Chapter 10
General Relativity and Cosmology

The general theory of relativity is considered to be Albert Einstein's masterpiece in theoretical physics. In contrast with special relativity, where scientists like Hendrik Lorentz and Henri Poincaré worked in parallel, motivated by the unsolved physical problems existing at the beginning of the twentieth century (for instance, motion with respect to the æther and the negative result of the Michelson–Morley experiment), there was no such motivation for general relativity. With the exception of an anomaly in the precession of Mercury's orbit, the Newtonian theory of gravitation did not manifest symptoms of obsolescence.

The general theory of relativity was constructed by Einstein in a purely deductive form, using as basic postulates the principles of covariance and equivalence. A suitable mathematical tool had just been invented, thanks to the works of the Italian mathematicians Gregorio Ricci-Curbastro (1853–1925) and Tullio Levi-Civita (1873–1941), who had developed the so-called absolute differential calculus. Einstein was introduced to the formal aspects of non-Euclidean geometry by his friend, the mathematician Marcel Grossmann (1878–1936).

In the summer of 1915, Einstein was invited by David Hilbert (1862–1943), an outstanding mathematician, to visit Göttingen in order to lecture on his work on the theory of gravitation. In November 1915, independently, Einstein and Hilbert presented the equations of the gravitational field, which Hilbert had derived by variational principle. Therefore, the gravitational field action is customarily called Einstein–Hilbert action. However, the scheme of general relativity was developed by Einstein, therefore the new theory of gravity is Einstein's general relativity.

The final version of the theory was published by Einstein in 1916. The most spectacular confirmation was obtained in 1919, when Arthur Eddington (1882–1944) together with a team observed the bending of light from a distant star as it passed close by the Sun during a solar eclipse. This and other predictions of general relativity were subsequently confirmed in several experiments, making it an essential tool in cosmological research.

M. Chaichian et al., *Basic Concepts in Physics*, Undergraduate Lecture Notes in Physics,
https://doi.org/10.1007/978-3-662-62313-8_10

10.1 Principle of Equivalence and General Relativity

It is customary to distinguish between two forms of the principle of equivalence, referred to as weak and strong. The weak principle of equivalence establishes the equality of the inertial and gravitational masses. The inertial mass m_i of a body is the coefficient of the acceleration **a** in Newton's second law:

$$\mathbf{F} = m_i \mathbf{a}. \tag{10.1}$$

The gravitational mass of the same body, for example, in its interaction with the Earth, is the one which appears in the expression for the force of gravitational attraction, i.e.,

$$\mathbf{F} = -\frac{GMm_g}{r^2}\mathbf{r}_0, \tag{10.2}$$

between, say, the Earth, of mass M, and the body of interest, of mass m_g. The unit vector $\mathbf{r}_0$ is along the line joining the body with the Earth's centre. We have the equivalence between these two masses expressed by means of the equality $m_i = m_g$. As a consequence, the acceleration due to gravity is the same for all bodies, if air resistance is neglected.

Imagine an elevator falling freely. An observer inside it would feel weightless. If the observer has a ball and lets go of it, without pushing it in any way, it will hang in the air, falling together with the system. When falling freely under the action of gravity, everything happens for the observer as if gravity were zero inside the elevator. For an observer inside an artificial satellite, this produces the effect of feeling weightless.

Returning to the elevator, if we accelerated it, for instance, by doubling the acceleration produced by the Earth attraction, our observer would feel weight in the opposite direction, that is, he would feel attracted toward the ceiling, as though there were a gravitational field in that direction. We see in this way that an accelerated system and a gravitational field produce similar effects, or in other words, motion in accelerated systems is equivalent to motion produced by a gravitational field (Fig. 10.1).

If the elevator were to ascend with some acceleration g', however, the observer of mass m would experience an increase in weight by an amount mg'. That is, it would seem as though the Earth's gravitational field had increased, and the observer's weight would now be $m(g + g')$, instead of mg. In conclusion, a local equivalence exists (that is, in a small region of space) between an accelerated reference frame and a gravitational field.

The *strong principle of equivalence* establishes that in every gravitational field, an elevator falling freely turns *locally* into a system in which the laws of physics are the same as in special relativity, that is, in an inertial system. The case is the same for an artificial satellite, in which the weightlessness effect is produced as a consequence of the satellite falling continuously toward the Earth as it moves around its orbit (as we pointed out in Chap. 1, the closed orbits result from the combination of this free-fall effect with a large enough tangential velocity).

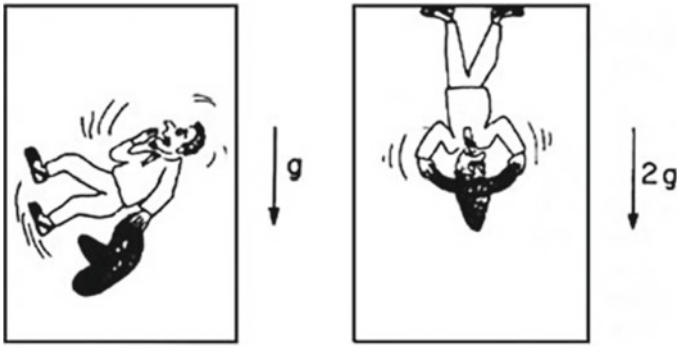
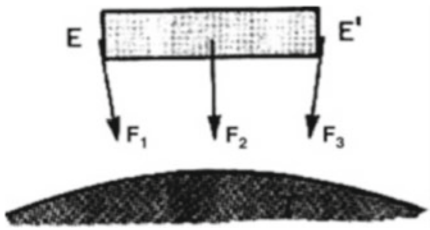


Fig. 10.1 For an observer inside an elevator falling freely under the action of gravity, the gravitational field force acting on him is canceled, and he feels as though he is floating or weightless. If the elevator is accelerated with twice the acceleration due to gravity, the observer inside feels a force equal to the Earth's gravity acting on him, but directed toward the ceiling of the elevator.

Fig. 10.2 For a very large elevator falling freely, the forces $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3$ are not parallel, whence the elevator tends to adopt the form of an arch.



So both the falling elevator and the satellite can be treated as inertial systems, if their dimensions are small (strictly, pointlike). In the same way the Earth could be considered as an inertial system with respect to the Sun if its dimensions were negligibly small. In the case of the Earth, the fact of not being pointlike causes the tidal forces due to the solar attraction (there are also tidal forces due to the Moon). The atmospheric and oceanic masses are more sensitive to the tidal forces.

It is easy to understand the origin of tidal forces if we consider an extremely large elevator (Fig. 10.2). Its centre of mass M moves with the acceleration due to gravity, and falls freely. But the forces exerted on the ends E and E' are not parallel to the one which acts on M, since, due to the curvature of the Earth's surface, $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$ are directed toward the Earth's centre, whence the elevator tends to adopt the form of an arch.

Similarly, the trajectory followed by the Earth in its motion around the Sun (without considering the effect of the Moon) corresponds to a pointlike mass located at the Earth's centre of mass. The centre of mass behaves like a freely falling body during its motion. But because of the Earth's extension, the points distant from the centre of mass do not rigorously follow the *free-falling* motion. The result is that a small residual force is exerted on them by the Sun, producing tides. It must be emphasized, however, that the most notable tides are produced by the Moon, and have a similar origin. For artificial satellites, this tidal effect is very small, and it can be neglected in the first approximation. One can thus consider that the satellite satisfies the condition of the principle of equivalence: for observers inside it, there the Earth's gravitational field vanishes.

10.2 Gravitational Field and Geometry

The potential of the gravitational field near the surface of the Earth is

$$V(r) = -\frac{GM}{r}, \tag{10.3}$$

where r is the Earth's radius, G is the constant of gravitation, and M is the Earth's mass.

Imagine now the following *Gedanken experiment*: suppose that at some height l with respect to some reference system on the Earth's surface we have an electron and a positron at rest. The mass of each is m. The potential energy of the two particles at that height, putting $l \ll r$, is

$$E = -\frac{2mGM}{r+l} = -\frac{2mGM}{r}\frac{1}{1+\frac{l}{r}} \approx 2mV\left(1-\frac{l}{r}\right) = 2m(V+\Delta V), \tag{10.4}$$

where $\Delta V = GMl/r^2$. If the two particles now fall to the Earth's surface, their potential energy decreases to $2mV$, and their kinetic energy will be equal to $2m\Delta V$. If now the electron and positron annihilate to produce two photons of angular frequency ω, the following equation will be satisfied:

$$2\hbar\omega = 2mc^2 + 2m\Delta V. \tag{10.5}$$

That is, the energy of the two photons will be equal to the sum of the rest energy of the electron and positron, plus their kinetic energy. We assume that the velocity of these particles is not very large, so that we can use the approximation

$$\frac{mc^2}{\sqrt{1-v^2/c^2}} \approx mc^2 + \frac{1}{2}mv^2,$$

where $\frac{1}{2}mv^2 = m\Delta V$. Now, by means of a suitable mirror, let the two photons be reflected back up to the initial level of height l. At this height l, let the two photons create the electron–positron pair again. The pair will be at rest, since otherwise there would be a gain of energy in the cyclic process, implying the possibility of constructing a perpetual motor of the second kind.

