This is quantified by the two-point correlation functions for galaxies to different distances,

$$n(r) = n_0[1 + \xi(r)] . \quad (1)$$



Note: we are no longer observing the distant Universe through a non-distorting screen.

Hubble's Law



The second result we need is the redshift-distance relation for galaxies – often called the [Hubble diagram](#).

A modern version of Hubble's law for the brightest galaxies in rich clusters of galaxies, $v = H_0 r$.

All classes of galaxy follow the same Hubble's law. H_0 is [Hubble's constant](#).

This means that the Universe as a whole is expanding uniformly. [Run simulation](#).

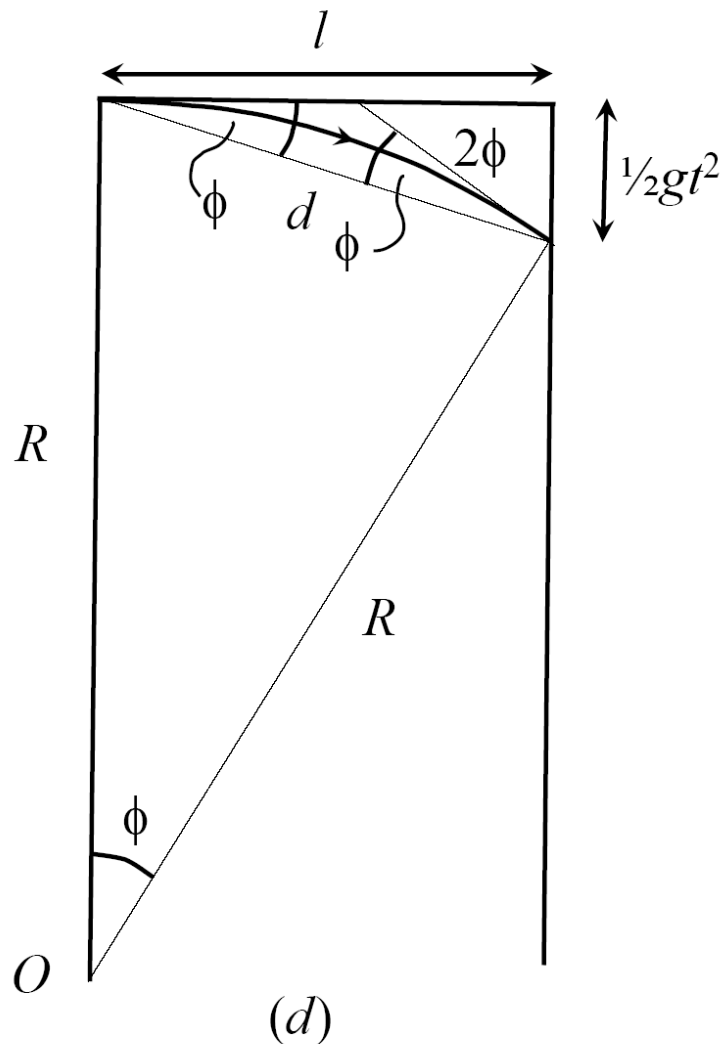
Ingredients of the Standard Cosmological Models

The standard models contain three essential ingredients:

- The *cosmological principle* – **we are at a typical location in the Universe**. Combined with the observations that the Universe is isotropic, homogeneous and uniformly expanding, this leads to the Robertson–Walker metric, **requiring only special relativity + postulates of isotropy and homogeneity**;
- *Weyl's postulate*– the world lines of particles meet at a singular point in the finite or infinite past. This solves the clock synchronisation problem and means that there is a unique world line passing through every point in space-time. The fluid moves along streamlines in the universal expansion and behaves like a perfect fluid with energy–momentum tensor is given by the $T^{\alpha\beta}$;
- Einstein appreciated that, in General Relativity, he had a theory which enabled fully self-consistent models for the Universe as a whole to be constructed.

Let us build up the standard equations by physical arguments.

Isotropic Curved Spaces



A lift accelerated upwards.

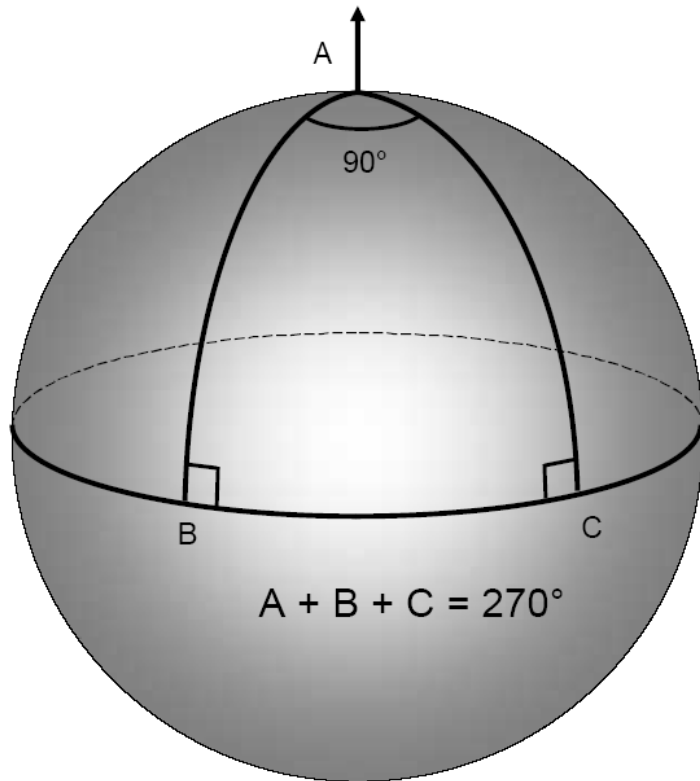
Einstein's great insight in developing General Relativity was that space-time could not be flat if the Principle of Equivalence was to hold good.

In the time the light ray propagates across the lift, a distance l , the lift moves upwards a distance $\frac{1}{2}|g|t^2$. Therefore, in the frame of reference of the accelerated lift, and also in the stationary frame in the gravitational field, the light ray follows a parabolic path. Approximating the light path by a circular arc of radius R , it is straightforward to show that

$$R = \frac{2l^2}{|g|t^2} = \frac{2c^2}{|g|}. \quad (2)$$

The radius of curvature of the path of the light ray depends only upon the local gravitational acceleration $|g|$.

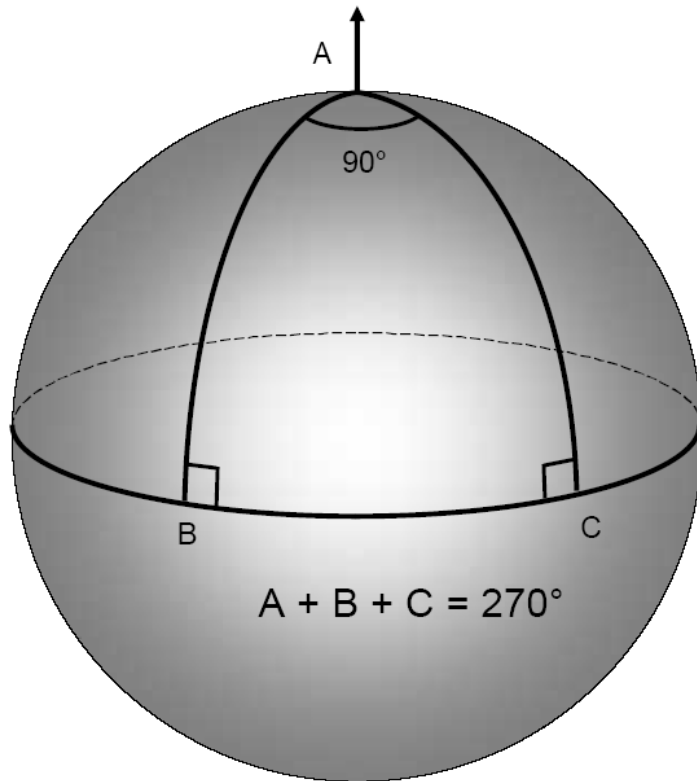
Isotropic Curved Spaces



Consider first the simplest two-dimensional curved geometry, the surface of a sphere. The three sides of this triangle are all segments of great circles on the sphere and so are the shortest distances between the three corners of the triangle. The three lines are **geodesics** in the curved geometry.

We need a procedure for working out how non-Euclidean the curved geometry is. The way this is done in general is by the procedure known as the **parallel displacement** or **parallel transport** of a vector on making a complete circuit around a closed figure such as the triangle. The total rotation of the vector is 270° . Clearly, the surface of the sphere is a non-Euclidean space.

