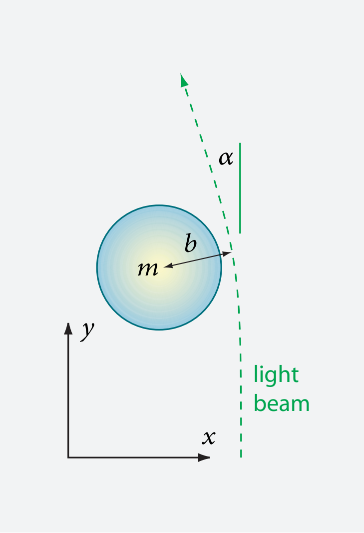


FIGURE 69 Calculating the bending of light by a mass.

BENDING OF LIGHT AND RADIO WAVES

Gravity influences the motion of light. In particular, *gravity bends light beams*. Indeed, the detection of the bending of light beams by the Sun made Einstein famous. This happened because the measured bending angle differed from the one predicted by universal gravitation and confirmed that of general relativity which takes into account the curvature of space. Page 143

The bending of light by a mass is easy to calculate. The bending of light is observed because any *distant* observer measures a changing value for the *effective* light speed v near a mass. (Measured at a location *nearby*, the speed of light is of course always c.) It turns out that a distant observer measures a *lower* speed, so that for him, gravity has the same effects as a dense optical medium. It takes only a little bit of imagination to see that this effect will thus *increase* the bending of light near masses already deduced in 1801 by Soldner from universal gravity. In short, relativistic light bending differs from non-relativistic light bending.*

Let us calculate the bending angle. As usual, we use the coordinate system of flat space-time at spatial infinity, shown in Figure 69. The idea is to do all calculations to first order, as the value of the bending is very small. The angle of deflection α, to first order, is simply Ref. 158

$$\alpha = \int_{-\infty}^{\infty} \frac{\partial v}{\partial x} \mathrm{d}y \,, \tag{151}$$

Challenge 245 e where v is the speed of light measured by a distant observer. (Can you confirm this?) For the next step we use the Schwarzschild metric around a spherical mass

$$\mathrm{d}\tau^2 = \left(1 - \frac{2GM}{rc^2}\right)\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{c^2 - \frac{2GM}{r}} - \frac{r^2}{c^2}\mathrm{d}\varphi^2 \tag{152}$$

Page 171 * In the vocabulary defined below, light bending is a pure gravitoelectric effect.

Challenge 246 ny

and transform it into (x, y) coordinates to first order. This gives

$$d\tau^2 = \left(1 - \frac{2GM}{rc^2}\right) dt^2 - \left(1 + \frac{2GM}{rc^2}\right) \frac{1}{c^2}(dx^2 + dy^2) \qquad (153)$$

which, again to first order, leads to

$$\frac{\partial v}{\partial x} = \left(1 - \frac{2GM}{rc^2}\right) c \, . \qquad (154)$$

This expression confirms what we know already, namely that distant observers see light *slowed down* when passing near a mass. Thus we can also speak of a height-dependent index of refraction. In other words, constant *local* light speed leads to a *global* slowdown.

Challenge 247 ny

Inserting the last result into expression (151) and using a clever substitution, we get a deviation angle α given by

$$\alpha = \frac{4GM}{c^2} \frac{1}{b} \qquad (155)$$

where the distance b is the so-called *impact parameter* of the approaching light beam.

▷ The light deviation angle α due to general relativity is *twice* the result for universal gravity.

Vol. I, page 201

For a beam just above the surface of the Sun, the bending angle has the famous value of $1.75'' = 8.5\,\mu\text{rad}$. This small value was spectacularly confirmed by the measurement expedition of 1919. (How did the astronomers measure the deviation angle?) The result showed that *universal gravity is wrong.* Since then, the experiment has been repeated hundreds of times, even by hobby astronomers.

Challenge 248 s

In fact, Einstein was lucky. Two earlier expeditions organized to measure the value had failed. In 1912, it was impossible to take data because of rain, and in 1914 in Crimea, scientists were arrested (by mistake) as spies, because the First World War had just begun. But in 1911, Einstein had already published an *incorrect* calculation, giving only the Soldner value with half the correct size; only in 1915, when he completed general relativity, did he find the correct result. Therefore Einstein became famous only because of the failure of the two expeditions that took place before he published his correct calculation!

Ref. 159

Vol. I, page 201

For high-precision experiments around the Sun, it is more effective to measure the bending of *radio* waves, as they encounter fewer problems when they propagate through the solar corona. So far, hundreds of independent experiments have done so, using radio sources in the sky which lie on the path of the Sun. All the measurements have confirmed general relativity's prediction within a few per cent or less. A beautiful example of such a measurement is shown in Figure 70. The left curve shows the measured values for expression (155); the right graph shows how the image of the radio source moves in the sky. Note the small angles that can be measured with the method of very long baseline interferometry nowadays.

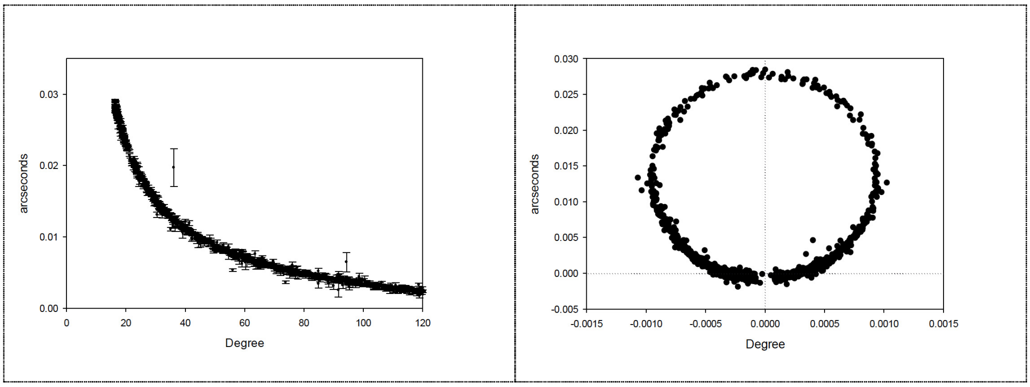
Ref. 151, Ref. 128

Ref. 129

Ref. 160

The bending of electromagnetic beams has also been observed near Jupiter, near certain stars, near several galaxies and near galaxy clusters. For the Earth itself, the angle

Page 252



Ref. 160 FIGURE 70 How the image of radio source 0552+398 changes in position over the course of ten years. Left: how the deviation changes with angular distance (impact parameter) from the Sun; right: how the position of the image in the sky changes from (0,0), the position in the sky when the quasar is far from the Sun (large impact parameter), to cases when the quasar image approaches the Sun (smaller impact parameter).

is at most 3 nrad, too small to be measured yet, even though this may be feasible in the near future. There is a chance to detect this value if, as Andrew Gould proposes, the data of the satellite Hipparcos, which was taking precision pictures of the night sky for many years, are analysed properly in the future.

Page 189 By the way, the bending of light also confirms that in a triangle, the sum of the angles does not add up to π (two right angles), as is predicted for curved space. What is the sign of the curvature? Challenge 249 e

TIME DELAY

The calculation of the bending of light near masses shows that for a distant observer, light is slowed down near a mass. Constant *local* light speed leads to a *global* light speed slowdown. If light were not slowed down near a mass, it would have to go faster than c for an observer near the mass!*

▷ Masses lead to a *time delay* of passing electromagnetic waves.

Ref. 161 In 1964, Irwin Shapiro had the idea to measure this effect. He proposed two methods. The first was to send radar pulses to Venus, and measure the time taken for the reflection to get back to Earth. If the signals pass near the Sun, they will be delayed. The second method was to use a space probe communicating with Earth.

Ref. 162 The first measurement was published in 1968, and directly confirmed the prediction of general relativity within experimental errors. All subsequent tests of the same type, such as the one shown in Figure 71, have also confirmed the prediction within experimental

---

* A nice exercise is to show that the bending of a *slow* particle gives the Soldner value, whereas with increasing speed, the value of the bending approaches twice that value. In all these considerations, the rotation of the mass has been neglected. As the effect of frame dragging shows, rotation also changes the deviation angle; however, in all cases studied so far, the influence is below the detection threshold. Challenge 250 e

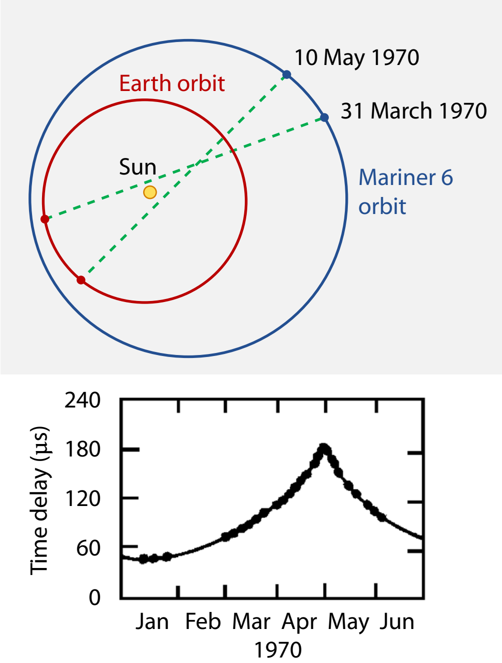


FIGURE 71 Time delay in radio signals – one of the experiments by Irwin Shapiro.

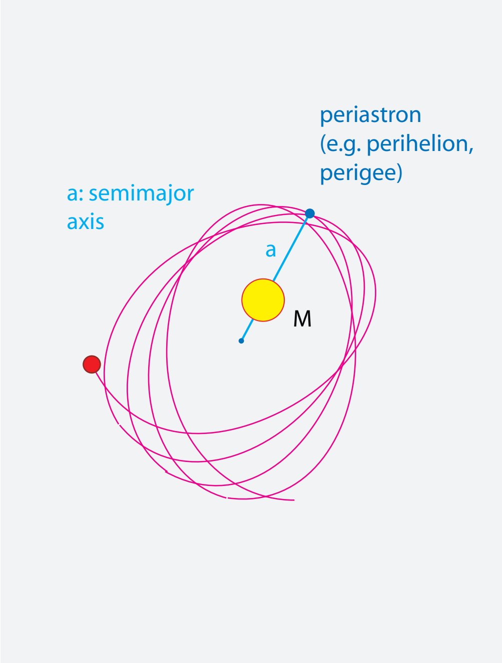


FIGURE 72 The orbit around a central body in general relativity.

errors, which nowadays are of the order of one part in a thousand. The delay has also Ref. 163 been measured in binary pulsars, as there are a few such systems in the sky for which the line of sight lies almost precisely in the orbital plane.