The frequency ω' of the two photons at the height l is different from the frequency ω at the level of the Earth's surface, and should satisfy

$$2\hbar\omega' = 2mc^2. \tag{10.6}$$

Comparing (10.5) and (10.6), we deduce that

$$\frac{\omega - \omega'}{\omega'} = \frac{\Delta V}{c^2}. \tag{10.7}$$

So the frequency of the radiation varies in a gravitational field. Since ΔV is positive in our case, (10.7) implies that radiation emitted away from the surface of the Earth has frequency diminished by an amount

$$\Delta\omega = \omega - \omega' = \frac{\Delta V}{c^2}\omega, \tag{10.8}$$

where in writing the second equality we have assumed that ω and ω' are much larger than their difference.

Assume that a source on the Earth emits radiation at some frequency. The observer at some height will measure a lower frequency, i.e., shifted toward the red. This effect was measured for the first time by the American physicists Robert Pound and Glen Rebka in 1960, using a source of γ rays and the Mössbauer effect. These and other experiments reached an accuracy of 7×10^{-5}. In 2010, a much more exact measurement of the gravitational red shift based on quantum interference of matter waves within an accuracy of 7×10^{-9} was reported by H. Müller, A. Peters, and S. Chu.

Let us now examine the phenomenon from the wave point of view. If the frequency varies in a gravitational field, this should be caused by a time dilation. Actually, if a train of waves is sent from the Earth's surface, containing n complete oscillations during the time T_1, the relation between the angular frequency and the interval T_1 is

$$T_1 = 2\pi n/\omega. \tag{10.9}$$

The angular frequency of the same train of waves at the height l can be measured by dividing n by the duration of the train. The number obtained, ω', is different from ω, and this means that the interval T_2 that corresponds to n oscillations is

$$T_2 = 2\pi n/\omega'. \tag{10.10}$$

From (10.8) to (10.10), it follows that

$$\frac{T_2 - T_1}{T_1} = \frac{\Delta V}{c^2}. \tag{10.11}$$

That is, a clock at a height l measures for the duration of the wave train an interval of time longer than a clock located at the Earth's surface, and $T_2 = (1 + \Delta V/c^2)T_1$.

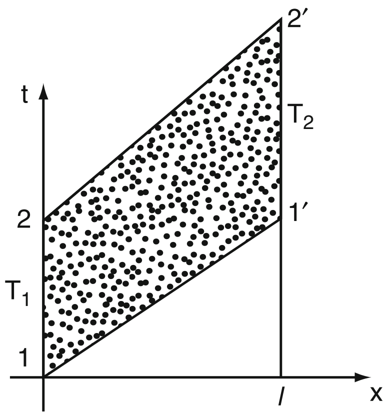


Fig. 10.3 Spacetime diagram of the propagation of a wave train in the gravitational field of the Earth from its surface to some height l. The duration of the wave train is different for observers located at the two points.

A clock on the Earth goes more slowly than another one placed at some height above the Earth's surface. In general, a clock located in a gravitational field goes more slowly than another clock located where the field is zero.

Let us draw a picture in which we mark on the horizontal axis x the height above the Earth and on the vertical axis the time t (Fig. 10.3). The event marked as 1 corresponds to the origin of the wave train as measured by the observer on the Earth. The event $1'$ corresponds to the origin of the train as measured by a second observer located at the height l. Similarly, the point 2 marks the end of the wave train as measured by the terrestrial observer, and $2'$ the same event as measured by the second observer. The lines $11'$ and $22'$ are the graphs of the propagation of the origin and the end of the wave train in spacetime. But as we have seen, the duration of the train, considered as the segments $12 = T_1$, $1'2' = T_2$, are different for the two observers:

$$T_2 > T_1. \tag{10.12}$$

On the other hand, the lines $11'$ and $22'$ should be parallel, since they correspond to the same phenomenon (the propagation of the signal) in a static gravitational field (it does not vary in time), and they differ only in that they have been measured by two different observers.

But the figure $11'2'2$ is not a parallelogram. The only solution to this paradox is that, in the presence of a gravitational field, the spacetime is curved. Hence, instead of taking the axes x, t on a plane, they must be taken on a curved surface. Then, by redefining the condition of parallelism on the surface, the lines $11'$ and $22'$ can be made to satisfy it on this surface.

A fundamental consequence of the general theory of relativity is that the effect of a gravitational field is described by spacetime curvature. Let us compare a plane and a curved surface like the surface of a sphere. Mark two points in the plane. Geometry demonstrates that the geodesic or shortest distance between those two points is the straight line segment joining them. In the geometry of the plane, the geodesics are straight lines extending across the whole plane toward infinity. Three points that are not aligned determine a triangle, the sum of whose internal angles is 180°. In other words, we can say that the plane is a two-dimensional Euclidean space.

Considering the same problem on the surface of the sphere leads to the conclusion that the geodesics are arcs of great circles (a great circle on the spherical surface is one whose centre coincides with the centre of the sphere). On the sphere, geodesics are finite in extent, and so is the total area of the sphere. Furthermore, a triangle on the spherical surface has the property that the sum of its internal angles is greater than 180°. The spherical surface is an example of a two-dimensional non-Euclidean space.

If α, β, and γ are the internal angles of a spherical triangle, A the area of this triangle, and R the radius of the sphere, we have the relation

$$\frac{A}{R^2} = \alpha + \beta + \gamma - \pi. \tag{10.13}$$

If A is kept constant and R tends to infinity, (10.13) gives the planar limit

$$\alpha + \beta + \gamma = \pi. \tag{10.14}$$

On the other hand, from (10.13), one can define the reciprocal of the square of the radius of the sphere, $K = 1/R^2$, by

$$K = \frac{\alpha + \beta + \gamma - \pi}{A}. \tag{10.15}$$

If the area A tends to zero in the expression (10.15), the resulting expression allows us to define the *curvature* in the neighbourhood of any point on the surface as

$$K = \lim_{A \to 0} \frac{\alpha + \beta + \gamma - \pi}{A}, \tag{10.16}$$

i.e., the excess over π of the sum of the internal angles of a triangle divided by the area of such triangle, in the limit of the area going to zero.

At a given point the curvature can be positive, zero, or negative. For example, K is positive everywhere in the case of a sphere, zero in the case of a plane, and negative on a saddle-shaped surface (Fig. 10.4).

Our intuition suggests that the three-dimensional physical space has the geometric properties resulting from generalizing the plane by adding one more dimension to obtain a three-dimensional Euclidean space. In this case, if we start from a point and move along a geodesic, that is, in a straight line, we move away from our starting point toward infinity.

In contrast, if our three-dimensional physical space had the geometrical properties which result from the generalization of the spherical surface to three dimensions, the geodesics would be closed curves. In contrast, if the geometry of space were of saddle

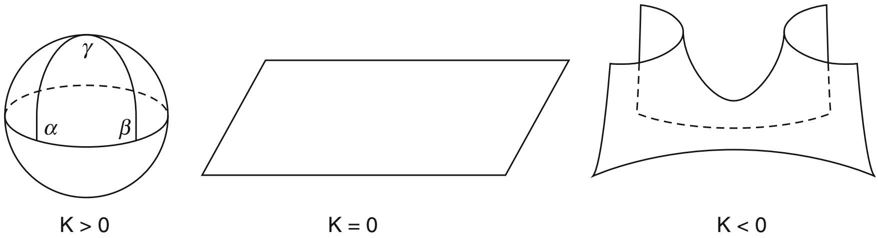


Fig. 10.4 A sphere has positive curvature, the plane has zero curvature, and a saddle-shaped surface has negative curvature.

type (negative curvature), the geodesics would not be closed, but open curves, and they would extend toward infinity.

According to general relativity, the planets, in their orbital motion around the Sun, describe geodesics in a four-dimensional curved spacetime, which is deformed by the mass of the Sun. In addition, when the rays of light emitted by a distant star pass close to the Sun, they follow a geodesic curve and hence deviate from the straight line trajectory. By defining $b = cL/E$, where L is the angular momentum of the beam and E its energy, the shifted angle is given approximately by

$$\delta\phi = 4GM_{\odot}/bc^2,$$

where $M_{\odot}$ is the mass of the Sun. Notice that b has dimension of length. The effect is $1.75''$ for light coming from distant stars and grazing the Sun's limb.

Actually, according to classical Newtonian mechanics and special relativity, some deviation of the light rays would be expected near the large solar mass, and it is not difficult to calculate this effect, which has been mentioned also in Sect. 1.5.3. But general relativity predicts a result twice as large, and this was confirmed by the observations made later by Eddington and other observers. The doubling of the deviation can be explained only in the framework of the general relativity, as a consequence of the curvature of space. It is found from the solution of the equation of motion for a light ray (the so-called eikonal equation) in a centrally symmetric gravitational field.