Isotropic Curved Spaces



This procedure illustrates how we can work out the geometrical properties of any two-space, entirely by making measurements within the two-space.

Suppose the angle at A was not 90° but some arbitrary angle θ . Then, if the radius of the sphere is R_C , the surface area of the triangle ABC is

$A = \theta R_C^2$. Thus, if $\theta = 90^\circ$, the area is $\pi R_C^2/2$ and

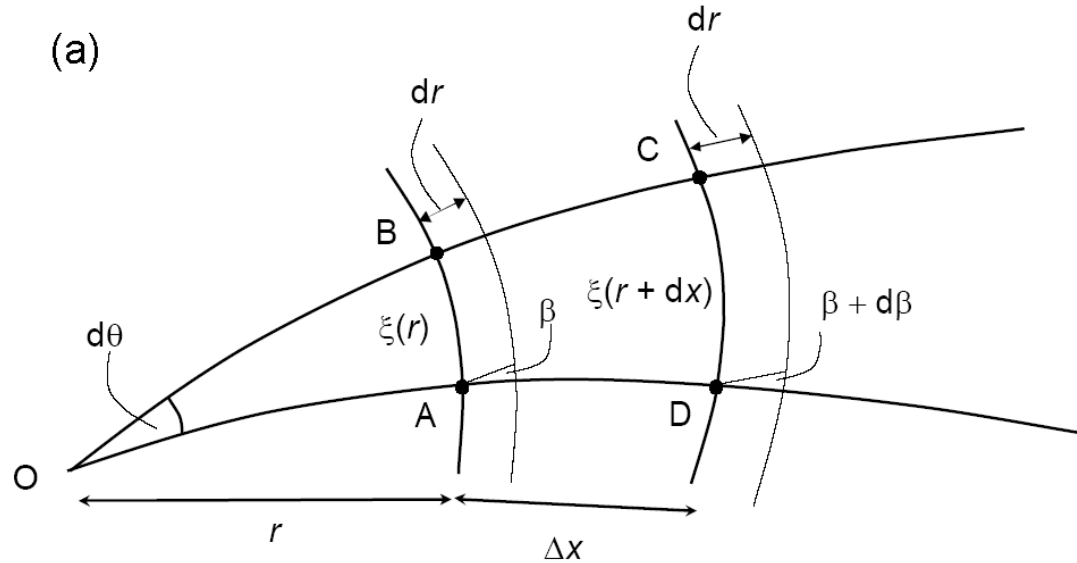
the sum of the angles of the triangle is 270° ; if

$\theta = 0^\circ$, the area is zero and the sum of the angles of the triangle is 180° . The difference of the sum of

the angles of the triangle from 180° is proportional to the area of the triangle, that is

$$(\text{Sum of angles of triangle} - 180^\circ) \propto (\text{Area of triangle}).$$

Isotropic Curved Spaces

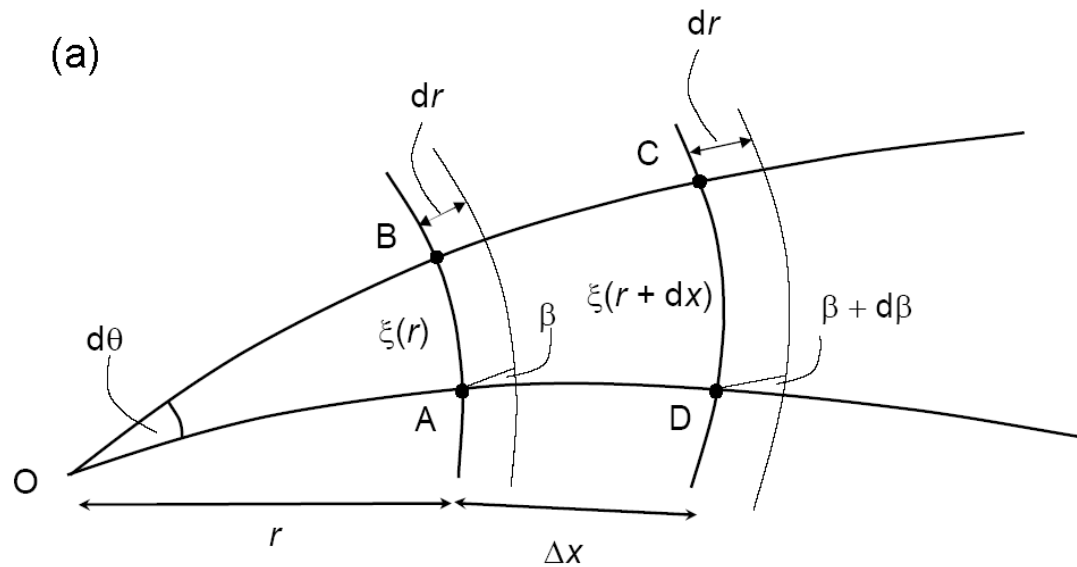


We now work out the sum of the angles round a closed figure in an isotropic curved space. The procedure is shown schematically in the diagram which shows two geodesics from the origin at O being crossed by another pair of geodesics at distances r and $r + \Delta x$ from the origin. The angle $d\theta$ between the geodesics at O is assumed to be small. In Euclidean space, the length of the segment of the geodesic AB would be $\xi = r d\theta$. However, this is no longer true in non-Euclidean space and instead, we write

$$\xi(r) = f(r) d\theta . \quad (3)$$

We now to work out the angle between the diverging geodesics at distance r from the origin.

Isotropic Curved Spaces



It can be seen that the angle between the geodesics is

$$\beta = \frac{\xi(r + dr) - \xi(r)}{dr} = \frac{d\xi(r)}{dr} = d\theta \frac{df(r)}{dr}. \quad (4)$$

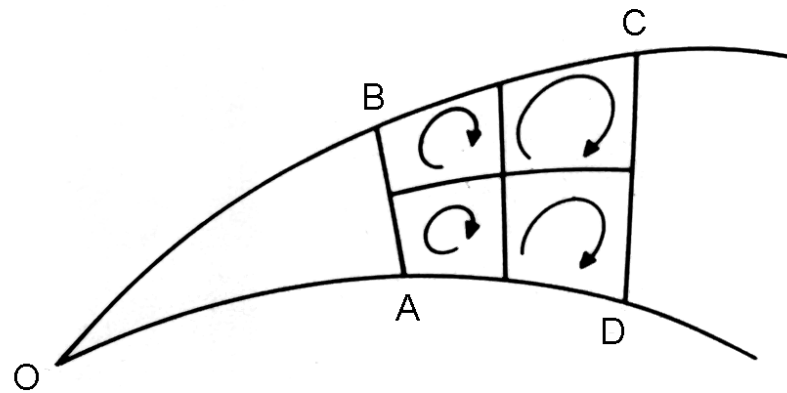
Now move a distance Δx further along the geodesics. The change in the angle β , $\Delta\beta$ is

$$\Delta\beta = \frac{d\xi(r + \Delta x)}{dr} - \frac{d\xi(r)}{dr} = \frac{d^2\xi(r)}{dr^2} \Delta x = \frac{d^2f(r)}{dr^2} \Delta x d\theta. \quad (5)$$

Isotropic Curved Spaces

In Euclidean space, $\xi(r) = f(r) d\theta = r d\theta$, $f(r) = r$ and hence (5) becomes $\beta = d\theta$. Furthermore, in Euclidean space, $d^2 f(r)/dr^2 = 0$ and so $\Delta\beta = 0$, $\beta = d\theta$ remains true for all values of r .

(b)



Now, the rotation of the vector $d\beta$ depends upon the area of the quadrilateral ABCD. In the case of an isotropic space, we should obtain the same rotation wherever we place the loop in the two-space. Furthermore, if we were to split the loop up into a number of sub-loops, the rotations around the separate sub-loops must add up linearly to the total rotation $d\beta$. Thus, in an isotropic two-space, the rotation $d\beta$ should be proportional to the area of the loop ABCD and must be a constant everywhere in the two-space, just as we found in the particular case of a spherical surface.