In short, relativistic gravitation is also confirmed by time delay measurements; in contrast, universal gravitation predicts no such effect. The simple calculations presented here suggest a challenge: Is it also possible to describe *full* general relativity – thus gravitation in *strong* fields – as a change of the speed of light with position and time induced by mass and energy? Challenge 251 ny

RELATIVISTIC EFFECTS ON ORBITS

Astronomy allows the most precise measurements of motions known. This is especially valid for planet motion. So, Einstein first of all tried to apply his results on relativistic gravitation to the motion of planets. He looked for deviations of their motions from the predictions of universal gravity. Einstein found such a deviation: *the precession of the perihelion of Mercury.* The effect is shown in Figure 72. Einstein said later that the moment he found out that his calculation for the precession of Mercury matched observations was one of the happiest moments of his life.

The calculation is not difficult. In universal gravity, orbits are calculated by setting $a_{\text{grav}} = a_{\text{centri}}$, in other words, by setting $GM/r^2 = \omega^2 r$ and fixing energy and angular momentum. The mass of the orbiting satellite does not appear explicitly. In general relativity, the mass of the orbiting satellite is made to disappear by rescaling energy and Ref. 128, Ref. 129 angular momentum as $e = E/c^2 m$ and $j = J/m$. Next, we include space curvature. We

Page 145
Challenge 252 e

use the Schwarzschild metric (152) mentioned above to deduce that the initial condition for the energy e, together with its conservation, leads to a relation between proper time τ and time t at infinity:

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \frac{e}{1 - 2GM/rc^2} , \tag{156}$$

whereas the initial condition on the angular momentum j and its conservation imply that

$$\frac{\mathrm{d}\varphi}{\mathrm{d}\tau} = \frac{j}{r^2} . \tag{157}$$

These relations are valid for any particle, whatever its mass m. Inserting all this into the Schwarzschild metric, we find that the motion of a particle follows

$$\left(\frac{\mathrm{d}r}{c\mathrm{d}\tau}\right)^2 + V^2(j,r) = e^2 \tag{158}$$

where the effective potential V is given by

$$V^2(J,r) = \left(1 - \frac{2GM}{rc^2}\right)\left(1 + \frac{j^2}{r^2c^2}\right) . \tag{159}$$

Challenge 253 e
Challenge 254 e

The expression differs slightly from the one in universal gravity, as you might want to check. We now need to solve for $r(\varphi)$. For *circular* orbits we get *two* possibilities

$$r_\pm = \frac{6GM/c^2}{1 \pm \sqrt{1 - 12(\frac{GM}{cj})^2}} \tag{160}$$

where the minus sign gives a stable and the plus sign an unstable orbit. If $cj/GM < 2\sqrt{3}$, no stable orbit exists; the object will impact the surface or, for a black hole, be swallowed. There is a stable circular orbit *only* if the angular momentum j is larger than $2\sqrt{3}GM/c$. We thus find that in general relativity, in contrast to universal gravity, there is a *smallest* stable circular orbit. The radius of this smallest stable circular orbit is $6GM/c^2 = 3R_S$.

What is the situation for *elliptical* orbits? Setting $u = 1/r$ in (158) and differentiating, the equation for $u(\varphi)$ becomes

$$u' + u = \frac{GM}{j^2} + \frac{3GM}{c^2}u^2 . \tag{161}$$

Challenge 255 e

Without the nonlinear correction due to general relativity on the far right, the solutions are the famous *conic sections*

$$u_0(\varphi) = \frac{GM}{j^2}(1 + \varepsilon\cos\varphi) , \tag{162}$$

i.e., ellipses, parabolas or hyperbolas. The type of conic section depends on the value of the parameter ε, the so-called *eccentricity*. We know the shapes of these curves from universal gravity. Now, general relativity introduces the nonlinear term on the right-hand side of equation (161). Thus the solutions are not conic sections any more; however, as the correction is small, a good approximation is given by

Vol. I, page 193

Challenge 256 e

$$u_1(\varphi) = \frac{GM}{j^2}\left(1 + \varepsilon\cos(\varphi - \frac{3G^2M^2}{j^2c^2}\varphi)\right) . \tag{163}$$

The hyperbolas and parabolas of universal gravity are thus slightly deformed.

▷ Instead of elliptical orbits, general relativity leads to the famous rosetta path shown in Figure 72.

Such a path is above all characterized by a periastron shift. The *periastron*, or *perihelion* in the case of the Sun, is the *nearest* point to the central body reached by an orbiting body. The periastron turns around the central body by an angle

Challenge 257 e

$$\alpha \approx 6\pi\frac{GM}{a(1-\varepsilon^2)c^2} \tag{164}$$

for every orbit, where a is the *semimajor axis*. For Mercury, the value is $43\,''= 0.21$ mrad per century. Around 1900, this was the only known effect that was unexplained by universal gravity; when Einstein's calculation led him to exactly that value, he was overflowing with joy for many days.

To be sure about the equality between calculation and experiment, all other effects leading to rosetta paths must be eliminated. For some time, it was thought that the quadrupole moment of the Sun could be an alternative source of this effect; later measurements ruled out this possibility.

In the past century, the perihelion shift has been measured also for the orbits of Icarus, Venus and Mars around the Sun, as well as for several binary star systems. In binary pulsars, the periastron shift can be as large as several degrees per year. In all cases, expression (164) describes the motion within experimental errors.

Ref. 163

We note that even the rosetta orbit itself is not really stable, due to the emission of gravitational waves. But in the solar system, the power lost this way is completely negligible even over thousands of millions of years, as we saw above, so that the rosetta path remains an excellent description of observations.

Page 179

The geodesic effect

Relativistic gravitation has a further effect on orbiting bodies, predicted in 1916 by Willem de Sitter.* When a *pointed* body orbits a central mass m at distance r, the *direction* of the tip will *change* after a full orbit. This effect, shown in Figure 73, exists only in general relativity. The angle α describing the direction change after one orbit is given

---

* Willem de Sitter (b. 1872 Sneek, d. 1934 Leiden) was mathematician, physicist and astronomer.

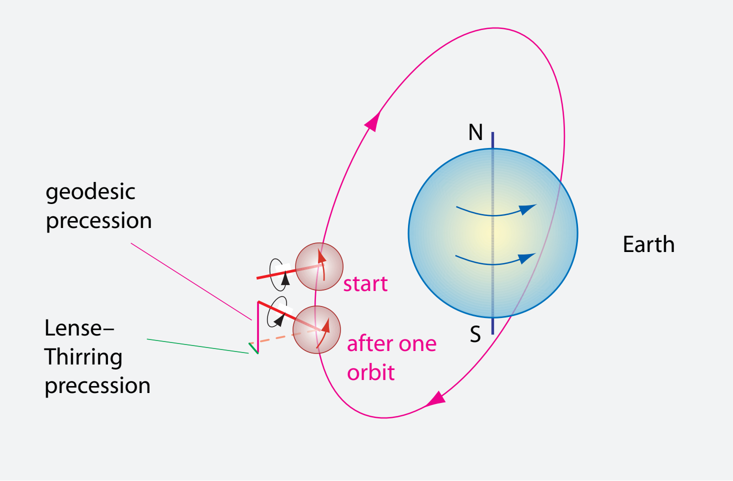


FIGURE 73 The geodesic effect.

by

$$\alpha = 2\pi \left(1 - \sqrt{1 - \frac{3Gm}{rc^2}} \right) \approx \frac{3\pi Gm}{rc^2} \ . \qquad (165)$$

This angle change is called the *geodesic effect* – 'geodetic' in other languages. It is a further consequence of the split into gravitoelectric and gravitomagnetic fields, as you may want to show. Obviously, it does not exist in universal gravity. Challenge 258 e

In cases where the pointing of the orbiting body is realized by an intrinsic rotation, such as a spinning satellite, the geodesic effect produces a *geodesic precession* of the axis. Thus the effect is comparable to spin–orbit coupling in atomic theory. (The Thirring–Lense effect mentioned below is analogous to spin–spin coupling.)

When Willem de Sitter predicted the geodesic effect, or geodesic precession, he proposed detecting that the Earth–Moon system would change its pointing direction in its fall around the Sun. The effect is tiny; for the axis of the Moon the precession angle is about 0.019 arcsec per year. Ref. 164 The effect was first measured in 1987 by an Italian team for the Earth–Moon system, through a combination of radio-interferometry and lunar ranging, making use of the Cat's-eyes, shown in Figure 74, deposited by Lunokhod and Apollo on the Moon. In 2005, the geodesic effect was confirmed to high precision with the help of an artificial satellite around the Earth that contained a number of high precision gyroscopes. Ref. 169

At first sight, geodesic precession is similar to the Thomas precession found in special relativity. Page 62 In both cases, a transport along a closed line results in the loss of the original direction. However, a careful investigation shows that Thomas precession can be *added* to geodesic precession by applying some additional, non-gravitational interaction, so the analogy is shaky.