The geodesic curves described by the planets according to general relativity are not ellipses (as predicted by Newtonian mechanics), but more complicated curves in the form of almost-ellipses whose major axes precess around their focus (Fig. 10.5). An anomalous effect of this sort had been known since the nineteenth century in the orbit of Mercury. The observed precession of Mercury's perihelion (the point of closest approach to the Sun on the orbit) is 574" (arc-seconds) per century. The gravitational tugs of other planets, calculated by Newton's theory, could explain a precession of about 531". The origin of such a difference was not known. The calculations performed by Einstein in 1915 in the framework of general relativity provided the extra amount of + 43", in perfect agreement with the observed data. This was the first observational fact explained by the theory of general relativity. The

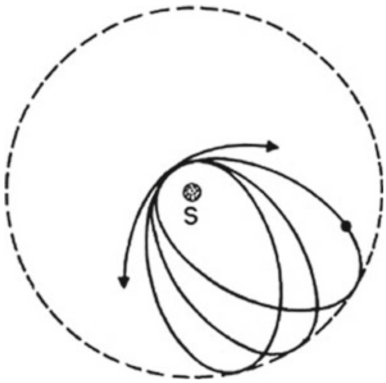


Fig. 10.5 The true trajectories of the planets around the Sun are precessing ellipses, resulting in curves in the shape of rosettes. This precession effect is very small, and is more perceptible in the case of Mercury owing to its proximity to the Sun.

effect is more perceptible in Mercury's orbit because of its high eccentricity and its proximity to the Sun, but other planets display it in smaller amounts. In particular, for the Earth, this effect is about 3.84" per century.

It is interesting to note that, according to general relativity, a body moving with some velocity $\mathbf{V}$ in a gravitational field is under the action of two forces: one corresponding to the usual gravitational attraction of Newtonian mechanics, and another one perpendicular to its velocity. This has a close analogy with the electromagnetic case, in which a charged particle in motion suffers the action of the Lorentz force, with two components: the electric force, independent of the velocity of the particle, and the magnetic force, perpendicular to its velocity. The additional force exerted by the gravitational field on a particle in motion in that field is the analog of the magnetic force. This second gravitational force is not very significant for low velocities since, as in the magnetic case, the term describing it contains the factor V/c.

According to the principle of equivalence, this second force of gravity corresponds more properly to the Coriolis force, appearing in a rotating (non-inertial) system of reference as a force perpendicular to the velocity of a particle moving in such a system.

General relativity also predicts that massive rotating bodies "drag" spacetime in their vicinity. This effect was first derived from general relativity by Josef Lense (1890–1985) and Hans Thirring (1888–1976) in 1918, and is also known as the Lense–Thirring effect.

Lensing effect. The lensing effect is due to the deflection of light coming from a distant object by a massive body. For small angles, it can be expressed as $\theta = 4GM_\odot/bc^2$, where $b = cL/E$ (see Sect. 10.2). Since the light is made up of photons, for which $p = E/c$, we have $L = Er/c$, which implies $b = r$, where r is the shortest distance from the photon beam to the body's centre. Thus, one can write

$$\theta = 2r_g/r. \tag{10.17}$$

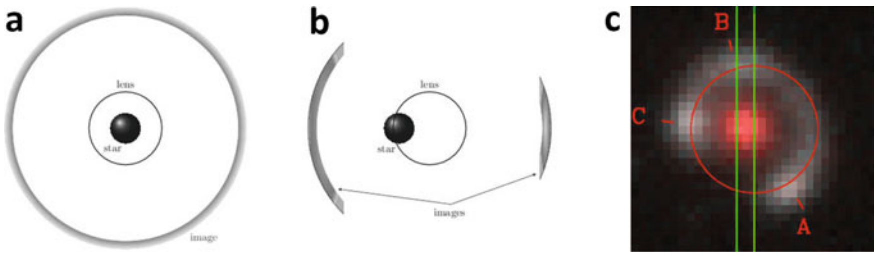


Fig. 10.6 **a** For perfect alignment of the observer, lens, and star we get a complete Einstein ring. **b** If this is not the case, only part of a ring will be observed. **c** An interesting example of the latter case is the Canarias ring, discovered by Margherita Bettinelli et al. in the constellation of the Sculptor in 2016. The maximal intensities are indicated by A, B, and C. The lens is a massive galaxy with redshift $z = 0.581$, and the source is also a galaxy, with $z = 1.165$. The ring covers around 300°.

Consider a massive object, which may be a star or a galaxy, and call it the lens. This is located between the observer and a still more distant star (galaxy). When it passes near the lens, light coming from this last object can be bent round toward the eye of the observer. This gravitational lensing phenomenon was first mentioned in 1924 by the physicist Orest Chwolson (1852–1934) in Saint Petersburg, and treated quantitatively by Albert Einstein in 1936. If the object, the lens, and the observer are perfectly aligned, the image of the body will be a circular ring, known as an Einstein ring, centred on the lens (Figs. 10.6).

GPS time correction due to general relativistic effects. GPS (Global Positioning System) satellites form a global navigation system. Each carries a very accurate atomic clock that provides geolocation and time information to a GPS receiver. GPS satellites are located at a height of approximately 26600 km from the centre of the Earth, and describe two full orbits every sidereal day. For a position accuracy of $\Delta x = 15$ m, the time aboard GPS satellites must be known to an accuracy of $\Delta t = \Delta x/c = 50 \times 10^{-9}$ s. The time measured by the satellite clocks must therefore be corrected due to effects from special and general relativity.

To calculate the relativistic effects on the time measured by a clock aboard a GPS satellite, which runs faster than a clock on Earth surface, we must compare the proper times measured on the satellite and on the Earth's surface. We take the Earth's mass to be 5.924×10^{27} g and its radius to be $R = 6378$ km. We need the velocities of the satellite v and the Earth's surface V, as well as r_g, to write the proper time at the satellite in the form $d\tau_S = ds/c$, where ds is obtained from the Schwarzschild metric (10.18). We should set r constant as well as $\theta = \pi/2$. Then, as $v = rd\phi/dt$, we have $-r^2 d\phi^2/c^2 = -(v^2/c^2)dt^2$ and

$$d\tau_S = \sqrt{1 - \frac{r_g}{r} - \frac{v^2}{c^2}}dt, \tag{10.18}$$

where the denominator of the third term in (10.28) has been approximated by one, i.e., $dr^2/(1 - \frac{r_g}{r}) \approx dr^2$, since it would only contribute to a small second order term.

(Note that, if r_g is neglected in (10.18), one is left with the special relativistic proper time relation in terms of v/c.) For the proper time at the Earth's surface, we have

$$d\tau_E = \sqrt{1 - \frac{r_g}{R} - \frac{V^2}{c^2}}dt. \tag{10.19}$$

Setting $\epsilon = r_g/r + v^2/c^2$ and $\eta = r_g/R + V^2/c^2$, we obtain an expression of the form

$$\frac{d\tau_S}{d\tau_E} = \frac{\sqrt{1-\epsilon}}{\sqrt{1-\eta}},$$

for the ratio of the proper times on the satellite and on Earth, where ϵ and η are very small quantities compared to 1.

Approximating by $\frac{d\tau_S}{d\tau_E} = (1 - \epsilon/2)(1 + \eta/2)$, and neglecting square terms, we get $d\tau_S - d\tau_E = (-\epsilon/2 + \eta/2)d\tau_E$. From this, we can integrate $d\tau$ over the interval of time to be corrected in one day. It is calculated as a problem below, leading to a value of 38×10^{-6} s.

10.3 Affine Connection and Metric Tensor

We saw in Chap. 5 how the concept of interval is used to characterize the distance between two events in spacetime. This concept remains valid in general relativity, and in fact the whole mathematical formulation of this theory starts from the expression for the infinitesimal interval between two events. In special relativity, if two events A and B have the spacetime coordinates $A = (x, y, z, ct)$ and $B = (x + dx, y + dy, z + dz, ct + cdt)$, the interval would have the form

$$ds_{AB}^2 = c^2dt^2 - dx^2 - dy^2 - dz^2. \tag{10.20}$$

Observe that the coefficients of the squares of the differentials of the coordinates are the constant numbers $(1, -1, -1, -1)$. It is customary to refer to (10.20) as the expression for the interval in the flat spacetime, and to call the set of four numbers $(1, -1, -1, -1)$ the Minkowski metric. With the notation introduced in Sect. 5.8, in the case of general relativity, the interval between events A and B would have the general form

$$ds_{AB}^2 = g_{00}dx_0^2 + g_{11}dx_1^2 + g_{22}dx_2^2 + g_{33}dx_3^2 + 2g_{12}dx_1dx_2 + 2g_{23}dx_2dx_3 + \cdots, \tag{10.21}$$

with ten general, spacetime-dependent coefficients $g_{\mu\nu} = g_{\mu\nu}(x)$, where $\mu, \nu = 0, 1, 2, 3$, and by x we denote the spacetime four-vector x^μ. In weak gravitational fields, $g_{\mu\nu}$ approach their special relativity values, i.e., the Minkowski metric. The quantities $g_{\mu\nu} = g_{\mu\nu}(x)$ form a mathematical entity, the *metric tensor of spacetime*. Recall that a tensor is an object which transforms as the product of vectors. The metric tensor is symmetric, i.e., $g_{\mu\nu} = g_{\nu\mu}$.