Isotropic Curved Spaces

The area of the loop is $dA = \xi(r)\Delta x = f(r)\Delta x d\theta$, and so we can write

$$\frac{d^2 f(r)}{dr^2} = -\kappa f(r), \quad (6)$$

where κ is a constant, the minus sign being chosen for convenience. This is the equation of simple harmonic motion which has solution

$$f(r) = A \sin \kappa^{1/2} r. \quad (7)$$

We find the value of A from the expression for $\xi(r)$ for small values of r , which must reduce to the Euclidean expression $d\theta = \xi/r$. Therefore, $A = \kappa^{-1/2}$ and

$$f(r) = \frac{\sin \kappa^{1/2} r}{\kappa^{1/2}}. \quad (8)$$

κ is the **curvature** of the two-space and can be positive, negative or zero. If it is negative, we can write $\kappa = -\kappa'$, where κ' is positive and then the circular functions become hyperbolic functions

$$f(r) = \frac{\sinh \kappa'^{1/2} r}{\kappa'^{1/2}}. \quad (9)$$

Isotropic Curved Spaces

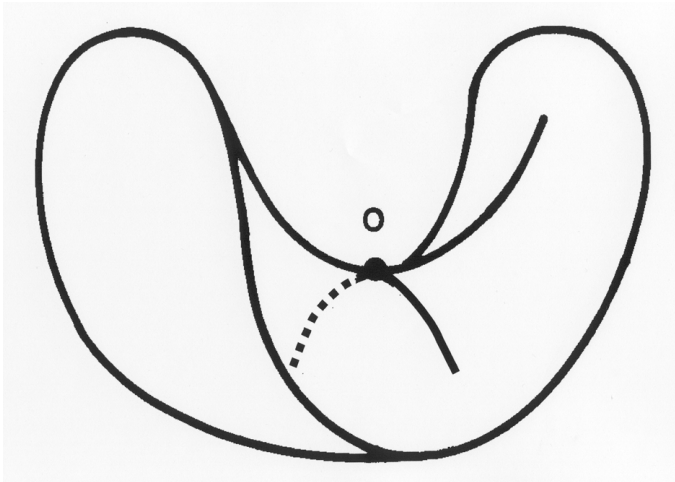
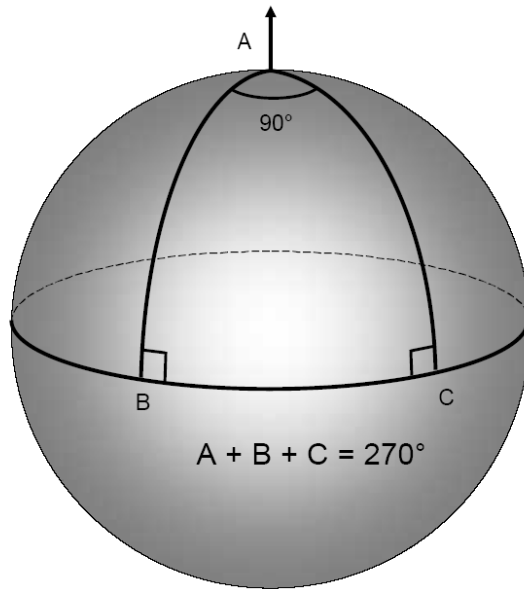
In the Euclidean case, $d^2f(r)/dr^2 = 0$ and so $\kappa = 0$.

The results we have derived include all possible isotropic curved two-spaces. The constant κ can be positive, negative or zero corresponding to spherical, hyperbolic and flat spaces respectively. In geometric terms, $R_C = \kappa^{-1/2}$ is the radius of curvature of a two-dimensional section through the isotropic curved space and has the same value at all points and in all orientations within the plane. It is often convenient to write the expression for $f(r)$ in the form

$$f(r) = R_C \sin \frac{r}{R_C}, \quad (10)$$

where R_C is real for closed spherical geometries, imaginary for open hyperbolic geometries and infinite for the case of Euclidean geometry.

Isotropic Curved Spaces



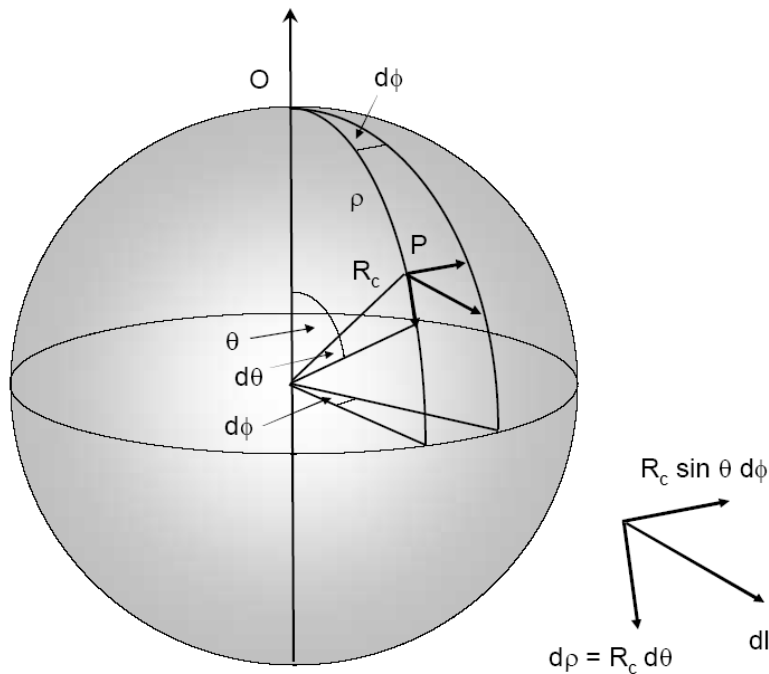
The simplest examples of such spaces are the spherical geometries in which R_C is just the radius of the sphere. The hyperbolic spaces are more difficult to envisage. The fact that R_C is imaginary can be interpreted in terms of the principal radii of curvature of the surface having opposite sign. The geometry of a hyperbolic two-sphere can be represented by a saddle-shaped figure, just as a two-sphere provides an visualisation of the properties of a spherical two-space.

The Space-time Metric for Isotropic Curved Spaces

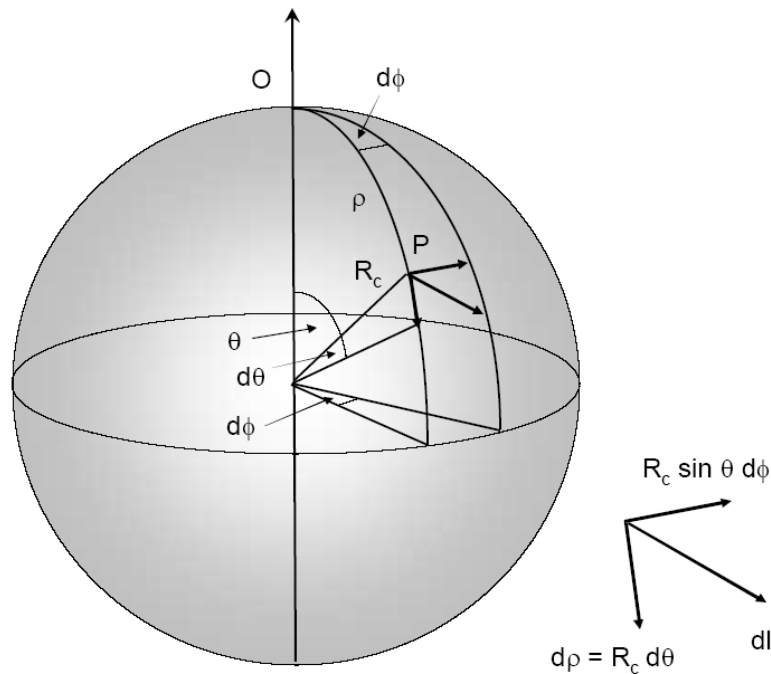
In flat space, the distance between two points separated by dx , dy , dz is

$$dl^2 = dx^2 + dy^2 + dz^2. \quad (11)$$

Let us now consider the simplest example of an isotropic *two-dimensional* curved space, namely the surface of a sphere. We can set up an orthogonal frame of reference at each point locally on the surface of the sphere. It is convenient to work in spherical polar coordinates to describe positions on the surface of the sphere.



The Space-time Metric for Isotropic Curved Spaces



In this case, the orthogonal coordinates are the angular coordinates θ and ϕ , and the expression for the increment of distance dl between two neighbouring points on the surface can be written

$$dl^2 = R_c^2 d\theta^2 + R_c^2 \sin^2 \theta d\phi^2, \quad (12)$$

where R_c is the radius of curvature of the two-space, which in this case is just the radius of the sphere.