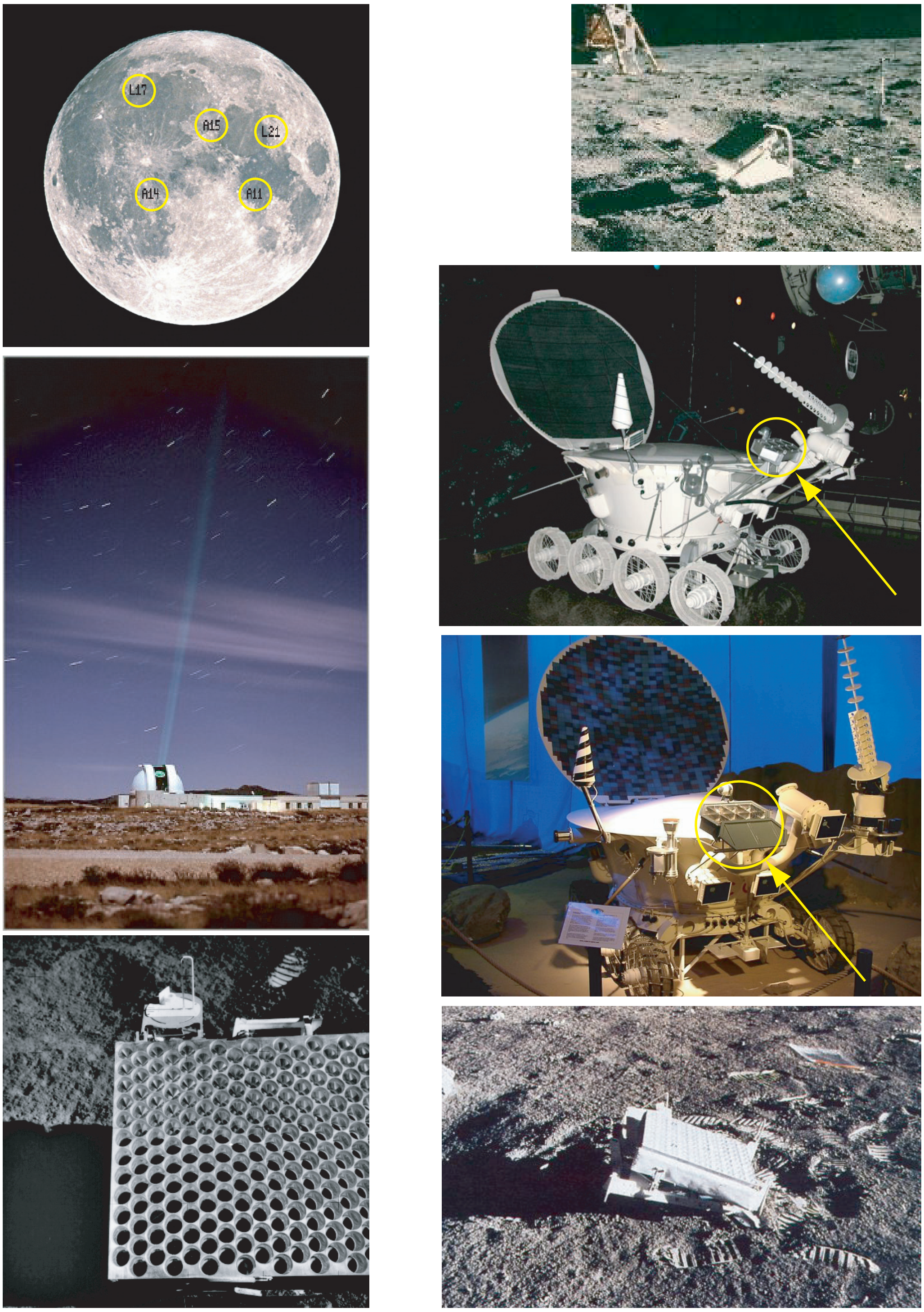


FIGURE 74 The lunar retroreflectors deposited by Apollo 11 (top right), the two Lunokhods (right), Apollo 14 (bottom right) and Apollo 15 (bottom left), their locations on the Moon (top left) and a telescope performing a laser distance measurement (© NASA, Wikimedia, Observatoire de la Côte d’Azur).

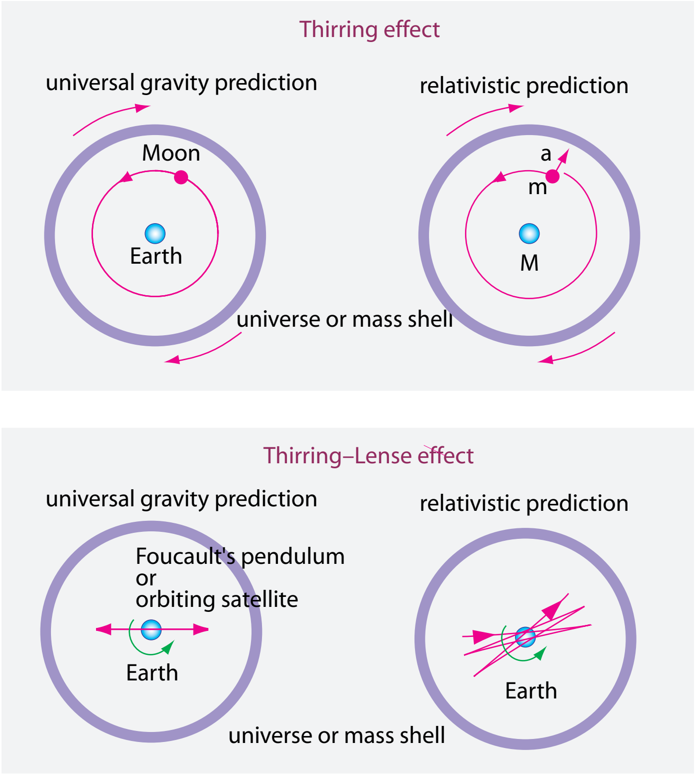


FIGURE 75 The Thirring and the Thirring–Lense effects.

THE THIRRING EFFECTS

In 1918, the Austrian physicist Hans Thirring published two new, simple and beautiful predictions of motions, one of them with his collaborator Josef Lense. Neither motion appears in universal gravity, but they both appear in general relativity. Figure 75 illustrates these predictions. Ref. 165, Ref. 166

The first example, nowadays called the *Thirring effect*, predicts centrifugal accelerations and Coriolis accelerations for masses in the interior of a rotating mass shell. Thirring showed that if an enclosing mass shell rotates, masses inside it are attracted towards the shell. The effect is very small; however, this prediction is in stark contrast to that of universal gravity, where a spherical mass shell – rotating or not – has no effect at all on masses in its interior. Can you explain this effect using the figure and the mattress analogy? Challenge 259 e

The second effect, the *Thirring–Lense effect*,* is more famous. General relativity predicts that an oscillating Foucault pendulum, or a satellite circling the Earth in a polar orbit, does not stay precisely in a fixed plane relative to the rest of the universe, but that the rotation of the Earth drags the plane along a tiny bit. This *frame-dragging*, as the effect is also called, appears because the Earth in vacuum behaves like a rotating ball in a foamy mattress. When a ball or a shell rotates inside the foam, it partly drags the foam

---

* Even though the order of the authors is Lense and Thirring, it is customary (but not universal) to stress the idea of Hans Thirring by placing him first.

FIGURE 76 The LAGEOS satellites: metal spheres with a diameter of 60 cm, a mass of 407 kg, and covered with 426 retroreflectors (courtesy NASA).

along with it. Similarly, the Earth drags some vacuum with it, and thus turns the plane of the pendulum. For the same reason, the Earth's rotation turns the plane of an orbiting satellite.

The Thirring–Lense or frame-dragging effect is extremely small. It might be that it was measured for the first time in 1998 by an Italian group led by Ignazio Ciufolini, and then again by the same group in the years up to 2004. The group followed the motion of two special artificial satellites – shown in Figure 76 – consisting only of a body of steel and some Cat's-eyes. The group measured the satellite's motion around the Earth with extremely high precision, making use of reflected laser pulses. This method allowed this experiment to be comparatively cheap and quick. Unfortunately, the size of the systematic effects and other reasons imply that the published results cannot be trusted. Ref. 167 Ref. 168

So far, only one other group tried the experiment around Earth. The satellite for the so-called Gravity Probe B experiment was put in orbit in 2005, after over 30 years of planning. These satellites were extremely involved and were carrying rapidly rotating superconducting spheres. Despite several broken systems, in 2009 the experiment confirmed the existence of frame dragging around Earth. The evaluation confirmed the predictions of general relativity within about 25 %. Ref. 169

In the meantime, frame dragging effects have also been measured in various other astronomical systems. The best confirmations have come from pulsars. Pulsars send out regular radio pulses, e.g. every millisecond, with extremely high precision. By measuring the exact times when the pulses arrive on Earth, one can deduce the details of the motion of these stars and confirm that such subtle effects as frame dragging do indeed take place. Ref. 170

Gravitomagnetism*

Frame-dragging, the geodesic effect and the Thirring effects can be seen as special cases of *gravitomagnetism*. (We will show the connection below.) This approach to gravity was already studied in the nineteenth century by Holzmüller and by Tisserand, long before general relativity was discovered. The approach has become popular again in recent years because it is simple to understand. As mentioned above, talking about a gravitational *field* is always an approximation. In the case of weak gravity, such as occurs in everyday life, the approximation is very good. Many relativistic effects can be described in terms of the Ref. 171

---

* This section can be skipped at first reading.

gravitational field, without using the concept of space curvature or the metric tensor. Instead of describing the complete space-time mattress, the gravitational-field model only describes the deviation of the mattress from the flat state, by pretending that the deviation is a separate entity, called the gravitational field. But what is the relativistically correct way to describe the gravitational field?

Vol. III, page 53

We can compare the situation to electromagnetism. In a relativistic description of electrodynamics, the electromagnetic field has an electric and a magnetic component. The electric field is responsible for the inverse-square Coulomb force. In the same way, in a relativistic description of (weak) gravity,* the gravitational field has an gravitoelectric and a gravitomagnetic component. The gravitoelectric field is responsible for the inverse square acceleration of gravity; what we call the gravitational field in everyday life is simply the gravitoelectric part of the full relativistic (weak) gravitational field.

Ref. 172, Ref. 173

What is the gravitomagnetic field? In electrodynamics, electric charge produces an electric field, and a *moving* charge, i.e., a *current*, produces a magnetic field. Similarly, in relativistic weak-field gravitation, mass–energy produces the gravitoelectric field, and *moving* mass–energy produces the gravitomagnetic field. In other words, frame-dragging is due to a gravitomagnetic effect and is due to mass *currents*.