As a consequence of the curvature of spacetime, systems of curvilinear coordinates are more convenient. Recall also the *contravariant* quantities, transforming like the coordinate differentials $dx^\mu = (dx^0, dx^1, dx^2, dx^3)$, and covariant quantities, transforming like the partial derivatives $\frac{\partial}{\partial x^\mu} = \left(\frac{\partial}{\partial x^0}, \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3}\right)$, where by x^μ we denote the generalized coordinate. As examples of curvilinear coordinates, we have cylindrical coordinates $x^\mu = ct, \rho, \varphi, z$, and spherical coordinates $x^\mu = ct, r, \theta, \varphi$. A typical case of a covariant vector is the vector formed by the derivative of a scalar function f with respect to the (contravariant) coordinates:

$$\frac{\partial f(x)}{\partial x^\mu}.$$

Another example of a covariant quantity is the metric tensor $g_{\mu\nu}$. Given a contravariant vector, (A^0, A^1, A^2, A^3), we can transform it to a covariant one by multiplying it by the matrix formed by the metric tensor. We write $A_\mu = \sum_\nu g_{\mu\nu} A^\nu$, but from now on we drop the summation symbol, understanding that when repeated indices appear, like ν in the previous expression, we sum over them. This is Einstein's summation convention, introduced by Albert Einstein in his general relativity paper of 1916. We define δ^ν_μ to be the unit four-dimensional tensor, or Kronecker symbol, with all components equal to zero but with units down the main diagonal. Then the contravariant metric tensor $g^{\mu\lambda}$ satisfies the property

$$g^{\mu\lambda} g_{\lambda\nu} = \delta^\mu_\nu.$$

The task of defining the derivative of a vector with respect to the coordinates is more complicated. We must bear in mind that the variation of each of the components of a vector depends also on the other components. That is, this derivative which we will represent by ∇_λ, and is called the covariant derivative, or affine connection (an affine transformation has the general form $y = ax + b$), has two terms:

$$\nabla_\lambda A^\mu = \frac{\partial A^\mu}{\partial x^\lambda} + \Gamma^\mu_{\eta\lambda} A^\eta, \tag{10.22}$$

where $\Gamma^\mu_{\eta\lambda} = g^{\mu\xi}\Gamma_{\xi\eta\lambda}$. Note that $\Gamma^\mu_{\eta\lambda}$ and $\Gamma_{\xi\eta\lambda}$ are not tensors. They are called Christoffel symbols, and are defined in terms of $g_{\mu\nu}$ by the relation

$$\Gamma_{\xi\eta\lambda} = \frac{1}{2}\left(\frac{\partial g_{\xi\eta}}{\partial x^\lambda} + \frac{\partial g_{\xi\lambda}}{\partial x^\eta} - \frac{\partial g_{\lambda\eta}}{\partial x^\xi}\right). \tag{10.23}$$

We would like to point out the analogy between (10.22) and (1.27). The latter equation can be written as $dx'_i/dt = dx_i/dt - \epsilon_{ijk}\omega_j x_k$, and expresses the transformation of the velocity of a body from an inertial to a rotating (non-inertial) frame in Newtonian mechanics. Actually, (10.22) contains a generalization of (1.27), as a covariant derivative since, due to the principle of equivalence, the rotating system is equivalent (locally) to a gravitational field. Incidentally, covariant derivatives related to gauge transformations are defined also in the theory of Yang–Mills fields (see Chap. 11).

The metric tensor $g_{\mu\nu}(x)$ describes the gravitational field in the general theory of relativity. If a falling elevator is used as a system of reference, in such a system the interval between two very close events will take the form (10.20). That is, by making a transformation of coordinates to such a system, the expression (10.21) takes the form (10.20), and the Einstein metric becomes locally Minkowskian. We say 'locally' since this transformation is only valid in an infinitesimally small region. The point is that a gravitational field can only be made to vanish in the neighbourhood of a given point. As pointed out by Einstein:

> In the immediate vicinity of an observer that falls freely in a gravitational field, the gravitational field does not exist.

This establishes an essential difference between a real gravitational field and a fictitious one (created by a non-inertial system). The fictitious gravitational field can simply be eliminated *at all spacetime points* by making an appropriate transformation of coordinates. A real gravitational field cannot be eliminated in this way.

Starting from the metric tensor $g_{\mu\nu}$ (and its contravariant associated tensor $g^{\lambda\eta}$), it is possible to build other mathematical entities, such as the Riemann–Christoffel tensor $R_{\mu\nu\lambda\eta}$, the Ricci tensor $R_{\mu\nu}$, which describes the curvature of spacetime, and the scalar curvature $R = g^{\mu\nu}R_{\mu\nu}$. The tensor $R_{\mu\nu}$ is defined by

$$R_{\mu\nu} = \frac{\partial \Gamma^{\eta}{}_{\mu\nu}}{\partial x^{\eta}} - \frac{\partial \Gamma^{\eta}{}_{\mu\eta}}{\partial x^{\nu}} + \Gamma^{\eta}{}_{\mu\nu}\Gamma^{\lambda}{}_{\eta\lambda} - \Gamma^{\lambda}{}_{\mu\eta}\Gamma^{\eta}{}_{\nu\lambda}, \tag{10.24}$$

and the scalar curvature is $R = g^{\mu\nu}R_{\mu\nu}$. Remark that in general relativity the tensors are covariant under general coordinate transformations. A non-vanishing Riemann tensor is the covariant criterion to define a curved spacetime, as this tensor is identically zero for the flat Minkowski spacetime.

A distribution of matter or radiation is described in general relativity by means of another mathematical entity: the energy–momentum tensor $T_{\mu\nu}$. For a relativistic fluid in thermal equilibrium having pressure p, energy density ϵ, and velocity four vector u_μ, one finds:

$$T_{\mu\nu} = (p + \epsilon)\frac{u_\mu u_\nu}{c^2} - p g_{\mu\nu}. \tag{10.25}$$

10.4 Gravitational Field Equations

The gravitational field equations in general relativity, named Einstein's equations, establish a relation between the geometrical properties of the spacetime, expressed by the metric tensor $g_{\mu\nu}$, the Ricci tensor $R_{\mu\nu}$, and the spacetime curvature R on the one hand, and the distribution of mass and energy, represented by the energy–momentum tensor of matter, $T_{\mu\nu}$, on the other hand:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu}, \qquad (10.26)$$

where G is the gravitational constant. Einstein's equations for the gravitational field are analogous to Maxwell's equations in classical electrodynamics. There are, however, three important differences:

1. Maxwell's equations apply to inertial systems. The equations of the gravitational field apply to arbitrarily moving systems;
2. Maxwell's equations do not contain the equations of motion of the charges which produce the electromagnetic field. However, the gravitational field equations provide the equations of motion for the particles producing the field;
3. Maxwell's equations are linear differential equations in the electromagnetic potential $A_\mu(x)$, while the gravitational field equations are highly non-linear in $g_{\mu\nu}(x)$, whose components represent the generalized gravitational potential.

In particular, from the latter feature, in the quantum version of the theory, we would expect the gravitons or quanta of the gravitational field (the gravitational analog of photons) to be able to split and generate other gravitons. Photons, on the other hand, do not split into pairs of photons (in vacuum), in standard quantum electrodynamics. Moreover, there is an analogy between the Lorentz force in electromagnetism and the gravitational force on a moving mass, as pointed out previously. If we denote $h = -g_{00}$ and if we define the three-dimensional vector $\mathbf{g}$ with components $g_i = g_{0i}/g_{00}$, where $i = 1, 2, 3$, for a constant gravitational field (the components of the metric tensor do not depend on time), one can write this force as

$$\mathbf{F} = \frac{mc^2}{\sqrt{1 - V^2/c^2}} \left\{ -\nabla \ln \sqrt{h} + \sqrt{h}\frac{\mathbf{v}}{c} \times (\nabla \times \mathbf{g}) \right\}. \qquad (10.27)$$

For small velocities, the first term corresponds to the well-known force of gravity, and it is the analog of the electrostatic attraction, while the second term depends on the velocity, as does the magnetic force, and it is equal to the Coriolis force in a rotating system with angular velocity $\Omega = \frac{c}{2}\sqrt{h}\nabla \times \mathbf{g}$. But for the latter to become significant, e.g., in the case of the planets, they would have to move at high speed, comparable with the speed of light.

In addition, as for the electromagnetic field, there should be gravitational waves, that is, deformations of the spacetime geometry propagating at the speed of light. But even for very massive astronomical objects, the amount of gravitational energy radiated is extremely small. For example, for a system of binary stars, the radiation emitted in a year would be 10^{-12} of the total energy of the system. The so-called Hulse–Taylor binary is a pair of stars, one of which is a pulsar. They each have masses around 1.4 $M_\odot$ and the distance between them is around 2×10^6 km, of the order of the Sun's diameter. They are expected to radiate 10^{22} times the gravitational energy radiated by the Earth–Sun system. This causes the stars to gradually move closer together, in what is known as an *inspiral*, and this has an effect on the observed pulsar's signals.

Russell Hulse (b. 1950) and Joseph Taylor (b. 1941) were awarded the Nobel Prize in 1993 for their measurements which led to the discovery of the first binary pulsar, and allowed them to show that the gravitational radiation predicted by general relativity matched the results of these observations with a precision within 0.2%. This was the first indirect evidence for gravitational energy radiation, which is understood as a wave phenomenon.

Observation of Gravitational Waves

The search for direct evidence of gravitational waves lead to a great success, using mainly detectors based on laser interferometry, like LIGO on Earth ground (Laser Interferometer Gravitational Wave Observatory) in Livingstone, Louisiana, and the Hanford Site in the state of Washington. The Laser Interferometer Space Antenna (LISA) is designed to detect gravitational waves at frequencies not observable by ground based interferometry, and planned to operate in the near future. LISA is a giant interferometer, composed of three satellites forming an equilateral triangle with the sides 2.5 million km long.