The Space-time Metric for Isotropic Curved Spaces

The expression (12) is known as the *metric* of the two-dimensional surface and can be written more generally in tensor form

$$dl^2 = g_{\mu\nu} dx^\mu dx^\nu. \quad (13)$$

It is a fundamental result of differential geometry that the *metric tensor* $g_{\mu\nu}$ contains all the information about the intrinsic geometry of the space. The problem is that we can set up a variety of different coordinate systems to define the coordinates of a point on any two dimensional surface. For example, in the case of a Euclidean plane, we could use rectangular *Cartesian coordinates* so that

$$dl^2 = dx^2 + dy^2, \quad (14)$$

or we could use *polar coordinates* in which

$$dl^2 = dr^2 + r^2 d\phi^2. \quad (15)$$

The Space-time Metric for Isotropic Curved Spaces

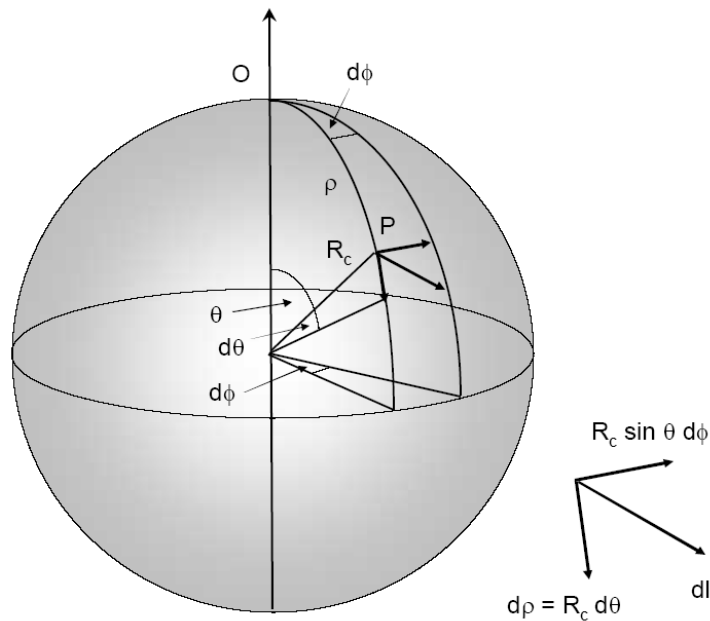
How do we determine the *intrinsic curvature* of the space in terms of the $g_{\mu\nu}$ of the metric tensor? For the case of two-dimensional metric tensors which can be reduced to diagonal form, the intrinsic curvature of the space is given by the quantity

$$\begin{aligned} \kappa = & \frac{1}{2g_{11}g_{22}} \left\{ -\frac{\partial^2 g_{11}}{\partial x_2^2} - \frac{\partial^2 g_{22}}{\partial x_1^2} + \frac{1}{2g_{11}} \left[\frac{\partial g_{11}}{\partial x_1} \frac{\partial g_{22}}{\partial x_1} + \left(\frac{\partial g_{11}}{\partial x_2} \right)^2 \right] \right. \\ & \left. + \frac{1}{2g_{22}} \left[\frac{\partial g_{11}}{\partial x_2} \frac{\partial g_{22}}{\partial x_2} + \left(\frac{\partial g_{22}}{\partial x_1} \right)^2 \right] \right\}. \end{aligned} \quad (16)$$

We can use (16) to show that metrics (14) and (15) have zero curvature and that, for the surface of a sphere, the metric (12) has positive curvature with $\kappa = R_C^{-2}$ at all points on the sphere. κ is known as the *Gaussian curvature* of the two-space and is the same as the definition of the curvature we have already introduced. In general curved spaces, the curvature κ varies from point to point in the space.

The extension to isotropic three-spaces is straightforward if we remember that any two-dimensional section through an isotropic three-space must be an isotropic two-space and we already know the metric tensor for this case.

The Space-time Metric for Isotropic Curved Spaces



The natural system of coordinates for an isotropic two-space is a spherical polar system in which a radial distance ρ round the sphere is measured from the pole and the angle ϕ measures angular displacements at the pole.

The distance ρ round the arc of a great circle from the point O to P is $\rho = \theta R_c$ and so the metric can be written

$$dl^2 = d\rho^2 + R_c^2 \sin^2 \left(\frac{\rho}{R_c} \right) d\phi^2 . \quad (17)$$

The distance ρ is the shortest distance between O and P on the surface of the sphere since it is part of a great circle and is therefore the *geodesic distance* between O and P in the isotropic curved space. Geodesics play the role of straight lines in curved space.

The Space-time Metric for Isotropic Curved Spaces

We can write the metric in an alternative form if we introduce a distance measure

$$x = R_C \sin \left(\frac{\varrho}{R_C} \right) . \quad (18)$$

Differentiating and squaring, we find

$$dx^2 = \left[1 - \sin^2 \left(\frac{\varrho}{R_C} \right) \right] d\varrho^2 \quad d\varrho^2 = \frac{dx^2}{1 - \kappa x^2} , \quad (19)$$

where $\kappa = 1/R_C^2$ is the curvature of the two space.

Therefore, we can rewrite the metric in the form

$$dl^2 = \frac{dx^2}{1 - \kappa x^2} + x^2 d\phi^2 . \quad (20)$$

From the metric (20) $dl = x d\phi$ is a **proper dimension** perpendicular to the radial coordinate ϱ and that it is the correct expression for the length of a line segment which subtends the angle $d\phi$ at geodesic distance ϱ from O. It is therefore what is known as an **angular diameter distance** since it is guaranteed to give the correct answer for the length of a line segment perpendicular to the line of sight.

The Space-time Metric for Isotropic Curved Spaces

We can use either ϱ or x in our metric but, if we use x , the increment of geodesic distance is $d\varrho = dx/(1 - \kappa x^2)^{1/2}$. We recall that the curvature $\kappa = 1/R_C^2$ can be *positive* as in the spherical two-space discussed above, *zero* in which case we recover flat Euclidean space ($R_C \rightarrow \infty$) and *negative* in which case the geometry becomes *hyperbolic* rather than spherical.

We now write down the expression for the spatial increment in any isotropic, three-dimensional curved space. Any two-dimensional section through an isotropic three-space must be an isotropic two-space for which the metric is (17) or (20). In spherical polar coordinates, the general angular displacement perpendicular to the radial direction is

$$d\Phi^2 = d\theta^2 + \sin^2 \theta d\phi^2 . \quad (21)$$

Note that the θ s and ϕ s in (21) are different from those used in the diagram. Thus, by extension of the formalism we have derived already, the spatial increment can be written

$$dl^2 = d\varrho^2 + R_C^2 \sin^2 \left(\frac{\varrho}{R_C} \right) [d\theta^2 + \sin^2 \theta d\phi^2] , \quad (22)$$

in terms of the three-dimensional spherical polar coordinates (ϱ, θ, ϕ) .

The Space-time Metric for Isotropic Curved Spaces

An exactly equivalent form is obtained if we write the spatial increment in terms of x, θ, ϕ in which case we find

$$dl^2 = \frac{dx^2}{1 - \kappa x^2} + x^2[d\theta^2 + \sin^2 \theta d\phi^2]. \quad (23)$$

We are now in a position to write down the *Minkowski metric* in any isotropic three-space. It is given by

$$ds^2 = dt^2 - \frac{1}{c^2}dl^2, \quad (24)$$

where dl is given by either of the above forms of the spatial increment, (22) or (23). Notice that we have to be careful about the meanings of the distance coordinates – x and ρ are equivalent but physically quite distinct distance measures. We can now proceed to derive from this metric the *Robertson–Walker metric*.

The Robertson–Walker Metric

In order to apply the metric (24) to isotropic, homogeneous world models, we need the *cosmological principle* and the concepts of *fundamental observers* and *cosmic time*.

- For uniform, isotropic world models, we define a set of *fundamental observers*, who move in such a way that the Universe always appears to be isotropic to them.
- *Cosmic time* is time measured on the clock of a fundamental observer.