Ref. 172

Vol. III, page 48

In the case of electromagnetism, the distinction between magnetic and electric field depends on the observer; each of the two can (partly) be transformed into the other. The same happens in the case of gravitation. Electromagnetism provides a good indication as to how the two types of gravitational fields behave; this intuition can be directly transferred to gravity. In electrodynamics, the motion $\boldsymbol{x}(t)$ of a charged particle is described by the Lorentz equation

$$m\ddot{\boldsymbol{x}} = q\boldsymbol{E} + q\dot{\boldsymbol{x}} \times \boldsymbol{B} \, , \qquad (166)$$

where the dot denotes the derivative with respect to time. In other words, the change of *speed* $\ddot{\boldsymbol{x}}$ is due to electric field $\boldsymbol{E}$, whereas the magnetic field $\boldsymbol{B}$ produces a velocity-dependent change of the *direction* of velocity, without changing the speed itself. Both changes depend on the value of the electric charge q. In the case of gravity this expression becomes

$$m\ddot{\boldsymbol{x}} = m\boldsymbol{G} + m\dot{\boldsymbol{x}} \times \boldsymbol{H} \, . \qquad (167)$$

The role of charge is taken by mass. The role of the electric field is taken by the gravitoelectric field $\boldsymbol{G}$ – which we simply call *gravitational field* in everyday life – and the role of the magnetic field is taken by the gravitomagnetic field $\boldsymbol{H}$. In this expression for the motion we already know the gravitoelectric field $\boldsymbol{G}$; it is given by

$$\boldsymbol{G} = \nabla\varphi = \nabla\frac{GM}{r} = -\frac{GM\boldsymbol{x}}{r^3} \, . \qquad (168)$$

As usual, the quantity φ is the (scalar) potential. The field $\boldsymbol{G}$ is the usual gravitational field of universal gravity, produced by every mass, and has the dimension of an acceleration. Masses are the sources of the gravitoelectric field. The gravitoelectric field obeys $\nabla\boldsymbol{G} = -4\pi G\rho$, where ρ is the mass density. A *static* field $\boldsymbol{G}$ has no vortices; it obeys $\nabla \times \boldsymbol{G} = 0$.

---

* The approximation requires low velocities, weak fields, and localized and stationary mass–energy distri-

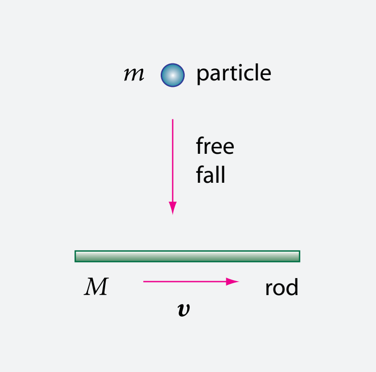


FIGURE 77 The reality of gravitomagnetism.

Ref. 174

Vol. III, page 53

It is not hard to show that if gravito*electric* fields exist, relativity requires that gravito*magnetic* fields must exist as well. The latter appear whenever we change from an observer at rest to a moving one. (We will use the same argument in electrodynamics.) A particle falling perpendicularly towards an infinitely long rod illustrates the point, as shown in Figure 77. An observer at rest with respect to the rod can describe the whole situation with gravitoelectric forces alone. A second observer, moving along the rod with constant speed, observes that the momentum of the particle *along the rod* also increases. This observer will thus not only measure a gravitoelectric field; he also measures a gravitomagnetic field. Indeed, a mass moving with velocity $\boldsymbol{v}$ produces a gravitomagnetic (3-) acceleration on a test mass m given by

$$m\boldsymbol{a} = m\boldsymbol{v} \times \boldsymbol{H} \tag{169}$$

Challenge 260 ny

where, *almost* as in electrodynamics, the static *gravitomagnetic field* $\boldsymbol{H}$ obeys

$$\boldsymbol{H} = 16\pi N\rho\boldsymbol{v} \tag{170}$$

where ρ is mass density of the source of the field and N is a proportionality constant. In nature, there are no sources for the gravitomagnetic field; it thus obeys $\nabla\boldsymbol{H} = 0$. The gravitomagnetic field has dimension of inverse time, like an angular velocity.

Challenge 261 ny

When the situation in Figure 77 is evaluated, we find that the proportionality constant N is given by

$$N = \frac{G}{c^2} = 7.4 \cdot 10^{-28}\,\mathrm{m/kg}\,, \tag{171}$$

Challenge 262 s

an extremely small value. We thus find that as in the electrodynamic case, the gravitomagnetic field is weaker than the gravitoelectric field by a factor of c^2. It is thus hard to observe. In addition, a second aspect renders the observation of gravitomagnetism even more difficult. In contrast to electromagnetism, in the case of gravity there is no way to observe *pure* gravitomagnetic fields (why?); they are always mixed with the usual, gravitoelectric ones. For these reasons, gravitomagnetic effects were measured for the first

---

butions.

time only in the 1990s. In other words, universal gravity is the weak-field approximation of general relativity that arises when all gravitomagnetic effects are neglected.

In summary, *if a mass moves, it also produces a gravitomagnetic field.* How can we imagine gravitomagnetism? Let's have a look at its effects. The experiment of Figure 77 showed that a moving rod has the effect to slightly accelerate a test mass in the same direction as its motion. In our metaphor of the vacuum as a mattress, it looks as if a moving rod drags the vacuum along with it, as well as any test mass that happens to be in that region. Gravitomagnetism appears as *vacuum dragging*. Because of a widespread reluctance to think of the vacuum as a mattress, the expression *frame dragging* is used instead.

In this description, *all frame dragging effects are gravitomagnetic effects*. In particular, a gravitomagnetic field also appears when a large mass *rotates*, as in the Thirring–Lense effect of Figure 75. For an angular momentum $\boldsymbol{J}$ the gravitomagnetic field $\boldsymbol{H}$ is a dipole field; it is given by

$$\boldsymbol{H} = \nabla \times \left(-2 \frac{\boldsymbol{J} \times \boldsymbol{x}}{r^3} \right) \qquad (172)$$

exactly as in the electrodynamic case. The gravitomagnetic field around a spinning mass has three main effects.

First of all, as in electromagnetism, a spinning test particle with angular momentum $\boldsymbol{S}$ feels a *torque* if it is near a large spinning mass with angular momentum $\boldsymbol{J}$. This torque $\boldsymbol{T}$ is given by

$$\boldsymbol{T} = \frac{\mathrm{d}\boldsymbol{S}}{\mathrm{d}t} = \tfrac{1}{2}\, \boldsymbol{S} \times \boldsymbol{H} \,. \qquad (173)$$

The torque leads to the mentioned *precession of gyroscopes* or *geodesic precession*. For the Earth, this effect is extremely small: at the North Pole, the precession has a conic angle of 0.6 milli-arcseconds and a rotation rate of the order of 10^{-10} times that of the Earth.

A second effect of gravitomagnetism is the following. Since for a torque we have $\boldsymbol{T} = \dot{\boldsymbol{\Omega}} \times \boldsymbol{S}$, the dipole field of a large rotating mass with angular momentum $\boldsymbol{J}$ has an effect on orbiting masses. An orbiting mass will experience *precession of its orbital plane*. Seen from infinity we get, for an orbit with semimajor axis a and eccentricity e,

Challenge 263 ny

$$\dot{\boldsymbol{\Omega}} = -\frac{\boldsymbol{H}}{2} = -\frac{G}{c^2}\frac{\boldsymbol{J}}{|\boldsymbol{x}|^3} + \frac{G}{c^2}\frac{3(\boldsymbol{J}\boldsymbol{x})\boldsymbol{x}}{|\boldsymbol{x}|^5} = \frac{G}{c^2}\frac{2\boldsymbol{J}}{a^3(1-e^2)^{3/2}} \qquad (174)$$

which is the prediction of Lense and Thirring.[*] The effect – analogous to spin–spin coupling in atoms – is extremely small, giving an angle change of only 8 ″ per orbit for a satellite near the surface of the Earth. This explains the difficulties and controversies around such Earth-bound experiments. As mentioned above, the effect is much larger in pulsar systems.

As a third effect of gravitomagnetism, not mentioned yet, a rotating mass leads to an additional *precession of the periastron*. This is a similar effect to the one produced by space curvature on orbiting masses even if the central body does not rotate. The rotation just reduces the precession due to space-time curvature. This effect has been fully confirmed

---

Challenge 264 ny

[*] A homogeneous spinning sphere has an angular momentum given by $J = \frac{2}{5} M\omega R^2$.

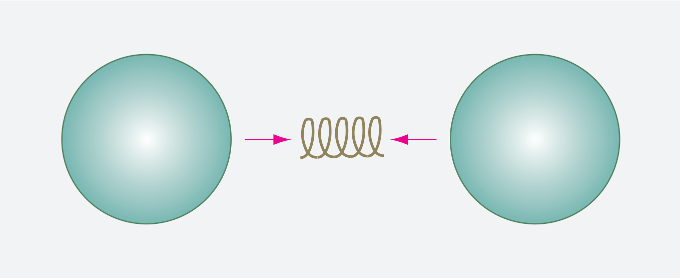


FIGURE 78 A Gedanken experiment showing the necessity of gravitational waves.

for the famous binary pulsar PSR 1913+16, discovered in 1974, as well as for the 'real' double pulsar PSR J0737-3039, discovered in 2003. This latter system shows a periastron precession of 16.9°/a, the largest value observed so far. Ref. 175

The split into gravitoelectric and gravitomagnetic effects is thus a useful approximation to the description of gravity. The split also helps to answer questions such as: How can gravity keep the Earth orbiting around the Sun, if gravity needs 8 minutes to get from the Sun to us? Challenge 265 s Above all, the split of the gravitational field into gravitoelectric and gravitomagnetic components allows a simple description of gravitational waves.

Gravitational waves

One of the most fantastic predictions of physics is the existence of gravitational waves. Gravity waves* prove that empty space itself has the ability to move and vibrate. The basic idea is simple. Since space is elastic, like a large mattress in which we live, space should be able to oscillate in the form of propagating waves, like a mattress or any other elastic medium.