As a gravitational wave passes through matter, a distortion in space-time produced by the gravitational wave leads to a tiny lengthening or contraction of objects, like the arms of an interferometer. This makes interferometry-based devices particularly useful for the detection of such waves. A modified Michelson interferometer is used to measure gravitational-wave strain through the difference in length of its orthogonal arms. LIGO is the largest interferometer ever built and the most sensitive detector, possessing a measurement sensitivity of about one part in 5×10^{22}. Each arm is formed by two mirrors, acting as test masses, separated by a distance $L_x = L_y = L = 4$ km. When a gravitational wave passes, it alters the arm lengths such that the measured difference is $\Delta L = \delta L_x - \delta L_y = h(t)L$, where h is the gravitational-wave strain amplitude projected onto the detector. This length variation produces a phase difference between the two light beams returning to the splitter, transmitting an optical signal proportional to the gravitational-wave strain to the output photodetector.

When a gravitational wave enters, one of the arms of the interferometer is lengthened. Mirrors placed near the beam splitter cause multiple reflections of the laser beam, increasing the distance traveled in each arm to 1120 km. This system of mirrors forms an optical resonator known as a Fabry–Pérot cavity. The output is the signal coming from the interference of the two beams, showing the shape of the incoming gravitational wave.

Up to 2015, evidence for black holes could only be obtained through electromagnetic signals, although evidence for the radiation of gravitational waves was provided by the Hulse–Taylor observations. However, the merging of two black holes by detection of the emitted gravitational waves was first observed on 14 September 2015. LIGO reported the observation of a signal corresponding to the wave predicted by general relativity for the merger of two black holes with masses $29M_\odot$ and $36M_\odot$ about 1.3 billion light years away. The final black hole mass was estimated to be of order $62M_\odot$. The difference of $3M_\odot$ was radiated as gravitational waves. A second set of gravitational waves was reported in December 2015. They represented the merger of two black holes about 1.4 billion light years aways, with masses of about 14.2 and 7.5 solar masses, yielding a final black hole of around 20.8 solar masses, with one solar mass radiated away as gravitational waves. In 2017, the Nobel Prize in Physics was awarded to Rainer Weiss (b. 1932), Kip Thorne (b. 1940), and Barry Barish (b. 1936) "for decisive contributions to the LIGO detector and the observation of gravitational waves".

On 17 August 2017, scientists also witnessed a process in which two neutron stars spiralled into each other and merged, producing a black hole. The event was first detected by the gravitational waves this generated. Scientists immediately knew it was due to two spiralling neutron stars, which were already emitting radiation before they merged. The radiation was detected by 70 observatories around the world, ranging from gamma ray detectors to radio telescopes. They confirmed several key astrophysical models, and revealed the birthplace of some heavy elements like gold and platinum. Above all, they were able to further test general theory of relativity.

10.5 Cosmology

If Einstein's equations (10.26) are solved for a gravitational field produced in vacuum by a body of mass M with spherical symmetry, and such that the metric does not depend on time and is asymptotically flat, the interval ds^2 is given by the expression obtained by Karl Schwarzschild (1873–1916) in 1915:

$$ds^2 = \left(1 - \frac{r_g}{r}\right) c^2 dt^2 - r^2(\sin^2\theta d\varphi^2 + d\theta^2) - \frac{dr^2}{1 - \frac{r_g}{r}}, \tag{10.28}$$

where $r_g = 2GM/c^2$ is the Schwarzschild radius of a spherical body of mass M. For $r = r_g$, $g_{00} = 0$ and $g_{11} \to \infty$ with the formation of the so-called event horizon of a black hole. An event horizon is a boundary in spacetime beyond which events cannot affect an outside observer. Such a region of spacetime is called a *black hole*. In 2020, the British mathematician Roger Penrose was awarded the Nobel Prize in Physics "for the discovery that black hole formation is a robust prediction of the general theory of relativity".

The Russian physicist Alexander A. Friedmann (1888–1925) studied the Einstein equations as applied to the Universe, assuming a homogeneous and isotropic density, and he concluded that there are two possible solutions: the closed and the open models. The latter leads to a perpetual expansion. At the boundary between the open and the closed models, there is the flat solution. Physically, the condition for open, closed, or flat Universe is determined by the density (of matter or energy).

If the distance between two galaxies is taken as $d(t) = R(t)d_0$, their relative speed can be written as $v = [\dot{R}(t)/R(t)]d(t)$, i.e., the speed is proportional to the separation between the two galaxies, with a proportionality factor $H(t) = \dot{R}(t)/R(t)$ which is called the Hubble parameter. Its present value is usually represented by H_0 and called Hubble's constant. We call $R(t)$ the *cosmic scale factor*, and here we take it to be dimensionless, while d_0 has the dimension of length. Below we shall consider $R(t)$ frequently as containing implicitly the d_0 factor and having dimensions of length. Concerning $H(t)$, it has the dimension of inverse time.

We shall discuss the problem of the motion of a galaxy by using the Newtonian mechanics of Chap. 1, but taking into account Hubble's law. Let us consider the mass of the galaxy as m, under the gravitational attraction of the rest of the Universe, of mass M. As $M \gg m$, one has $M + m \simeq M$ and the total energy is

$$\frac{1}{2}mv^2 - \frac{GMm}{r} = E. \tag{10.29}$$

Let us write $v = \dot{R}(t) = H(t)R(t)$ and $r = R$, where $H(t)$ is the Hubble parameter and R is the radius of the Universe. For a spherical mass distribution, the total mass is $M = \frac{4}{3}\pi R^3\rho$, where ρ is the average mass density of the Universe, and we substitute this expression into (10.29). This gives

$$\frac{\dot{R}^2(t)}{2} - \frac{4\pi\rho G R^2(t)}{3} = \frac{E}{m} = \frac{-K}{2}. \tag{10.30}$$

This is a non-relativistic way of obtaining Einstein's equation from the Friedmann model for the expansion of the homogeneous and isotropic Universe. The latter is identical to the one obtained using the relativistic formalism starting from the Robertson–Walker metric, which is a metric compatible with the conditions of homogeneity and isotropy (these conditions are sometimes called *cosmological principle*):

$$ds^2 = c^2dt^2 - R^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\varphi^2)\right]. \tag{10.31}$$

Here $k = -1, 0, 1$ correspond to open, flat, and closed cosmologies, respectively. Observe that K in (10.30) has the dimension of the square of a velocity, while k in (10.31) is dimensionless, because $R(t)$ has the dimension of length, and r is dimensionless. Then we have $K \sim kc^2$. According to (10.30), the critical condition to bring the expansion asymptotically to a halt occurs for $k = 0$, that is to say, for the density

$$\rho_c = \frac{3H^2}{8\pi G}. \tag{10.32}$$

With the present-day value of the Hubble parameter, H_0, the value of ρ_c is of the order of 10^{-29}g cm^{-3}.

But the Robertson–Walker metric does not tell us anything about the time dependence of the scale factor $R(t)$. To obtain this information, one must solve not only the Einstein equations, that is, (10.30) and (10.34) below, but also the equation of conservation of energy and the equation of state. Let us discuss the simplest case of a flat Universe. If we expand $R(t)$ in a power series around the reference time t_0, taken as the present time, we get $R(t) = R(t_0)[1 + H_0(t - t_0) - \frac{1}{2}q(t_0)H_0^2(t - t_0)^2 + \cdots]$, where the so-called *deceleration parameter* is given by

$$q(t) = -\frac{\ddot{R}(t)R(t)}{\dot{R}^2(t)}. \tag{10.33}$$

This quantity was estimated to be of the order of -0.5 at present, indicating that the expansion of the Universe is accelerated. The value of the deceleration parameter is a major topic in the present day cosmological research.

Together with (10.30) we must consider the other Einstein equation,

$$\ddot{R}(t) = -\frac{4\pi G}{3}R(t)\left(\rho + \frac{3p}{c^2}\right). \tag{10.34}$$

For $\rho > 0$ and $p > 0$, the acceleration $\ddot{R}$ is negative, and consistent with a positive deceleration. But as will be pointed out later, dark energy may provide a negative value for the factor $(\rho + 3p/c^2)$, producing an accelerated expansion of the Universe. We postpone the discussion of this case and continue with the solutions for standard cosmology. We denote $\Omega = \rho/\rho_c$. Then we can write (10.30) in terms of the Hubble parameter as follows:

$$H^2(\Omega - 1) = KR^{-2}(t). \tag{10.35}$$

If one assumes the pressure to be negligible compared with the density, that is to say $p \simeq 0$, simple solutions of the Friedmann model are found. In the flat case ($k = 0$, $q_0 < 0.5$, $\Omega = 1$), one has

$$R(t) = [3GM/\pi]^{1/3}t^{2/3}, \qquad H = 2/3t. \tag{10.36}$$

In the closed case ($k = +1$, $q_0 > 0$, $\Omega > 1$), the Universe has a finite volume, but it is unbounded (this corresponds to the previously mentioned space which can be regarded as a generalization of the spherical surface to three dimensions). In such a case, one obtains solutions in terms of a parameter η, defined by $d\eta = R(t)dt$:

$$R(\eta) = (2GM/3\pi c^2)(1 - \cos\eta), \qquad t(\eta) = (2GM/3\pi c^3)(\eta - \sin\eta). \tag{10.37}$$

In both the open ($k = -1$, $\Omega < 1$) and the flat cases, the Universe is infinite and unbounded. In the open case, one has

$$R(\eta) = (2GM/3\pi c^2)(\cosh\eta - 1), \qquad t(\eta) = (2GM/3\pi c^3)(\sinh\eta - \eta). \tag{10.38}$$

In none of the three cases is the Universe static, and it should be either expanding or contracting. Expansion is interpreted as meaning that the galaxies separate with increasing speed because their mutual separation increases. But if this occurs, there should be a *redshift* in the spectra of light coming from remote galaxies. The effect was observed for the first time in 1912 by Vesto Slipher (1875–1969) at the Lowell Observatory in Flagstaff, Arizona.