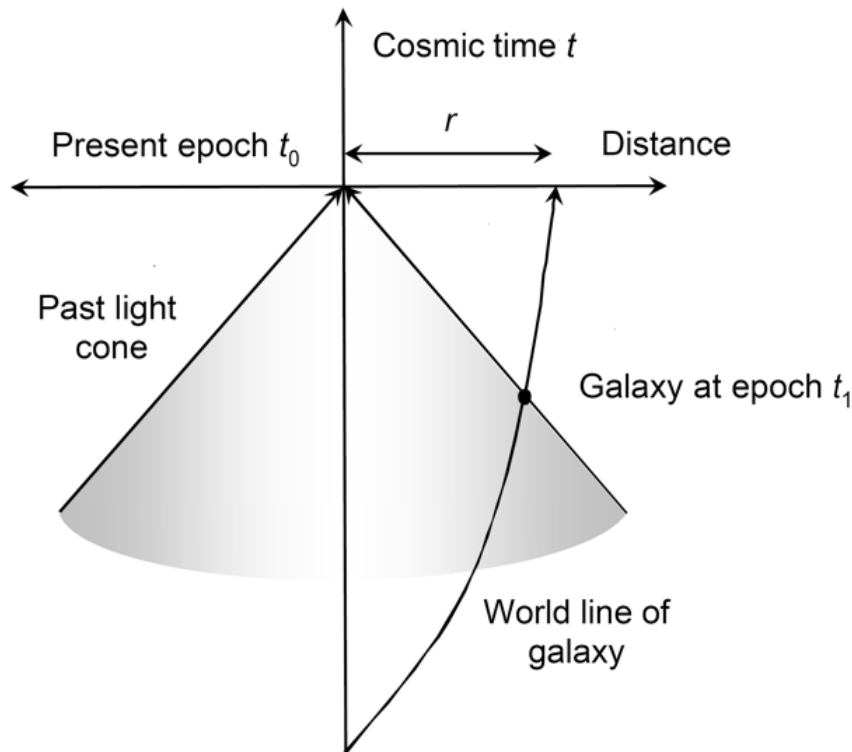
There are no problems of synchronisation of the clocks carried by the fundamental observers because, according to Weyl's postulate, the geodesics of all observers meet at one point in the past and cosmic time can be measured from that reference epoch.

From (22) and (24), the metric can be written in the form

$$ds^2 = dt^2 - \frac{1}{c^2} [d\varrho^2 + R_c^2 \sin^2(\varrho/R_c) (d\theta^2 + \sin^2 \theta d\phi^2)] . \quad (25)$$

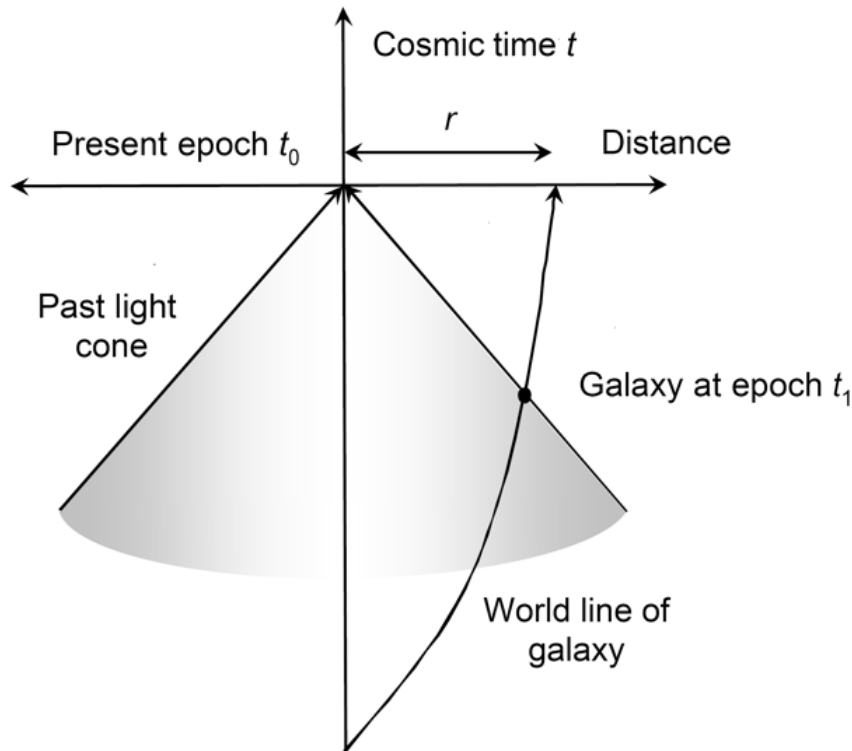
t is cosmic time and $d\varrho$ is an increment of proper distance in the radial direction.

The Robertson-Walker Metric



There is a problem in applying the metric to the expanding Universe as is illustrated by the space-time diagram. Since light travels at a finite speed, we observe all astronomical objects along a *past light cone* which is centred on the Earth at the present epoch t_0 . Therefore, when we observe distant objects, we do not observe them at the present epoch but rather at an earlier epoch t_1 when the distances between fundamental observers were smaller and the spatial curvature different. The problem is that we can only apply the metric (25) to an isotropic curved space defined *at a single epoch*.

The Robertson-Walker Metric



To resolve this problem, we perform the following thought experiment. To measure a proper distance which can be included in the metric (25),

we line up a set of fundamental observers between the Earth and the galaxy whose distance we wish to measure. The observers are instructed to measure the distance d_{ϱ} to the next fundamental observer at a particular cosmic time t . By adding together all the d_{ϱ} s, we can find a proper distance ϱ which is measured *at a single epoch* and which can be used in the metric (25). Notice that ϱ is a *fictitious distance* since we do not know how to project their positions to the present epoch until we know the kinematics of the expanding Universe. Thus, *the distance measure ϱ depends upon the choice of cosmological model.*

The Comoving Distance Coordinate

The definition of a uniform expansion is that between two cosmic epochs, t_1 and t_2 , the distances of any two fundamental observers, i and j , change such that

$$\frac{\varrho_i(t_1)}{\varrho_j(t_1)} = \frac{\varrho_i(t_2)}{\varrho_j(t_2)} = \text{constant} , \quad (26)$$

that is,

$$\frac{\varrho_i(t_1)}{\varrho_i(t_2)} = \frac{\varrho_j(t_1)}{\varrho_j(t_2)} = \dots = \text{constant} = \frac{a(t_1)}{a(t_2)} . \quad (27)$$

For isotropic world models, $a(t)$ is a universal function known as the *scale factor* which describes how the relative distances between *any* two fundamental observers change with cosmic time t . We set $a(t)$ equal to 1 at the present epoch t_0 and let the value of ϱ at the present epoch be r , that is, we can rewrite (27) as

$$\varrho(t) = a(t)r . \quad (28)$$

r thus becomes a *distance label* which is attached to a galaxy or fundamental observer *for all time* and the variation in proper distance in the expanding Universe is taken care of by the scale factor $a(t)$; r is called the *comoving radial distance coordinate*.

The Comoving Distance Coordinate

Proper distances perpendicular to the line of sight must also change by a factor a between the epochs t and t_0 .

$$\frac{\Delta l(t)}{\Delta l(t_0)} = a(t) . \quad (29)$$

From the metric (25),

$$a(t) = \frac{R_c(t) \sin [\varrho/R_c(t)] d\theta}{R_c(t_0) \sin[r/R_c(t_0)] d\theta} . \quad (30)$$

Reorganising this equation and using (28),

$$\frac{R_c(t)}{a(t)} \sin \left[\frac{a(t)r}{R_c(t)} \right] = R_c(t_0) \sin \left[\frac{r}{R_c(t_0)} \right] . \quad (31)$$

This is only true if

$$R_c(t) = a(t) R_c(t_0) , \quad (32)$$

that is, the radius of curvature of the spatial sections is proportional to the scale factor $a(t)$. Thus, in order to preserve isotropy and homogeneity, *the curvature of space changes as the Universe expands as $\kappa = R_c^{-2} \propto a^{-2}$* . κ cannot change sign and so, if the geometry of the Universe was once, say, hyperbolic, it will always remain so.

The Robertson-Walker Metric

Let us call the value of $R_c(t_0)$, that is, the radius of curvature of the spatial geometry at the present epoch, \mathfrak{R} . Then

$$R_c(t) = a(t) \mathfrak{R} . \quad (33)$$

Substituting (28) and (33) into the metric (25), we obtain

$$ds^2 = dt^2 - \frac{a^2(t)}{c^2} [dr^2 + \mathfrak{R}^2 \sin^2(r/\mathfrak{R})(d\theta^2 + \sin^2 \theta d\phi^2)] . \quad (34)$$

This is the *Robertson–Walker metric* in the form we will use in much of our future analysis. Notice that it contains one unknown function $a(t)$, the scale factor, which describes the dynamics of the Universe and an unknown constant \mathfrak{R} which describes the spatial curvature of the Universe at the present epoch.

The metric can be written in different ways. For example, if we use a *comoving angular diameter distance* $r_1 = \mathfrak{R} \sin(r/\mathfrak{R})$, the metric becomes

$$ds^2 = dt^2 - \frac{a^2(t)}{c^2} \left[\frac{dr_1^2}{1 - \kappa r_1^2} + r_1^2 (d\theta^2 + \sin^2 \theta d\phi^2) \right] , \quad (35)$$

where $\kappa = 1/\mathfrak{R}^2$.