Gravitational waves were predicted by Poincaré in 1905.** The waves were deduced from an approximation of general relativity by Einstein in 1916. For a certain time period, Ref. 176 Einstein – and many others – believed that his calculation was mistaken. He was convinced about the existence of gravitational waves only in 1937, when several people pointed out errors to him in his draft paper with Nathan Rosen on how to deduce waves from general relativity without any approximation. He then revised the manuscript. Therefore, only the paper published in 1937 showed unambiguously, for the first time, that gravitational waves exist in general relativity. A number of side issues had to be clarified even after this paper; in the 1950s the issue was definitively settled. Ref. 177

Starting from the existence of a maximum energy speed, Jørgen Kalckar and Ole Ulfbeck have given a simple argument for the necessity of gravitational waves. Ref. 178 They studied two equal masses falling towards each other under the effect of gravitational attraction, and imagined a spring between them. The situation is illustrated in Figure 78. Such a spring will make the masses bounce towards each other again and again. The central

---

* To be strict, the term 'gravity wave' has a special meaning: *gravity waves* are the surface waves of the sea, where gravity is the restoring force. However, in general relativity, the term is used interchangeably with 'gravitational wave'.

** In fact, the question of the speed of gravity was discussed long before him, by Laplace, for example. However, these discussions did not envisage the existence of waves.

TABLE 4 The predicted spectrum of gravitational waves.

Frequency	Wavelength	Name	Expected appearance
$< 10^{-4}$ Hz	> 3 Tm	extremely low frequencies	slow binary star systems, supermassive black holes
10^{-4} Hz–10^{-1} Hz	3 Tm–3 Gm	very low frequencies	fast binary star systems, massive black holes, white dwarf vibrations
10^{-1} Hz–10^{2} Hz	3 Gm–3 Mm	low frequencies	binary pulsars, medium and light black holes
10^{2} Hz–10^{5} Hz	3 Mm–3 km	medium frequencies	supernovae, pulsar vibrations
10^{5} Hz–10^{8} Hz	3 km–3 m	high frequencies	unknown; maybe future human-made sources
$> 10^{8}$ Hz	< 3 m		maybe unknown cosmological sources

spring stores the kinetic energy from the falling masses. The energy value can be measured by determining the length by which the spring is compressed. When the spring expands again and hurls the masses back into space, the gravitational attraction will gradually slow down the masses, until they again fall towards each other, thus starting the same cycle again.

However, the energy stored in the spring must get *smaller* with each cycle. Whenever a sphere detaches from the spring, it is decelerated by the gravitational pull of the other sphere. Now, the value of this deceleration depends on the distance to the other mass; but since there is a maximal propagation velocity, the effective deceleration is given by the distance the other mass *had* when its gravity effect started out towards the second mass. For two masses departing from each other, the effective distance is thus somewhat smaller than the actual distance. In short, while departing, the real deceleration is *larger* than the one calculated without taking the time delay into account.

Similarly, when one mass falls back towards the other, it is accelerated by the other mass according to the distance it had when the gravity effect started moving towards it. Therefore, while approaching, the acceleration is *smaller* than the one calculated without time delay.

Therefore, the masses arrive with a *smaller* energy than they departed with. At every bounce, the spring is compressed a little less. The difference between these two energies is lost by each mass: the energy is taken away by space-time. In other words, the energy difference is radiated away as gravitational radiation. The same thing happens with mattresses. Remember that a mass deforms the space around it as a metal ball on a mattress deforms the surface around it. (However, in contrast to actual mattresses, there is no friction between the ball and the mattress.) If two metal balls repeatedly bang against each other and then depart again, until they come back together, they will send out surface waves on the mattress. Over time, this effect will reduce the distance that the two balls depart from each other after each bang. As we will see shortly, a similar effect has already been measured; the two masses, instead of being repelled by a spring, were orbiting each other.

A simple mathematical description of gravity waves follows from the split into gravitomagnetic and gravitoelectric effects. It does not take much effort to extend gravitomagnetostatics and gravitoelectrostatics to *gravitodynamics*. Just as electrodynamics can be deduced from Coulomb's attraction by boosting to all possible inertial observers, gravitodynamics can be deduced from universal gravity by boosting to other observers. One gets the four equations

Ref. 179

Challenge 266 ny

$$\begin{aligned} \nabla\cdot\boldsymbol{G} = -4\pi G\rho \quad , \quad \nabla\times\boldsymbol{G} = -\frac{1}{4}\frac{\partial\boldsymbol{H}}{\partial t} \\ \nabla\cdot\boldsymbol{H} = 0 \quad , \quad \nabla\times\boldsymbol{H} = -16\pi N\rho\boldsymbol{v} + 4\frac{N}{G}\frac{\partial\boldsymbol{G}}{\partial t} \ . \end{aligned} \qquad (175)$$

We have met two of these equations already. The two other equations are expanded versions of what we have encountered, taking time-dependence into account. Except for the various factors of 4, the equations for gravitodynamics are the same as Maxwell's equations for electrodynamics. The additional factors of 4 appear because the ratio between angular momentum L and energy E of gravity waves is different from that of electromagnetic waves. The ratio determines the *spin* of a wave. For gravity waves

$$L = \frac{2}{\omega}E \ , \qquad (176)$$

whereas for electromagnetic waves the factor is $1/\omega$. It is worth recalling that the spin of radiation is a *classical* property. The *spin of a wave* is defined as the ratio $E/L\omega$, where E is the energy, L the angular momentum, and ω is the angular frequency. For electromagnetic waves, the spin is equal to 1; for gravitational waves, it is 2.

The spin is, of course, also a property of the – so far undetected – quantum particle that makes up gravitational waves. Interestingly, since gravity is universal, there can exist only a *single* kind of spin 2 radiation particle in nature. This is in strong contrast to the spin 1 case, of which there are *several* examples in nature: photons, weak bosons and gluons.

The equations of gravitodynamics must be complemented by the definition of the fields through the acceleration they produce:

$$m\ddot{\boldsymbol{x}} = m\boldsymbol{G} + m\dot{\boldsymbol{x}}\times\boldsymbol{H} \ . \qquad (177)$$

Definitions with different numerical factors are also common and then lead to different numerical factors in the equations of gravitodynamics.

The equations of gravitodynamics have a simple property: in vacuum, we can deduce from them a *wave equation* for the gravitoelectric and the gravitomagnetic fields $\boldsymbol{G}$ and $\boldsymbol{H}$. (It is not hard: try!) In other words, *gravity can behave like a wave: gravity can radiate.* All this follows from the expression of universal gravity when applied to moving observers, with the requirement that neither observers nor energy can move faster than c. Both the above argument involving the spring and the present mathematical argument use the

Challenge 267 e

same assumptions and arrive at the same conclusion.

Challenge 268 e

A few manipulations show that the speed of gravitational waves is given by

$$c = \sqrt{\frac{G}{N}} \ . \tag{178}$$

Vol. III, page 106

This result corresponds to the electromagnetic expression

$$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \ . \tag{179}$$

The same letter has been used for the two speeds, as they are identical. Both influences travel with the speed common to all energy with vanishing rest mass. We note that this is, strictly speaking, a prediction: the value of the speed of gravitational waves has been confirmed directly, despite claims to the contrary, only in 2016. Ref. 180

Ref. 181

How should we imagine gravitational waves? We sloppily said above that a gravitational wave corresponds to a surface wave of a mattress; now we have to do better and imagine that we live *inside* the mattress. Gravitational waves are thus moving and oscillating deformations of the mattress, i.e., of space. Like (certain) mattress waves, it turns out that gravity waves are *transverse*. Thus they can be polarized. In fact, gravity waves can be polarized in two ways. The effects of a gravitational wave are shown in Figure 79, for both linear and circular polarization.* We note that the waves are invariant under a rotation by π and that the two linear polarizations differ by an angle π/4; this shows that the particles corresponding to the waves, the gravitons, are of spin 2. (In general, the classical radiation field for a spin S particle is invariant under a rotation by $2\pi/S$. In

---

[*] A (small amplitude) plane gravity wave travelling in the z-direction is described by a metric g given by

$$g = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 + h_{xx} & h_{xy} & 0 \\ 0 & h_{xy} & -1 + h_{xx} & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{180}$$

where its two components, whose amplitude ratio determine the polarization, are given by

$$h_{ab} = B_{ab} \sin(kz - \omega t + \varphi_{ab}) \tag{181}$$

as in all plane harmonic waves. The amplitudes B_{ab}, the frequency ω and the phase φ are determined by the specific physical system. The general dispersion relation for the wave number k resulting from the wave equation is

$$\frac{\omega}{k} = c \tag{182}$$

and shows that the waves move with the speed of light.

In another gauge, a plane wave can be written as

$$g = \begin{pmatrix} c^2(1 + 2\varphi) & A_1 & A_2 & A_3 \\ A_1 & -1 + 2\varphi & h_{xy} & 0 \\ A_2 & h_{xy} & -1 + h_{xx} & 0 \\ A_3 & 0 & 0 & -1 \end{pmatrix} \tag{183}$$

where φ and $\boldsymbol{A}$ are the potentials such that $\boldsymbol{G} = \nabla\varphi - \frac{\partial \boldsymbol{A}}{c\partial t}$ and $\boldsymbol{H} = \nabla \times \boldsymbol{A}$.