If ν_E is the emitted frequency and ν_O the observed one, the redshift is measured by a quantity $z = (\nu_E/\nu_O) - 1$. If $\nu_E > \nu_O$, the light is redshifted and $z > 0$. In the opposite case, if $\nu_E < \nu_O$, then $z < 0$, and the spectrum is shifted to the blue.

Edwin Hubble (1889–1953) discovered that the distances to the far-away galaxies are roughly proportional with their redshifts, which is now known as Hubble's law. Hubble reached this conclusion by interpreting his own measurements of galaxy distances and the galactic redshift measurements of Slipher. George Lemaître (1894–1966) had been the first to report this result in 1927 and to propose the theory of the expansion of the Universe. As pointed out before, as our Universe expands, the galaxies recede from each other with increasing speed. This expansion suggests that there was necessarily an initial moment in which all the matter composing these galaxies, and all intergalactic matter, was concentrated in a small region of the Universe. A great explosion, the Big Bang, occurred at some time around 10 to 20 billion years ago. The most recent estimate by the Planck collaboration for the age of the Universe, i.e. the time since the Big Bang, is 13.79 billion years. The Big Bang theory was proposed by Lemaître in 1931, but the term Big Bang was coined later.

Over the last few decades a theory has been proposed on the hypothesis that, in the early stages of the Universe, there was an exponential expansion. This phase was called *inflation* in the 1980s. It has been suggested that this could be described by a coupling between the gravitational field and some scalar field which is displaced from its equilibrium configuration. This point will be discussed further in Chap. 11.

With regard to the distribution of galaxies, moving away from each other in all space directions, observations indicate that they are grouped into clusters or superclusters, separated by empty space, with a cellular distribution. This in turn suggests a three-dimensional structure of these clusters separated by empty space, with some regularity, on a gigantic scale of 390 million light-years, in a form similar to a honeycomb.

The temperature of the *primeval fireball* in which the matter composing our visible Universe was concentrated was extraordinarily large, of the order of 10^{32} K, but it would have decreased quickly to values between 10^{10} and 10^9 K a few seconds after the Big Bang. This stage is said to be radiation dominated, because the density of the radiation was significantly greater than the density of matter. For instance, the photon density was much higher than the baryon density. As the initial ball cooled down in the process of expansion, the atoms of the light elements would have condensed out, while heavier atoms would have formed later inside the stars.

With the expansion of the Universe, the average temperature has decreased, and the whole system has cooled down, going through a matter-dominated era, when most of the energy of the Universe was concentrated in the masses of the nuclear particles. At present, the Universe is dominated by dark energy, which drives the cosmic acceleration.

As a result of this cosmological process, one may expect some fingerprint of the radiation-dominated era during the first stages after the Big Bang. George Gamow (1904–1968) predicted the existence of a background radiation, corresponding to a black body at very low temperatures. In 1964, Arno Allan Penzias (b. 1933) and Robert Woodrow Wilson (b. 1936) discovered this fossil radiation, a discovery for which they were awarded the Nobel Prize in 1978. The background radiation comes from all directions of space and it corresponds to a black body radiation at a temperature of about 2.725 K. It is called the cosmic microwave background (CMB). This radiation has a density of 4.40×10^{-34} g/cm^3, while the density of matter is of the order of 10^{-29} g/cm^3, that is, 10^5 times greater. For a certain time this justified the claim that we live in a matter-dominated era. At the present time, this view has changed due to current hypotheses about dark matter and dark energy, which we shall come back to in Sect. 10.6.2.

10.6 Gravitational Radius and Collapse

The idea of escape velocity is well known: it is the minimum velocity one must give a body so that it can escape from the Earth's gravitational field. If one neglects air resistance, the problem reduces to solving the equation in which the total energy of the particle in the gravitational field is equal to zero, viz.,

$$\frac{1}{2}mv^2 - \frac{GMm}{r} = E = 0. \tag{10.39}$$

Taking M and r as the Earth's mass and radius, this gives

$$v = \sqrt{\frac{2GM}{r}}. \tag{10.40}$$

If the Earth's radius decreased to one quarter (but keeping the same total mass), the escape velocity is doubled. But one can also consider the opposite problem: to which radius would we have to compress the Earth to reach a given value of the escape velocity? Let us suppose $v = c$, the speed of light. Then the value obtained for the radius R is the gravitational or Schwarzschild radius mentioned above,

$$R = \frac{2GM}{c^2} \equiv r_g. \tag{10.41}$$

For M of the order of the Earth's mass (6×10^{27} g), $R \approx 0.9$ cm. So if the Earth's mass were compressed to such an incredibly small size, no object could escape from inside, and only light emitted vertically would be able to get outside.

For a radius smaller than this value of R, the Earth would be transformed into a black hole, and even light could not escape from it. A black hole would absorb all the substance and radiation in its surrounding space. The existence of black holes, which we have argued mainly from non-relativistic mechanics, is a consequence of the general theory of relativity. For every body of mass M, a corresponding gravitational radius can be calculated by dividing its mass M (multiplied by the gravitational constant G), by the square of the speed of light. We have already seen that the Earth's gravitational radius is of the order of 0.45 cm. A similar calculation carried out for the Sun would give a sphere of radius about 3 km. Assuming a spherical shape and density ρ, its mass would be $M = \frac{4}{3}\pi R^3 \rho$. Then

$$R = \frac{8\pi}{3c^2} G R^3 \rho, \tag{10.42}$$

which implies that $\rho = 3c^2/(8\pi G R^2)$, that is, the density required to achieve the gravitational radius condition decreases as the reciprocal of the square of the radius. In other words, the larger the mass, the smaller the density required to achieve the gravitational radius condition. For instance, for our galaxy, if we assume a mass 10^{44} g (that is 10^{11} times that of the Sun, whose mass is about 2×10^{33} g), the gravitational radius is

$$R \approx 10^{11} \text{ km}, \tag{10.43}$$

which is about a hundredth of a light-year (one light-year is approximately 9.4×10^{12} km). The radius of our galaxy is about 55,000 light-years, i.e., $\sim 5.2 \times 10^{17}$ km. The gravitational radius would be reached by reducing the galactic radius to one millionth of its present size.

If for the Universe we estimate a mass of 10^{80} times the proton mass, that is, about 10^{56} g, the corresponding gravitational radius would be of the order of 10^{10} light-years. This is of the same order as the estimated radius of the Universe, the distance of the most remote cosmic objects. It has thus been speculated that the whole of our visible Universe is a black hole. Such an idea is in contradiction with the current cosmology.

If a star explodes in a supernova, its nucleus may be compressed to such a density that it becomes a neutron star, with a density of about 10^{15} g/cm^3. If its mass is greater than 2.5 times the mass of the Sun, gravity dominates over any other force resisting the compression. A gravitational collapse then occurs leading to the formation of a black hole. The gravitational radius determines the so-called event horizon. All the radiation and matter surrounding it would be absorbed by the black hole, and it would disappear below the horizon (Fig. 10.7). An observer inside a black hole (if it could survive the forces generated inside) could find out about what happens outside, but could never communicate with external observers, since it would be impossible to send out a signal.

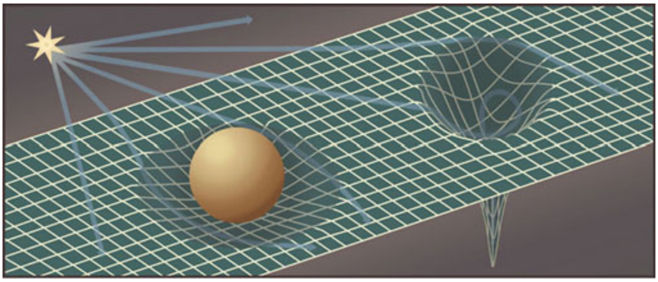


Fig. 10.7 Light from a distant star is deviated by a heavy body, which deforms spacetime around it (part of it is represented schematically by a two-dimensional mattress). A black hole captures light as well as matter incident on it.

Under such extreme conditions in which the gravitational force becomes so large, classical ideas cease to be valid, and we are in a situation similar to that of atomic theory as described by classical electrodynamics, according to which the atom would disappear in a collapse. Under such extreme conditions, quantum effects would thus enter the game in a predominant way.