The Robertson-Walker Metric

By a suitable rescaling of the r_1 coordinate $\kappa r_1^2 = r_2^2$, the metric can equally well be written

$$ds^2 = dt^2 - \frac{R_1^2(t)}{c^2} \left[\frac{dr_2^2}{1 - \kappa r_2^2} + r_2^2 (d\theta^2 + \sin^2 \theta d\phi^2) \right], \quad (36)$$

with $k = +1, 0$ and -1 for universes with spherical, flat and hyperbolic geometries respectively. Notice that, in this rescaling, the value of $R_1(t) = R_c(t_0)a = \mathfrak{R}a$ and so the value of $R_1(t)$ at the present epoch is \mathfrak{R} rather than unity. This is a popular form for the metric, but I will normally use (34) because the r coordinate has an obvious and important physical meaning.

The importance of the metrics (34), (35) and (36) is that they enable us to define the invariant interval ds^2 between events at any epoch or location in the expanding Universe.

The Robertson-Walker Metric

To summarise, the *Robertson-Walker metric* can be written in the following form:

$$ds^2 = dt^2 - \frac{a^2(t)}{c^2} [dr^2 + \mathfrak{R}^2 \sin^2(r/\mathfrak{R}) (d\theta^2 + \sin^2 \theta d\phi^2)] .$$

The metric contains one unknown function $a(t)$, the **scale factor**, and the constant \mathfrak{R} which is the radius of curvature of the geometry of the Universe at the present epoch.

- t is cosmic time as measured by a clock carried by a fundamental observer;
- r is the *comoving radial distance coordinate* which is fixed to a galaxy for all time.
- $a(t) dr$ is the element of proper (or geodesic) distance in the radial direction at the epoch t ;
- $a(t) [\mathfrak{R} \sin(r/\mathfrak{R})] d\theta$ is the element of proper distance perpendicular to the radial direction subtended by the angle $d\theta$ at the origin;
- Similarly, $a(t) [\mathfrak{R} \sin(r/\mathfrak{R})] \sin \theta d\phi$ is the element of proper distance in the ϕ -direction.

The Cosmological Redshift

By cosmological redshift, we mean the shift of spectral lines to longer wavelengths associated with the isotropic expansion of the system of galaxies. If λ_e is the wavelength of the line as emitted and λ_0 the observed wavelength, the redshift z is defined to be

$$z = \frac{\lambda_0 - \lambda_e}{\lambda_e} . \quad (37)$$

If the redshift z were interpreted as the recession velocity v of a galaxy, these would be related by the Newtonian Doppler shift formula

$$v = cz . \quad (38)$$

This is the type of velocity which Hubble used in deriving the velocity–distance relation, $v = H_0 r$. It is however incorrect to use the special relativistic Doppler shift formula

$$1 + z = \left(\frac{1 + v/c}{1 - v/c} \right)^{1/2} , \quad (39)$$

at large redshifts. Rather, because of the requirements of isotropy and homogeneity, the relation $v \propto r$ applies at all distances, **including those at which the recession velocity would exceed the speed of light.**

The Real Meaning of Redshift

Consider a wave packet of frequency ν_1 emitted between cosmic times t_1 and $t_1 + \Delta t_1$ from a distant galaxy. This wave packet is received by an observer at the present epoch in the interval of cosmic time t_0 to $t_0 + \Delta t_0$. The signal propagates along null-cones, $ds^2 = 0$, and so, considering radial propagation from source to observer, $d\theta = 0$ and $d\phi = 0$, the metric (34) gives us the relation

$$dt = -\frac{a(t)}{c} dr \quad \frac{c dt}{a(t)} = -dr . \quad (40)$$

$a(t) dr$ is simply the interval of proper distance at cosmic time t . The minus sign appears because the origin of the r coordinate is the observer at $t = t_0$. Considering first the leading edge of the wave packet, the integral of (40) is

$$\int_{t_1}^{t_0} \frac{c dt}{a(t)} = - \int_r^0 dr . \quad (41)$$

The end of the wave packet must travel the same distance in units of comoving distance coordinate since the r coordinate is fixed to the galaxy for all time. Therefore,

$$\int_{t_1 + \Delta t_1}^{t_0 + \Delta t_0} \frac{c dt}{a(t)} = - \int_r^0 dr , \quad (42)$$

The Meaning of Redshift

Therefore,

$$\int_{t_1}^{t_0} \frac{c dt}{a(t)} + \frac{c \Delta t_0}{a(t_0)} - \frac{c \Delta t_1}{a(t_1)} = \int_{t_1}^{t_0} \frac{c dt}{a(t)}. \quad (43)$$

Since $a(t_0) = 1$, we find that

$$\Delta t_0 = \Delta t_1 / a(t_1). \quad (44)$$

This is the cosmological expression for the phenomenon of *time dilation*. Distant galaxies are observed at an earlier cosmic time t_1 when $a(t_1) < 1$ and so phenomena are observed to take longer in our frame of reference than in that of the source.

Expression (44) also provides an expression for the *redshift*. If $\Delta t_1 = \nu_1^{-1}$ is the period of the emitted waves and $\Delta t_0 = \nu_0^{-1}$ that of observed waves, $\nu_0 = \nu_1 a(t_1)$. In terms of redshift z ,

$$z = \frac{\lambda_0 - \lambda_e}{\lambda_e} = \frac{\lambda_0}{\lambda_e} - 1 = \frac{\nu_1}{\nu_0} - 1, \quad (45)$$

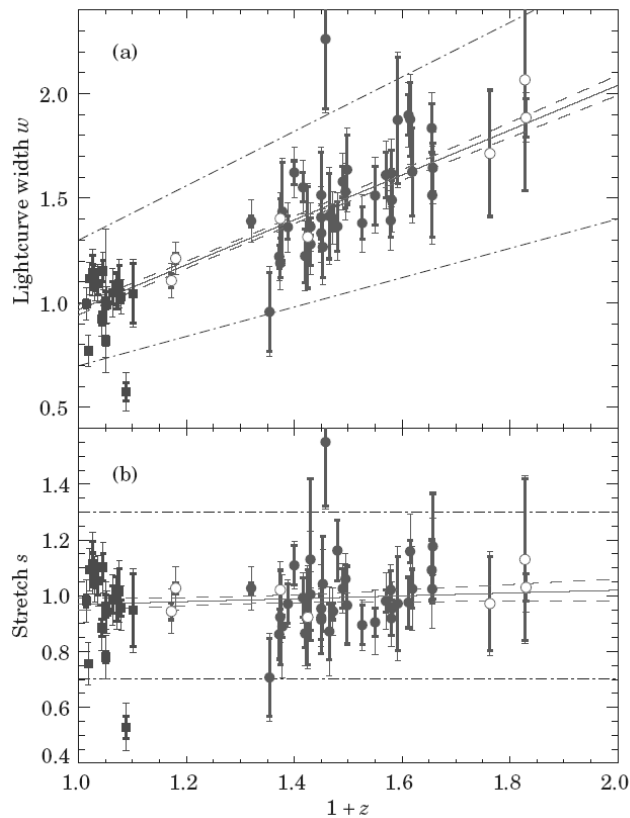
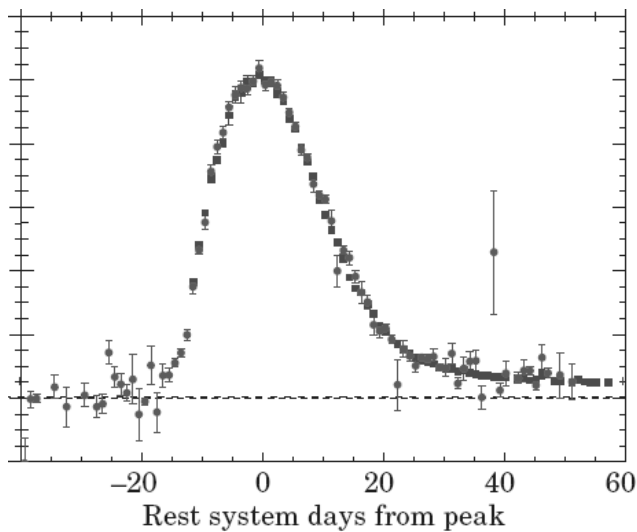
$$\boxed{a(t_1) = \frac{1}{1 + z}}. \quad (46)$$

The Time Dilation Test

A key test of the Robertson-Walker metric is that the same formula which describes the redshift of spectral lines should also apply to time intervals in the emitted and received reference frames.

This has been possible through the use of Type 1a supernova which have remarkably similar light curves (upper panel).

The lower panel shows the width of the light curves of Type 1a supernovae as a function of redshift. (a) A clear time dilation effect is observed exactly proportional to $(1 + z)$, as predicted by the Robertson-Walker metric. (b) The width of the light curves divided by $(1 + z)$.