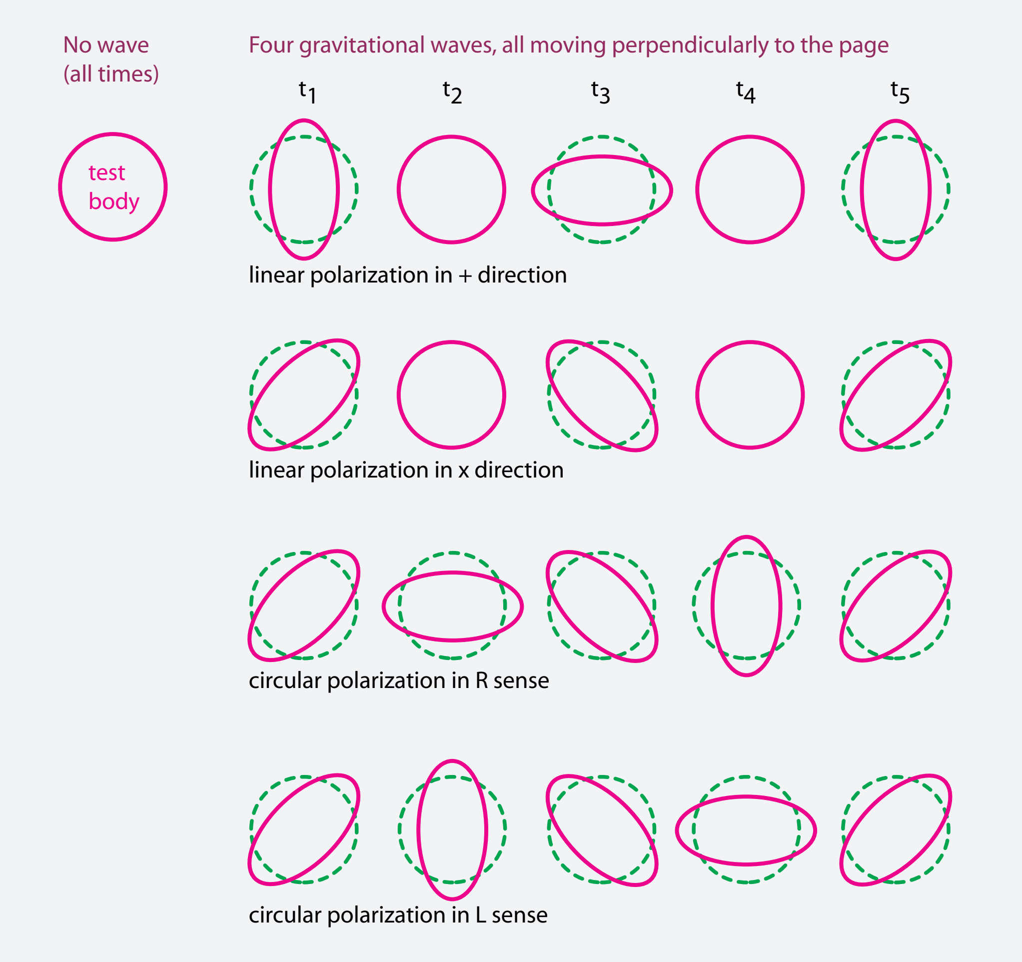


FIGURE 79 Effects on a circular or spherical body due to a plane gravitational wave moving in a direction perpendicular to the page.

addition, the two orthogonal linear polarizations of a spin S particle form an angle $\pi/2S$. For the photon, for example, the spin is 1; indeed, its invariant rotation angle is 2π and the angle formed by the two polarizations is $\pi/2$.)

If we image empty space as a mattress that *fills* space, gravitational waves are wobbling deformations of the mattress. More precisely, Figure 79 shows that a wave of circular polarization has the same properties as a corkscrew advancing through the mattress. We will discover later on why the analogy between a corkscrew and a gravity wave with circular polarization works so well. Indeed, in the last part of our adventure we will find a specific model of the space-time mattress that automatically incorporates corkscrew waves (instead of the spin 1 waves shown by ordinary latex mattresses). Vol. VI, page 298

PRODUCTION AND DETECTION OF GRAVITATIONAL WAVES

How does one produce gravitational waves? Obviously, masses must be accelerated. But how exactly? The conservation of energy forbids mass monopoles from varying in strength. We also know from universal gravity that a spherical mass whose radius oscillates would not emit gravitational waves. In addition, the conservation of momentum forbids mass dipoles from changing.

Challenge 269 ny

As a result, only *changing quadrupoles* can emit gravitational waves.* For example, two masses in orbit around each other will emit gravitational waves. Also, any rotating object that is not cylindrically symmetric around its rotation axis will do so. As a result, rotating an arm leads to gravitational wave emission. Most of these statements also apply to masses in mattresses. Can you point out the differences?

Challenge 270 ny

Einstein found that the amplitude h of waves at a distance r from a source is given, to a good approximation, by the second derivative of the retarded quadrupole moment Q:

Ref. 182

$$h_{ab} = \frac{2G}{c^4}\frac{1}{r}\mathrm{d}_{tt}Q^{\mathrm{ret}}_{ab} = \frac{2G}{c^4}\frac{1}{r}\mathrm{d}_{tt}Q_{ab}(t - r/c)\ . \tag{184}$$

This expression shows that the amplitude of gravity waves *decreases only with* $1/r$, in contrast to naive expectations. This feature is the same as for electromagnetic waves. In addition, the small value of the prefactor, $1.6 \cdot 10^{-44}$ Wm/s, shows that truly gigantic systems are needed to produce quadrupole moment changes that yield any detectable length variations in bodies. To be convinced, just insert a few numbers, keeping in mind that the best present detectors are able to measure length changes down to $h = \delta l/l = 10^{-21}$. The production by humans of detectable gravitational waves is probably impossible.

Challenge 271 e

Gravitational waves, like all other waves, transport energy.** If we apply the general formula for the emitted power P to the case of two masses m_1 and m_2 in circular orbits around each other at distance l and get

Ref. 129

$$P = -\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{G}{45c^5}\,\dddot{Q}^{\mathrm{ret}}_{ab}\,\dddot{Q}^{\mathrm{ret}}_{ab} = \frac{32}{5}\frac{G}{c^5}\left(\frac{m_1 m_2}{m_1 + m_2}\right)^2 l^4\omega^6 \tag{185}$$

which, using Kepler's relation $4\pi^2 r^3/T^2 = G(m_1 + m_2)$, becomes

$$P = \frac{32}{5}\frac{G^4}{c^5}\frac{(m_1 m_2)^2(m_1 + m_2)}{l^5}\ . \tag{186}$$

For elliptical orbits, the rate increases with the ellipticity, as explained in the text by Goenner. Inserting the values for the case of the Earth and the Sun, we get a power of about 200 W, and a value of 400 W for the Jupiter–Sun system. These values are so small

Ref. 129

---

* A *quadrupole* is a symmetrical arrangement, on the *four* sides of a square, of four alternating poles. In gravitation, a monopole is a point-like or spherical mass, and, since masses cannot be negative, a quadrupole is formed by *two* monopoles. A flattened sphere, such as the Earth, can be approximated by the sum of a monopole and a quadrupole. The same is valid for an elongated sphere.

** Gravitoelectromagnetism allows defining the gravitational Poynting vector. It is as easy to define and use as in the case of electrodynamics.

Vol. III, page 89
Ref. 174

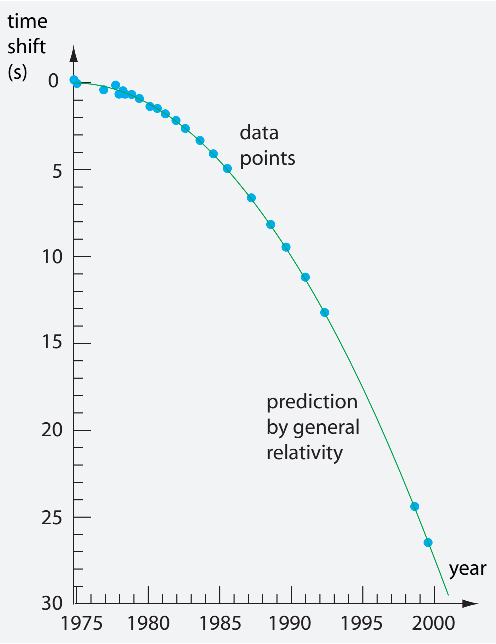


FIGURE 80 Comparison between measured time delay for the periastron of the binary pulsar PSR 1913+16 and the prediction due to energy loss by gravitational radiation.

that their effect cannot be detected at all.

For all orbiting systems, the frequency of the waves is twice the orbital frequency, as you might want to check. These low frequencies make it even more difficult to detect them. Challenge 272 ny

As a result of the usually low power of gravitational wave emission, the first observation of their effects was in binary pulsars. Pulsars are small but extremely dense stars; even with a mass equal to that of the Sun, their diameter is only about 10 km. Therefore they can orbit each other at small distances and high speeds. Indeed, in the most famous binary pulsar system, PSR 1913+16, the two stars orbit each other in an amazing 7.8 h, even though their semimajor axis is about 700 Mm, just less than twice the Earth–Moon distance. Since their orbital speed is up to 400 km/s, the system is noticeably relativistic.

Pulsars have a useful property: because of their rotation, they emit extremely regular radio pulses (hence their name), often in millisecond periods. Therefore it is easy to follow their orbit by measuring the change of pulse arrival time. In a famous experiment, a team of astrophysicists led by Joseph Taylor* measured the speed decrease of the binary pulsar system just mentioned. Eliminating all other effects and collecting data for 20 years, they found a decrease in the orbital frequency, shown in Figure 80. The slowdown is due to gravity wave emission. The results exactly fit the prediction by general relativity, *without any adjustable parameter*. (You might want to check that the effect must be quadratic in time.) This was the first case in which general relativity was tested up to $(v/c)^5$ precision. To get an idea of the precision, consider that this experiment detected a reduction of the orbital diameter of 3.1 mm per orbit, or 3.5 m per year! The measurements were possible only because the two stars in this system are neutron stars with small size, Ref. 183 Ref. 184 Challenge 273 ny Page 156 Ref. 183

---

* In 1993 he shared the Nobel Prize in Physics for his life's work.