Stephen Hawking (1942–2018) suggested in 1974 that black holes can evaporate in a gas of photons and other particles, by a quantum mechanism: the tunnel effect. In Chap. 7, we saw that particle and antiparticle pairs are created and annihilated spontaneously in vacuum. The process of pair formation has a characteristic time, given by the Heisenberg uncertainty principle:

$$\tau = h/E, \tag{10.44}$$

where E is the energy required for pair formation. Associated with the black hole, there is also a characteristic time τ' given by

$$\tau' = \frac{R}{c}, \tag{10.45}$$

where $R = 2GM/c^2$ is the gravitational radius of the black hole. If $\tau' < \tau$, the pair production process may be possible, at the expense of the mass (energy) of the black hole, and one particle of the pair can tunnel out of the black hole. The black hole temperature is inversely proportional to its mass and it would radiate energy proportionally to the fourth power of the temperature. Jacob Bekenstein (1947–2015) conjectured in 1972 that the area of the event horizon is proportional to the black hole entropy. This is intuitively comprehensible if one remembers that, when two black holes of masses M_1 and M_2 (and radii r_1 and r_2) join together, the area of the event horizon of the resulting black hole is always *larger* than the sum of the areas of the original black holes, because $(r_1 + r_2)^2 > r_1^2 + r_2^2$. As we see, the horizon surface area always grows with mass. In his work on black hole radiation of 1974, Hawking confirmed Bekenstein's conjecture and fixed the proportionality constant.

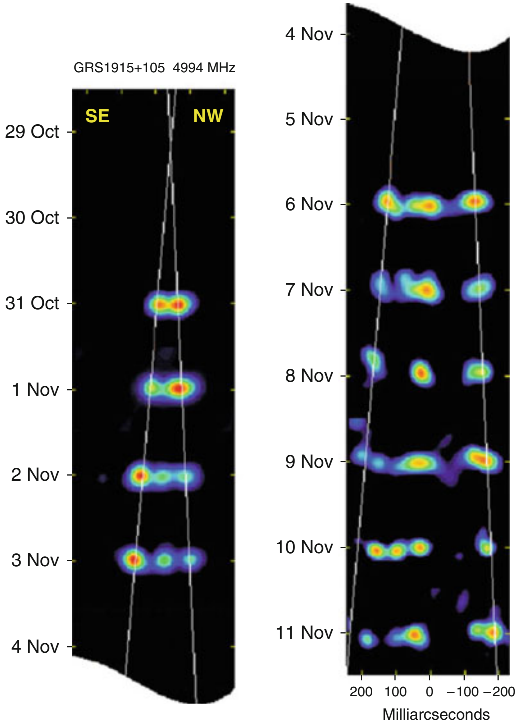


Fig. 10.8 MERLIN imaging of relativistic jets in the microquasar GRS1915+105, which is an X-ray binary system assumed to be composed of a rotating black hole and a normal star. The black hole has an accretion disk fed by gas from the star. It was the first known galactic source which ejects material with apparently superluminal velocities. This seems to be due to a relativistic effect known as Doppler boosting, produced by jets of particles moving with the speed of 0.9c (Courtesy of Fender et al., University of Manchester, Jodrell Bank Centre for Astrophysics).

Let us assume then that the entropy of a black hole is proportional to the area of the event horizon, $S = kA$. But $A = k'R^2 = 4k'G^2M^2/c^4$, where k and k' are constants of proportionality. Then $S \sim U^2$, with U the internal energy, which is proportional to the mass M. From this, one has $1/T = \partial S/\partial U \sim U$, and evidently $T \sim 1/M$. From here we deduce that a big black hole will radiate less than a small one, according to the law $1/M^4$, since the radiation power is proportional to $T^4 \sim 1/M^4$. A black hole is a system which loses information. If the black hole is in a pure quantum state when it begins, as it radiates thermal energy, it will pass to a mixed state with consequent loss of *quantum coherence*. This hypothesis is due to Hawking.

It is believed that the first indirect observation of a black hole was the binary system GRO J1655-40, assumed to comprise a black hole and a star, like the system GRS1915+105 (Fig. 10.8). Orbiting around the black hole, there is an accretion disk made up of material fed to it by the normal star, and this disk radiates in the X-ray region. Both binary systems are galactic 'microquasars' and may provide a link between the supermassive black holes which are believed to power extragalactic quasars and more local, accreting black hole systems.

In 2008, astrophysicists found compelling evidence that a supermassive black hole, called Sagittario, of more than 4 million solar masses is located at the centre of the Milky Way. Supermassive black holes were subsequently found at the centre of all known galaxies.

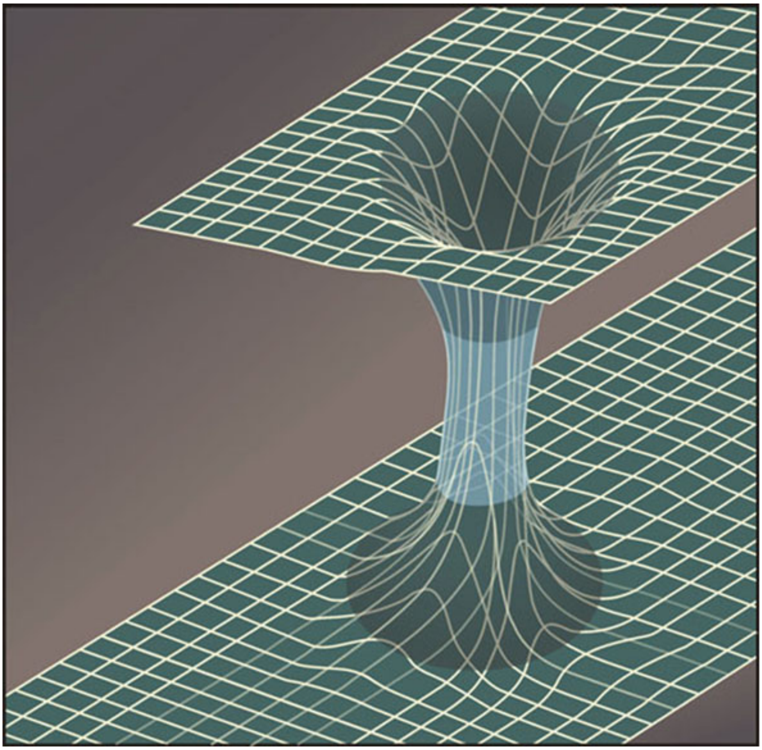


Fig. 10.9 A wormhole is a shortcut between separate regions of spacetime.

It has been suggested that instead of an infinite compression, the substance contained inside the black hole may emerge in another region of the spacetime. This would lead to a *white hole*, which would be a black hole running backward in time. In other words, gravitational collapse might cause an interconnection between two remote regions of spacetime.

10.6.1 Wormholes

The hypothetical bridges between separate regions of spacetime are called wormholes. These would be shortcuts between areas of space otherwise separated by long distances (Fig. 10.9). The wormhole would have two mouths (which are spheres in 3D space) and a throat between them. Standard theory indicates that they would not generally be stable. To be stable, or *traversable*, some exotic matter with negative energy density would be required in the throat. But this point remains open, since the assumption of extra dimensions would provide new scenarios. It should be noted that, although faster-than-light speeds remain forbidden locally, through a wormhole, even moving at speeds smaller than c, it would be possible to connect two events A, B by an interval of time $\tau < t_{AB}$, where $t_{AB} = l_{AB}/c$ and l_{AB} is the distance across standard space. This leads to an effective faster-than-light communication. Even time travel seems possible through wormholes.

10.6.2 Dark Matter, Dark Energy, and Accelerated Expansion

In an open Universe there would be perpetual expansion, whereas in a closed one the expansions and contractions would alternate in huge cycles. The time required for each of these cycles defies all imagination.

We have seen from (10.32) how to find ρ_c. The most recent value measured for the Hubble constant H_0 is 67.8 km/s Mpc^{-1}, where 1 Mpc $= 3.26 \times 10^6$ light-years. The critical density ρ_c is of order 10^{-29} g/cm^3. There is strong evidence, most recently from the Planck space mission, that the observable Universe is flat, i.e. $\rho = \rho_c$, which is consistent with inflationary models. But the flatness implies the existence of *dark matter* and *dark energy*, in significant amounts compared to the usual matter.

At present, dark matter appears to be an unavoidable hypothesis, providing some missing matter needed to explain the observed rotational velocities of galaxies, orbital velocities of galaxies in clusters, gravitational lensing of background objects by galaxy clusters, and other observable phenomena.

Most dark matter does not interact with electromagnetic radiation. It is thus transparent. However, there is not yet any satisfactory model for dark matter. For instance, it could be that some as-yet undiscovered weakly interacting particles were created during the Big Bang and today remain in significantly large amounts to account for the dark matter. The name of *weakly interacting massive particles* (WIMPs) has been suggested for some of these candidates for dark matter, assuming that it is nonbaryonic, i.e., that it contains no atoms. In addition to WIMPs, the nonbaryonic candidates for dark matter include neutrinos and hypothetical particles such as axions or supersymmetric particles (see Chap. 11).

However, certain astronomical objects may constitute the dark matter, but escape detection. For instance, brown dwarf stars with very small mass or black hole remnants of an early generation of stars would be similarly invisible. A small fraction of this hypothetical dark matter is referred to as MACHO an acronym for *massive (astrophysical) compact halo object*, made up of baryonic matter. Yet a large fraction of the dark matter has to be of a non-baryonic nature.

At present it is believed that ordinary matter constitutes only around 4.9% of the mass of the Universe, whereas dark matter would make up 26.8%, and the remaining 68.3% is thought to be due to dark energy. These percentages have varied over the last few years, since some measurements have been refined.

Since 1997, observations of supernovas of type Ia, which are excellent standard candles for measuring cosmological distances, suggest that the expansion of the Universe is actually accelerating. The Nobel Prize in Physics in 2011 was awarded to Saul Perlmutter (b. 1959), Brian Schmidt (b. 1967), and Adam Riess (b. 1969) for their discovery of the accelerated expansion of the Universe. This cannot be explained on the basis of the present gravitational interaction, and it requires an assumption of additional energies able to act as a repulsive force, for instance, whence the idea of dark energy. Such dark energy is assumed to be transparent.