Radiation Dominated Universes

For a gas of photons, massless particles or a relativistic gas in the ultrarelativistic limit $E \gg mc^2$, pressure p is related to energy density ε by

$$p = \frac{1}{3}\varepsilon$$

and the inertial mass density of the radiation ρ_r is related to its energy density ε by

$$\varepsilon = \rho_r c^2 .$$

If $N(\nu)$ is the number density of photons of energy $h\nu$, then the energy density of radiation is found by summing over all frequencies

$$\varepsilon = \sum_{\nu} h\nu N(\nu) . \quad (47)$$

If the number of photons is conserved, their number density varies as

$N = N_0 a^{-3} = N_0 (1+z)^3$ and the energy of each photon changes with redshift by the usual redshift factor $\nu = \nu_0 (1+z)$. Therefore, the variation of the energy density of radiation with cosmic epoch is

$$\varepsilon = \sum_{\nu_0} h\nu_0 N_0(\nu_0) (1+z)^4 ; \quad (48)$$

$$\varepsilon = \varepsilon_0 (1+z)^4 = \varepsilon_0 a^{-4} . \quad (49)$$

The Radiation Temperature Test

In the case of black-body radiation, the energy density of the radiation is given by the Stefan–Boltzmann law $\varepsilon = aT^4$ and its spectral energy density, that is, its energy density per unit frequency range, by the Planck distribution

$$\varepsilon(\nu) d\nu = \frac{8\pi h\nu^3}{c^3} \frac{1}{e^{h\nu/kT} - 1} d\nu . \quad (50)$$

It immediately follows that, for black-body radiation, the radiation temperature T_r varies with redshift as $T_r = T_0(1 + z)$ and the spectrum of the radiation changes as

$$\begin{aligned} \varepsilon(\nu_1) d\nu_1 &= \frac{8\pi h\nu_1^3}{c^3} [(e^{h\nu_1/kT_1} - 1)]^{-1} d\nu_1 \\ &= \frac{8\pi h\nu_0^3}{c^3} [e^{h\nu_0/kT_0} - 1]^{-1} (1 + z)^4 d\nu_0 \\ &= (1 + z)^4 \varepsilon(\nu_0) d\nu_0 . \end{aligned} \quad (51)$$

Radiation Dominated Universes

This provides another key test of the change of the time dilation formula since the temperature inferred from observations of fine structure lines in the spectra of distant quasars should increase as $(1 + z)$.

The fine-structure splittings of the ground state of neutral carbon atoms C I enable this test to be carried out. The photons of the background radiation excite the fine-structure levels of the ground state of the neutral carbon atoms and the relative strengths of the absorption lines originating from the ground and first excited states are determined by the energy density and temperature of the background radiation.

Author	quasar	redshift	predicted	observed
Songaila et al. (1994)	Q1331+170	$z_{\text{abs}} = 1.776$	7.58 K	7.4 ± 0.8
Ge et al. (1997)	QSO 0013-004	$z_{\text{abs}} = 1.9731$	8.105 K	7.9 ± 1.0 K
Ledoux et al. (2006)	PSS J1443+2724	$z_{\text{abs}} = 4.224$	14.2 K	consistency

Hubble's Law

In terms of proper distances, Hubble's Law can be written $v = H\varrho$ and so

$$\frac{d\varrho}{dt} = H\varrho . \quad (52)$$

We have written H rather than H_0 in Hubble's law since a 'Hubble's constant' H can be defined at any epoch as we show below. Substituting $\varrho = a(t)r$, we find that

$$r \frac{da(t)}{dt} = Ha(t)r , \quad (53)$$

that is,

$$H = \dot{a}/a . \quad (54)$$

Since we measure Hubble's constant H_0 at the present epoch, $t = t_0$, $a = 1$, we find

$$H_0 = (\dot{a})_{t_0} . \quad (55)$$

Thus, Hubble's constant H_0 defines the present expansion rate of the Universe. Notice that we can define a value of Hubble's constant at any epoch through the more general relation

$$H(t) = \dot{a}/a . \quad (56)$$

Angular Diameters

Next, we work out the angular size of an object of proper length d perpendicular to the radial coordinate at redshift z . The spatial component of the metric (34) is the term in $d\theta$. The proper length d of an object at redshift z , corresponding to scale factor $a(t)$, is given by the increment of proper length perpendicular to the radial direction in the metric (34), that is,

$$d = a(t) \mathfrak{R} \sin\left(\frac{r}{\mathfrak{R}}\right) \Delta\theta = a(t) D \Delta\theta = \frac{D \Delta\theta}{(1+z)}; \quad (57)$$

$$\Delta\theta = \frac{d(1+z)}{D}, \quad (58)$$

where we have introduced a *distance measure* $D = \mathfrak{R} \sin(r/\mathfrak{R})$. For small redshifts, $z \ll 1$, $r \ll \mathfrak{R}$, (58) reduces to the Euclidean relation $d = r \Delta\theta$.

The expression (58) can also be written in the form

$$\Delta\theta = \frac{d}{D_{\Delta}}, \quad (59)$$

so that the relation between d and $\Delta\theta$ looks like the standard Euclidean relation. To achieve this, we introduce another distance measure $D_{\Delta} = D/(1+z)$, the *angular diameter distance*.

Angular Diameters of objects expanding with the Universe

Another useful calculation is the angular diameter of an object which continues to partake in the expansion of the Universe. This is the case for infinitesimal perturbations in the expanding Universe. A good example is the angular diameter which large scale structures present in the Universe today would have subtended at an earlier epoch, say, the epoch of recombination, if they had simply expanded with the Universe. This calculation is used to work out physical sizes today corresponding to the angular scales of the fluctuations observed in the Cosmic Microwave Background Radiation. If the physical size of the object is $d(t_0)$ now and it expanded with the Universe, its physical size at redshift z was $d(t_0)a(t) = d(t_0)/(1+z)$. Therefore, the object subtended an angle

$$\Delta\theta = \frac{d(t_0)}{D}. \quad (60)$$

Notice that in this case the $(1+z)$ factor has disappeared from (58).

Apparent Intensities

Suppose a source at redshift z has luminosity $L(\nu_1)$ (measured in W Hz^{-1}), that is, the total energy emitted over 4π steradians per unit time per unit frequency interval. What is the flux density $S(\nu_0)$ of the source at the observing frequency ν_0 , that is, the energy received per unit time, per unit area and per unit bandwidth ($\text{W m}^{-2} \text{Hz}^{-1}$) where $\nu_0 = a(t_1)\nu_1 = \nu_1/(1+z)$?

Suppose the source emits $N(\nu_1)$ photons of energy $h\nu_1$ in the bandwidth ν_1 to $\nu_1 + \Delta\nu_1$ in the proper time interval Δt_1 . Then the luminosity $L(\nu_1)$ of the source is

$$L(\nu_1) = \frac{N(\nu_1) h\nu_1}{\Delta\nu_1 \Delta t_1}. \quad (61)$$

These photons are distributed over a 'sphere' centred on the source at epoch t_1 and, when the 'shell' of photons arrives at the observer at the epoch t_0 , a certain fraction of them is intercepted by the telescope. The photons are observed at the present epoch t_0 with frequency $\nu_0 = a(t_1)\nu_1$, in a proper time interval $\Delta t_0 = \Delta t_1/a(t_1)$ and in the waveband $\Delta\nu_0 = a(t_1)\Delta\nu_1$.

Apparent Intensities

We need to know how the photons spread out over a sphere between the epochs t_1 and t_0 , that is, we must relate the diameter of our telescope Δl to the angular diameter $\Delta\theta$ which it subtends at the source at epoch t_1 . The metric (34) provides the answer. The proper distance Δl refers to the present epoch at which $R(t) = 1$ and hence

$$\Delta l = D\Delta\theta, \quad (62)$$

where $\Delta\theta$ is the angle measured by a fundamental observer located at the source.

We can also understand this result by considering how the photons emitted by the source spread out over solid angle $d\Omega$, as observed from the source in the curved geometry. If the Universe were not expanding, the surface area over which the photons would be observed at a time t after their emission would be

$$dA = R_C^2 \sin^2 \frac{x}{R_C} d\Omega, \quad (63)$$

where $x = ct$. In the expanding Universe, R_C changes as the Universe expands and so, in place of the expression x/R_C , we write

$$\frac{1}{\mathfrak{R}} \int_{t_1}^{t_0} \frac{c dt}{a} = \frac{r}{\mathfrak{R}}, \quad (64)$$

where r is the comoving radial distance coordinate.