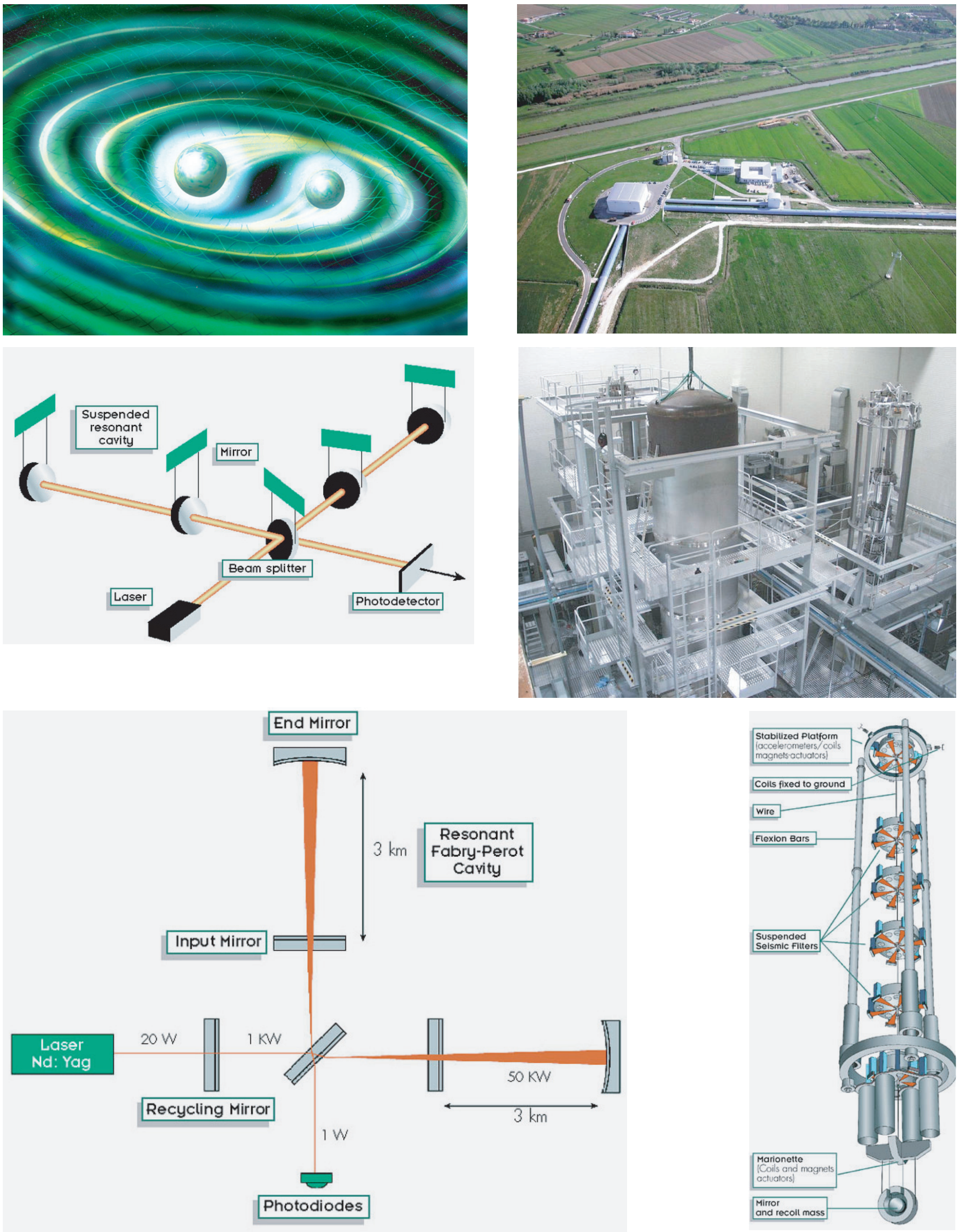


FIGURE 81 Detection of gravitational waves: an illustration of the merger of two black holes emitting such waves (top left). The other images show the VIRGO detector in Cascina, Italy, with one of its huge mirror suspensions, the mirror suspension details, and two drawings of the laser interferometer (© INFN).

large velocities and purely gravitational interactions. The pulsar rotation period around its axis, about 59 ms, is known to eleven digits of precision, the orbital time of 7.8 h is known to ten digits and the eccentricity of the orbit to six digits. Radio astronomy can Ref. 129

be spectacular.

The *direct* detection of gravitational waves was one of the long-term aims of experimental general relativity. The race has been on since the 1990s. The basic idea is simple, as shown in Figure 81: take four bodies, usually four mirrors, for which the line connecting one pair is perpendicular to the line connecting the other pair. Then measure the distance changes of each pair. If a gravitational wave comes by, one pair will increase in distance and the other will decrease, at the *same* time.

Since detectable gravitational waves cannot be produced by humans, wave detection first of all requires the patience to wait for a strong enough wave to come by. It turns out that even for a body around a black hole, only about 6 % of the rest mass can be radiated away as gravitational waves; furthermore, most of the energy is radiated during the final fall into the black hole, so that only quite violent processes, such as neutron star collisions or black hole mergers, are good candidates for detectable gravity wave sources. The waves produced by a black hole merger are shown in Figure 81.

In addition, a measurement system able to detect length changes of the order of 10^{-22} or better is needed – in other words, a lot of money. For mirrors spaced 4 km apart, the detectable distance change must be less than one thousandth of the diameter of a proton. Essential for a successful detection are the techniques to *eliminate noise* in the detection signal. Since decades, worlds's best noise reduction experts are all working on gravitational wave detectors. Understanding the noise mechanisms has become a research filed in its own.

Until 2015, gravitational waves had not been detected. The sensitivity of the detectors was not sufficient. In fact, the race to increase the sensitivity is still ongoing across the world. After over twenty years of constant improvements, finally, in 2016, a signal with a duration of 0.2 s – shown in Figure 82 – was published: it corresponds precisely to the signal expected from the merger of two black holes of 29 and 36 solar masses. The result of the merger is a black hole of 62 solar masses, and the 3 lost solar masses were radiated away, in large part as gravitational waves. This happened between 600 and 1800 million light-years away. The clarity of the signal, measured at two different locations, convinced everybody of the correctness of the interpretation. The astonishingly small peak length variation $\Delta l/l$ of below 10^{-21} remains a fascinating experimental feat, even when the large financial budget is taken into account. Several additional merger events have been measured after the first one. Ref. 112

Challenge 274 r Gravitational waves are a fascinating topic. Can you find a cheap method to measure their speed? A few astrophysical experiments had deduced bounds on the mass of the graviton before, and had confirmed the speed of gravity in an indirect way. The first direct measurement was the discovery of 2016; the result is the speed of light, within measurement precision. The observation of a candidate light flash that accompanied the black hole merger would, if confirmed in this or in a future observation, show that gravitational waves travel with the same speed as light waves to within one part in 10^{16}. Ref. 112

Another question on gravitational waves remains open at this point: If all change in nature is due to motion of particles, as the Greeks maintained, how do gravity waves fit into the picture? Quantum theory requires that gravitational waves must be made of particles. (These hypothetical particles are called *gravitons*.) Now, there is no real difference between empty space at rest and wobbling empty space. If gravitational waves were made of particles, space-time would also have to be! How can this be the case? We have

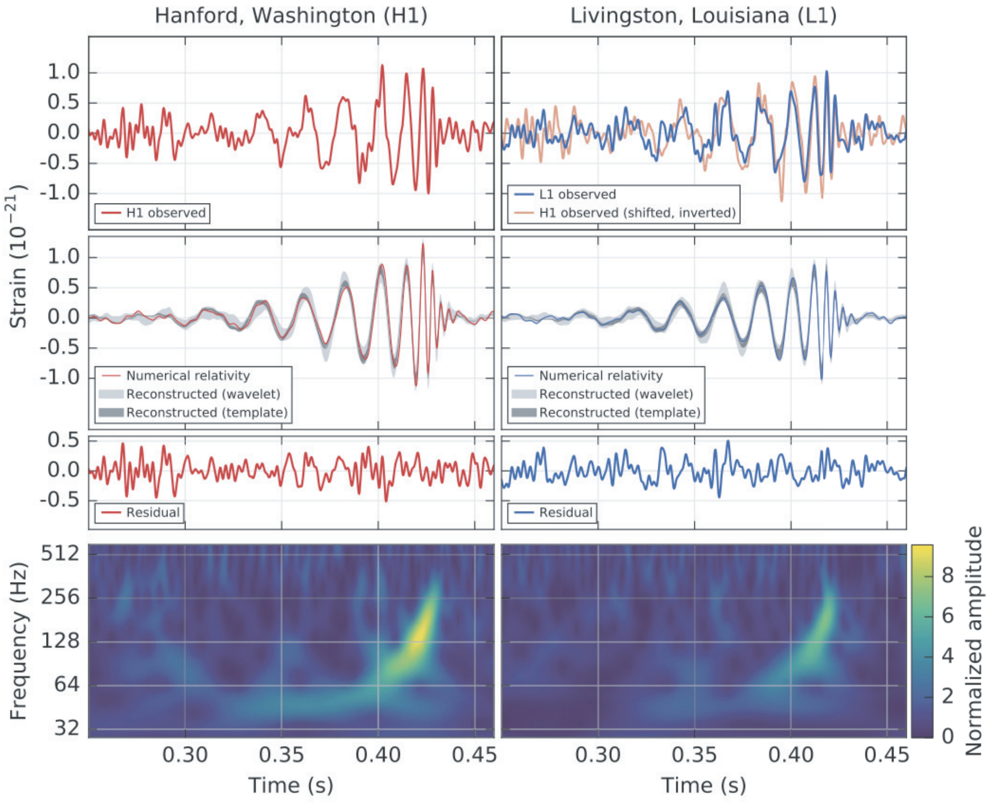


FIGURE 82 The first direct detection of gravitational waves through deformation of space, with a strain of the order of 10^{-21}, by two detectors spaced three thousand kilometres apart (© LIGO/Physical Review Letters).

to wait until the final part of our adventure to say more.

Curiosities and fun challenges about weak fields

Challenge 275 s Is there a static gravitational field that oscillates in space?

* *

If we explore the options for the speed of gravitational waves, an interesting connection appears. If the speed of gravitational waves were *smaller* than the speed of light, moving bodies that move almost as rapidly as the speed of light, like cosmic ray particles, would be slowed down by emitting *Vavilov–Čerenkov radiation*, until they reach the lower speed. This is not observed. Page 28

If on the other hand, the speed of gravitational waves were *larger* than that of light, the waves would not obey causality or the second principle of thermodynamics. In short, gravitational waves, if they exist, must propagate with the speed of light. (A speed very near to the speed of light might also be possible.)

* *

One effect that disturbs gravitational wave detectors are the tides. On the GEO600 detector in Hannover, tides change the distance of the mirrors, around 600 m, by 2 μm.

∗ ∗

Challenge 276 ny

Are narrow beams of gravitational waves, analogous to beams of light, possible? Would two parallel beams of gravitational waves attract each other?

∗ ∗

Challenge 277 e

As predicted in earlier editions of this book, the discovery of gravitational waves was announced in television and radio. Does the discovery help to improve the quality of life across the planet? Except for a number of scientists, other humans will almost surely not benefit at all. This situation is in stark contrast to scientific discoveries made in the twentieth century. What is the reason for this contrast?