The quantum vacuum was suggested in 1967 by Yakov Zel′dovich (1914–1987) as a candidate for dark energy, but present estimates give an extremely large figure for this, not compatible with what is expected by observation.

Let us consider the amount of dark energy inside a cylindrical cavity with a piston. The energy associated with a change of volume dV is $dE = -pdV$. If ρ_E is the energy density, we have $dE = \rho_E dV$. Thus, $\rho_E = -p$. The vacuum pressure is minus its energy density. In ordinary matter, we usually have $|p| \ll \rho_E$. This leads us to conclude that dark energy is essentially relativistic, able to interact in a repulsive way with ordinary matter. This would give a negative pressure term in the Einstein equations.

When Einstein wrote his equations, there was no knowledge of the expansion of the Universe. Hence, to make a static Universe from his model, Einstein introduced a cosmological constant. For years, this cosmological constant was taken as zero by cosmologists. The quantum vacuum effect is equivalent to assuming a nonvanishing cosmological constant.

Other researchers work with models based on appropriate scalar fields, called *quintessence*, able to generate similar effects. The problem is still open.

10.7 Gravitation and Quantum Effects

If we combine the constant of gravitation G, the reduced Planck constant $\hbar$, and the speed of light c, it is possible to estimate the order of magnitude at which quantum gravity phenomena are likely to manifest themselves. The combination with the dimension of length is

$$l_P = \sqrt{\frac{G\hbar}{c^3}} \approx 10^{-33}\,\text{cm}. \tag{10.46}$$

This is the so-called *Planck length*. It indicates the order of distances at which quantum gravitational effects are expected to appear. Starting from this value, it is possible to derive a number with dimensions of mass, named the *Planck mass*:

$$m_P = \sqrt{\frac{c\hbar}{G}} \approx 10^{-4}\,\text{g}. \tag{10.47}$$

The Planck mass can be interpreted as the mass of a body whose reduced Compton wavelength (characteristic of relativistic quantum effects) is equal to its

Schwarzschild or gravitational radius:

$$\frac{\hbar}{m_P c} = 2\frac{G m_P}{c^2},$$

which leads to the above expression for m_P. This mass has a macroscopic value, and can be used to obtain the equivalent energy

$$E_P = m_P c^2 = \sqrt{\frac{c^5 \hbar}{G}} \approx 10^{16} \text{erg} \approx 10^{19} \text{ GeV}. \tag{10.48}$$

This energy is so large that the gravitational field can give rise to the spontaneous creation of particle–antiparticle pairs. The average temperature associated with that energy is 10^{32} K which is believed to be the initial temperature of the *primeval fireball* from which the Big Bang was produced.

10.8 Cosmic Numbers

As pointed out earlier, the mass of the visible Universe is estimated as being 10^{80} times the proton mass. This is an incredibly large number. Other very large numbers (called cosmic numbers) appear in the physics of the microscopic as well as the macroscopic world. The first cosmic number is the ratio of the electromagnetic and gravitational forces exerted between an electron and a proton. Letting F_e and F_G be the moduli of these forces, one has

$$F_e = \frac{e^2}{r^2}, \quad F_G = \frac{G m_p m_e}{r^2}, \tag{10.49}$$

where e is the electron charge, G the constant of gravitation, and m_p and m_e the proton and electron masses. The first cosmic number N_1 is then

$$N_1 = \frac{F_e}{F_G} = \frac{e^2}{G m_p m_e} = 0.23 \times 10^{40}. \tag{10.50}$$

Being the ratio of two forces, it is a dimensionless number.

The second cosmic number is the quotient of the radius of the Universe L and the proton radius r_P. The number L is of order 10^{10} light-years, and one light-year is $\simeq 10^{18}$ cm, so that $L \approx 10^{28}$ cm. On the other hand, $r_P \simeq 10^{-13}$ cm. Dividing L by r_P, one obtains the second cosmic number

$$N_2 = \frac{L}{r_P} \approx 10^{40}. \tag{10.51}$$

The coincidence in the orders of magnitude of the two numbers is very striking, and Dirac suggested that there should be some relation between them. Now, as L increases with time, N_2 also increases, and if there is a relation between N_2 and N_1, the latter should vary with time. There is no evidence that e^2, m_e, or m_p vary with time, so this would leave open the possibility that G might be time-dependent. This topic remains open to speculation.

Problems

Problem 10.1 Calculate the relativistic effects on time measured by a clock on a GPS satellite, which runs faster than a clock on the Earth's surface, and show that it amounts to around 38×10^{-6} s per day.

Problem 10.2 Starting from $dl^2 = (1 - r_g/r)^{-1}dr^2 + r^2(\sin^2\theta d\varphi^2 + d\theta^2)$, which is the spatial part of ds^2 in the expression (10.28) for the metric outside a spherically symmetric gravitating body, (i) find an expression for the radial distance $(l_2 - l_1)$ between two circles of radii r_1 and r_2 concentric to the body's centre, in this geometry; and (ii) obtain the limit $r_2 > r_1 \gg r_g$. (iii) Apply to the case $r_1 = 7 \times 10^8$ m, which is the order of the average solar radius, $r_2 = 5.8 \times 10^{10}$ m, which is of the order of the average radius of Mercury's orbit, and $r_g = 3 \times 10^3$ m, which is approximately the Sun's gravitational radius.

Problem 10.3 On an intuitive quantum mechanical basis, justify the Hawking–Bekenstein expression for the black hole temperature $T = \frac{\hbar c^3}{8\pi GMk}$ (up to a constant factor).

Problem 10.4 Once a black hole has formed, its mass is extremely unevenly distributed within it. The usual concept of density applies more to the body undergoing gravitational collapse, so the density to which we refer in the present and following problems corresponds to the density of the collapsing body rather than to the resulting black hole. In any case, these problems serve to show the scales involved in black holes. As an example, calculate the size and density of the black hole at the centre of our Galaxy, called Sagittarius A^*, estimated to have a mass $M = 4 \times 10^6 M_\odot \sim 8 \times 10^{36}$ kg.

Problem 10.5 Calculate the Schwarzschild radius for a black hole with density equal to that of (a) water, (b) the estimated density of (normal) matter in the Universe, which is 10^{-29} g/cm^3.

Problem 10.6 The Hawking–Bekenstein black hole entropy formula is $S = \frac{kA}{4l^2}$, where $l = \sqrt{\frac{G\hbar}{c^3}}$ is the Planck length and $A = 4\pi R^2 = 16\pi G^2M^2/c^4$ is the area of the event horizon, since $R = 2GM/c^2$ is the gravitational radius. Use this to calculate the black hole temperature and internal energy.

Problem 10.7 Calculate the black hole heat capacity.

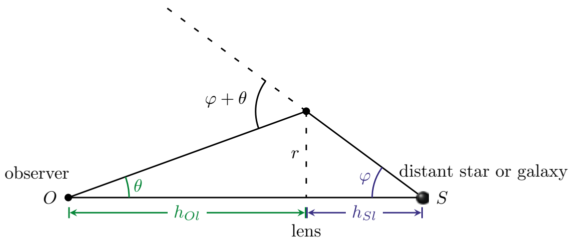


Fig. 10.10 We assume a perfect alignment and hence a complete Einstein ring.

Problem 10.8 Einstein ring due to a lens star. If the distance and mass of the lens are approximately known, along with the distance to the more remote object, the radius of the Einstein ring can be calculated. Consider a lens mass similar to the Sun's mass, and assume the distance from the observer to the remote star to be $H = 60\,\text{kpc}$. Assume also that the lens is located at a distance $H/3$ from the observer. Calculate the Einstein ring radius r assuming a perfect alignment. (1 parsec$=$ 3.2616 lyr $\approx 3.0857 \times 10^{16}$m) (Fig. 10.10).

Literature

1. M. Chaichian, I. Merches, D. Radu, A. Tureanu, *Electrodynamics. An Intensive Course* (Springer, Berlin, 2016). This is a very complete book, containing theory and applications, and the theoretical background and detection of gravitational waves are reported in a clear way
2. L.C. Epstein, *Relativity Visualized* (Insight Press, San Francisco, 1985). A highly appreciated intuitive approach to special and general relativity, replacing equations with diagrams and paper models
3. S. Hawking, *A Brief History of Time* (Bantam Books, New York, 1988). A unique book, recommended to all readers
4. L.D. Landau, E.M. Lifshitz, *The Classical Theory of Fields* (Pergamon, Oxford, 1975). The principles of general relativity are presented very clearly in this book
5. M.S. Longair, *Theoretical Concepts in Physics* (Cambridge University Press, London, 1984). There is a partial overlap in the scope of several sections of our book with that of this reference. The interested reader will find it useful to supplement several sections of the present chapter
6. T. Regge, *An Elementary Course on General Relativity* (CERN, Geneva, 1985). An excellent introduction to the topic, based on lectures given by the author
7. E.G. Thomas, D.J. Raine, *Physics to a Degree* (Gordon and Breach Science Publishers, 1998). Some of the proposed problems are adapted from those in this excellent book
8. S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972). This is a valuable book containing a large amount of information and adopting a very original approach, where the point of view of a particle physicist plays a more important role than in other books dealing with general relativity