Apparent Intensities

Thus,

$$dA = \mathfrak{R}^2 \sin^2 \frac{r}{\mathfrak{R}} d\Omega . \quad (65)$$

Therefore, the diameter of the telescope as observed from the source is $\Delta l = D\Delta\theta$. Notice how the use of the comoving radial distance coordinate takes account of the changing geometry of the Universe in this calculation. Notice also the difference between (58) and (62). They correspond to angular diameters measured in opposite directions along the light cone. The factor of $(1+z)$ difference between them is part of a more general relation concerning angular diameter measures along light cones which is known as the *reciprocity theorem*.

Therefore, the surface area of the telescope is $\pi\Delta l^2/4$ and the solid angle subtended by this area at the source is $\Delta\Omega = \pi\Delta\theta^2/4$. The number of photons incident upon the telescope in time Δt_0 is therefore

$$N(\nu_1)\Delta\Omega/4\pi , \quad (66)$$

but they are now observed with frequency ν_0 .

Apparent Intensities

Therefore, the flux density of the source, that is, the energy received per unit time, per unit area and per unit bandwidth is

$$S(\nu_0) = \frac{N(\nu_1) h\nu_0 \Delta\Omega}{4\pi \Delta t_0 \Delta\nu_0 (\pi/4) \Delta l^2} . \quad (67)$$

We can now relate the quantities in (67) to the properties of the source, using (61), (62) and (65).

$$S(\nu_0) = \frac{L(\nu_1) a(t_1)}{4\pi D^2} = \frac{L(\nu_1)}{4\pi D^2 (1+z)} . \quad (68)$$

If the spectra of the sources are of power law form, $L(\nu) \propto \nu^{-\alpha}$, this relation becomes

$$S(\nu_0) = \frac{L(\nu_0)}{4\pi D^2 (1+z)^{1+\alpha}} . \quad (69)$$

Apparent Intensities

We can repeat the analysis for *bolometric* luminosities and flux densities. In this case, we consider the total energy emitted in a finite bandwidth $\Delta\nu_1$ which is received in the bandwidth $\Delta\nu_0$, that is

$$\begin{aligned} L_{\text{bol}} &= L(\nu_1)\Delta\nu_1 = 4\pi D^2 S(\nu_0)(1+z) \times \Delta\nu_0(1+z) \\ &= 4\pi D^2(1+z)^2 S_{\text{bol}}, \end{aligned} \quad (70)$$

where the bolometric flux density is $S_{\text{bol}} = S(\nu_0)\Delta\nu_0$. Therefore,

$$S_{\text{bol}} = \frac{L_{\text{bol}}}{4\pi D^2(1+z)^2} = \frac{L_{\text{bol}}}{4\pi D_L^2}. \quad (71)$$

The quantity $D_L = D(1+z)$ is called the *luminosity distance* of the source since this definition makes the relation between S_{bol} and L_{bol} look like an inverse square law. The bolometric luminosity can be integrated over any suitable bandwidth so long as the corresponding redshifted bandwidth is used to measure the bolometric flux density at the present epoch.

$$\sum_{\nu_0} S(\nu_0)\Delta\nu_0 = \frac{\sum_{\nu_1} L(\nu_1)\Delta\nu_1}{4\pi D^2(1+z)^2} = \frac{\sum_{\nu_1} L(\nu_1)\Delta\nu_1}{4\pi D_L^2}. \quad (72)$$

K -corrections

The formula (68) is the best expression for relating the observed intensity $S(\nu_0)$ to the intrinsic luminosity of the source $L(\nu_1)$. We can also write (68) in terms of the luminosity of the source at the observing frequency ν_0 as

$$S(\nu_0) = \frac{L(\nu_0)}{4\pi D_L^2} \left[\frac{L(\nu_1)}{L(\nu_0)} (1+z) \right]. \quad (73)$$

but this now requires knowledge of the spectrum of the source $L(\nu)$. The last term in square brackets is a form of what is known as the *K-correction*. K-corrections were introduced by the pioneer optical cosmologists in the 1930s in order to ‘correct’ the apparent magnitude of distant galaxies for the effects of redshifting their spectra when observations are made through standard filters with a fixed mean observing frequency ν_0 . Taking logarithms and multiplying by -2.5 , we can rewrite (68) in terms of absolute (M) and apparent (m) magnitudes through the relations

$M = \text{constant} - 2.5 \log_{10} L(\nu_0)$ and $m = \text{constant} - 2.5 \log_{10} S(\nu_0)$. We find

$$M = m - 5 \log_{10}(D_L) - K(z) - 2.5 \log_{10}(4\pi), \quad (74)$$

where

$$K(z) = -2.5 \log_{10} \left[\frac{L(\nu_1)}{L(\nu_0)} (1+z) \right]. \quad (75)$$

K-corrections

This form of K-correction is correct for *monochromatic* flux densities and luminosities. In the case of observations in the optical waveband, apparent magnitudes are measured through standard filters which usually have quite wide pass-bands. Therefore, to determine the appropriate K-corrections, the spectral energy distribution of the galaxy has to be convolved with the transmission function of the filter in the rest-frame and at the redshift of the galaxy. This is a straightforward calculation once the spectrum of the object is known.

Although I prefer to work directly with (68) and take appropriate averages, K-corrections are rather firmly established in the literature and it is often convenient to use the term to describe the effects of shifting the emitted spectrum into the observing wavelength window.

Number Densities

We often need to know the number of objects in a particular redshift interval, z to $z + dz$. Since there is a one-to-one relation between r and z , the problem is straightforward because, by definition, r is a radial proper distance coordinate defined *at the present epoch*. Therefore, the number of objects in the interval of comoving radial distance coordinate distance r to $r + dr$ is given by results already obtained. The space-time diagram illustrates how we can evaluate the numbers of objects in the comoving distance interval r to $r + dr$ entirely by working in terms of *comoving volumes* at the present epoch. At the present epoch, the radius of curvature of the spatial geometry is \mathfrak{R} and so the volume of a spherical shell of thickness dr at comoving distance coordinate r is

$$dV = 4\pi\mathfrak{R}^2 \sin^2(r/\mathfrak{R}) dr = 4\pi D^2 dr . \quad (76)$$

Number Densities

Therefore, if N_0 is the present space density of objects and their number is conserved as the Universe expands,

$$dN = N(z) dz = 4\pi N_0 D^2 dr . \quad (77)$$

The definition of comoving coordinates automatically takes care of the expansion of the Universe. Another way of expressing this result is to state that (77) gives the number density of objects in the redshift interval z to $z + dz$, assuming the *comoving number density* of the objects is unchanged with cosmic epoch. If, for some reason, the comoving number density of objects changes with cosmic epoch as, say, $f(z)$ with $f(z = 0) = 1$, then the number of objects expected in the redshift interval dz is

$$dN = N(z) dz = 4\pi N_0 f(z) D^2 dr . \quad (78)$$

Age of the Universe

Finally, let us work out an expression for the age of the Universe, T_0 from a rearranged version of (40). The basic differential relation is

$$\frac{c dt}{a(t)} = -dr , \quad (79)$$

and hence

$$T_0 = \int_0^{t_0} dt = \int_0^{r_{\max}} \frac{a(t) dr}{c} , \quad (80)$$

where r_{\max} is the comoving distance coordinate corresponding to $a = 0, z = \infty$.

Summary

1. First work out from theory, or otherwise, the function $a(t)$ and the curvature of space at the present epoch $\kappa = \mathfrak{R}^{-2}$. Once we know $a(t)$, we know the redshift–cosmic time relation.
2. Now work out the *comoving radial distance coordinate* r from the integral

$$r = \int_{t_1}^{t_0} \frac{c dt}{a(t)}. \quad (81)$$

Recall what this expression means – the proper distance interval $c dt$ at epoch t is projected forward to the present epoch t_0 by the scale factor $a(t)$. This integration yields an expression for r as a function of redshift z .

3. Next, work out the *distance measure* D from

$$D = \mathfrak{R} \sin \frac{r}{\mathfrak{R}}. \quad (82)$$

This relation determines D as a function of redshift z .

4. If so desired, the *angular diameter distance* $D_A = D/(1 + z)$ and the *luminosity distance* $D_L = D(1 + z)$ can be introduced to relate physical sizes and luminosities to angular diameters and flux densities respectively.
5. The number of objects dN in the redshift interval dz and solid angle Ω can be found from the expression

$$dN = \Omega N_0 D^2 dr , \quad (83)$$

where N_0 is the number density of objects at the present epoch which are assumed to be conserved as the Universe expands.