∗ ∗

Ref. 185
Challenge 278 e

Can gravity waves be used to power a rocket? Yes, maintain Bonnor and Piper. You might ponder the possibility yourself.

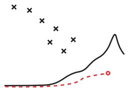
∗ ∗

Electromagnetism and gravity differ in one aspect: two equal charges repel, two equal masses attract. In more elaborate terms: for the exchange of spin 2 particles – gravitons – the effect of mass can be depicted with the mattress model. This is possible because the sign of the effect in the mattress is independent of other masses. In contrast, for electromagnetism, the sign of the potential depends on the other electric charges.

A SUMMARY ON ORBITS AND WAVES

In summary, the curvature of space and space-time implies:
- in contrast to universal gravity, masses deflect light more strongly;
- in contrast to universal gravity, light is effectively slowed down near masses;
- in contrast to universal gravity, elliptical orbits are not closed;
- in contrast to universal gravity, orbiting objects change their orientation in space;
- in contrast to universal gravity, empty vacuum can propagate gravitational waves that travel with the speed of light.

All experiments ever performed confirm these conclusions and verify the numerical predictions within measurement precision. Both the numerous experiments in weak gravitational fields and the less common experiments in strong fields fully confirm general relativity. All experiments also confirm the force and power limits.



Chapter 7

FROM CURVATURE TO MOTION

In the precise description of gravity, motion depends on space-time curvature. In order to quantify this idea, we first of all need to accurately describe curvature itself. To clarify the issue, we will start the discussion in two dimensions, and then move to three and four dimensions. Once we are able to explore curvature, we explore the precise relation between curvature and motion.

How to measure curvature in two dimensions

Obviously, a flat sheet of paper has no curvature. If we roll it into a cone or a cylinder, it gets what is called *extrinsic curvature*; however, the sheet of paper still looks flat for any two-dimensional animal living on it – as approximated by an ant walking over it. In other words, the *intrinsic curvature* of the sheet of paper is zero even if the sheet as a whole is extrinsically curved.

Intrinsic curvature is thus the stronger concept, measuring the curvature which can be observed even by an ant. We note that all intrinsically curved surfaces are also extrinsically curved. The surface of the Earth, the surface of an island, or the slopes of a mountain** are intrinsically curved. Whenever we talk about curvature in general relativity, we always mean *intrinsic* curvature, since any observer in nature is by definition in the same situation as an ant on a surface: their experience, their actions and plans always only concern their closest neighbourhood in space and time.

But how can an ant determine whether it lives on an intrinsically curved surface?*** One way is shown in Figure 83. The ant can check whether either the circumference of a circle bears a Euclidean relation to the measured radius. She can even use the difference between the measured and the Euclidean values as a measure for the local intrinsic curvature, if she takes the limit for vanishingly small circles and if she normalizes the values correctly. In other words, the ant can imagine to cut out a little disc around the point she is on, to iron it flat and to check whether the disc would tear or produce folds. Any two-dimensional surface is intrinsically curved whenever ironing is not able to make a flat street map out of it. The 'density' of folds or tears is related to the curvature. Folds imply negative intrinsic curvature, tears positive curvature.

Challenge 279 e

** Unless the mountain has the shape of a perfect cone. Can you confirm this?

*** Note that the answer to this question also tells us how to distinguish real curvature from curved coordinate systems on a flat space. This question is often asked by those approaching general relativity for the first time.

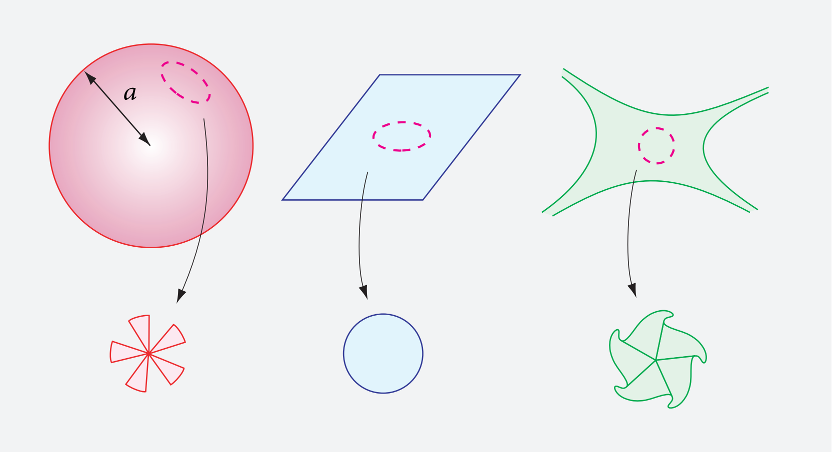


FIGURE 83 Positive, vanishing and negative curvature in two dimensions.

Challenge 280 s

Check your understanding: Can a one-dimensional space have intrinsic curvature? Is a torus intrinsically curved?

Page 188

Alternatively, we can recognize intrinsic curvature also by checking whether two parallel lines that are locally straight stay parallel, approach each other, or depart from each other. On a paper cylinder, parallel lines remain parallel; in this case, the surface is said to have *vanishing* intrinsic curvature. A surface with *approaching* parallels, such as the Earth, is said to have *positive* intrinsic curvature, and a surface with *diverging* parallels, such as a saddle, is said to have *negative* intrinsic curvature. Speaking simply, positive curvature means that we are more restricted in our movements, negative that we are less restricted. A *constant* curvature even implies being locked in a finite space. You might want to check this with Figure 83 and Figure 85. We can even measure instrinsic curvature by determining how rapidly to parallel lines depart or converge.

A third way to measure intrinsic curvature of surfaces uses triangles. On curved surfaces the sum of angles in a triangle is larger than π, i.e., larger than two right angles, for positive curvature, and smaller than π for negative curvature.

Ref. 186

Challenge 281 e

Let us see in detail how we can *quantify* and measure the curvature of surfaces. First a question of vocabulary: a sphere with radius a is said, by definition, to have an intrinsic curvature $K = 1/a^2$. Therefore a plane has zero curvature. You might check that for a circle on a sphere, the measured radius r, circumference C, and area A are related by

$$C = 2\pi r\left(1 - \frac{K}{6}r^2 + ...\right) \quad \text{and} \quad A = \pi r^2\left(1 - \frac{K}{12}r^2 + ...\right) \qquad (187)$$

where the dots imply higher-order terms. This allows us to define the intrinsic curvature K, also called the *Gaussian* curvature, for a general point on a two-dimensional surface in either of the following two equivalent ways:

$$K = 6\lim_{r\to 0}\left(1 - \frac{C}{2\pi r}\right)\frac{1}{r^2} \quad \text{or} \quad K = 12\lim_{r\to 0}\left(1 - \frac{A}{\pi r^2}\right)\frac{1}{r^2}\,. \qquad (188)$$

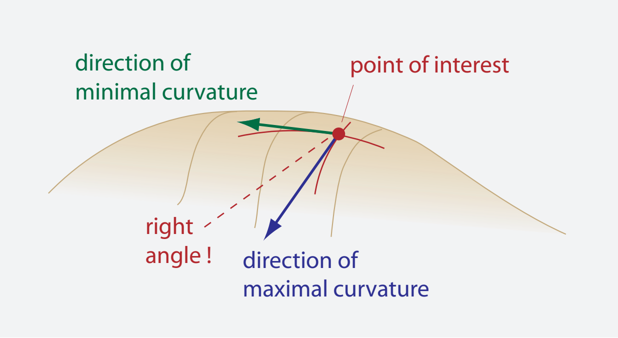


FIGURE 84 The maximum and minimum curvature of a surface are always at a right angle to each other.

These expressions allow an ant to measure the intrinsic curvature at each point for any smooth surface.*

From now on in this text, *curvature* will always mean *intrinsic* curvature, i.e., Gaussian curvature and its higher-dimensional analogs. Like an ant on a surface, also an observer in space can only detect intrinsic curvature. Therefore, only intrinsic curvature is of interest in the description of nature.

We note that the curvature of a surface can be different from place to place, and that it can be positive, as for an egg, or negative, as for the part of a torus nearest to the hole. A saddle is another example of negative curvature, but, unlike the torus, its curvature changes along all directions. In fact, it is not possible at all to fit a two-dimensional surface of *constant* negative curvature inside three-dimensional space; we need at least four dimensions to do so, as you can find out if you try to imagine the situation. Challenge 283 e

For any surface, at *every* point, the direction of maximum curvature and the direction of minimum curvature are *perpendicular* to each other. This relationship, shown in Figure 84, was discovered by Leonhard Euler in the eighteenth century. You might want to check this with a tea cup, with a sculpture by Henry Moore, or with any other curved object from your surroundings, such as a Volkswagen Beetle. Challenge 284 e The Gaussian curvature K defined in (188) is in fact the product of the two corresponding inverse curvature radii. Thus, even though *line* curvature is *not* an intrinsic property, the Gaussian curvature is.

The Gaussian curvature is an *intrinsic* property of a surface at each point. This means, as just explained, that bending the surface does not change its value at each point. For example, a flat sheet of paper, a paper rolled up into a cylinder and a folded paper all have zero intrinsic curvature. Because the intrinsic, Gaussian curvature of a flat sheet is zero, for every bent sheet, at every point, there is always a line with zero curvature. Bent sheets are made up of straight lines. This property follows from the shape-independence of the Gaussian curvature. The property makes *bent* sheets – but not flat sheets – stiff against bending attempts that try to bend the straight line. This property is the reason

---

* If the n-dimensional volume of a sphere is written as $V_n = C_n r^n$ and its $(n-1)$-dimensional ‘surface’ as $O_n = nC_n r^{n-1}$, we can generalize the expressions for curvature to Ref. 187

$$K = 3(n+2)\lim_{r\to 0}\left(1-\frac{V_n}{C_n r^n}\right)\frac{1}{r^2} \quad \text{or} \quad K = 3n\lim_{r\to 0}\left(1-\frac{O_n}{nC_n r^{n-1}}\right)\frac{1}{r^2}\,, \qquad (189)$$

Challenge 282 ny as shown by Vermeil. A famous riddle is to determine the number C_n.

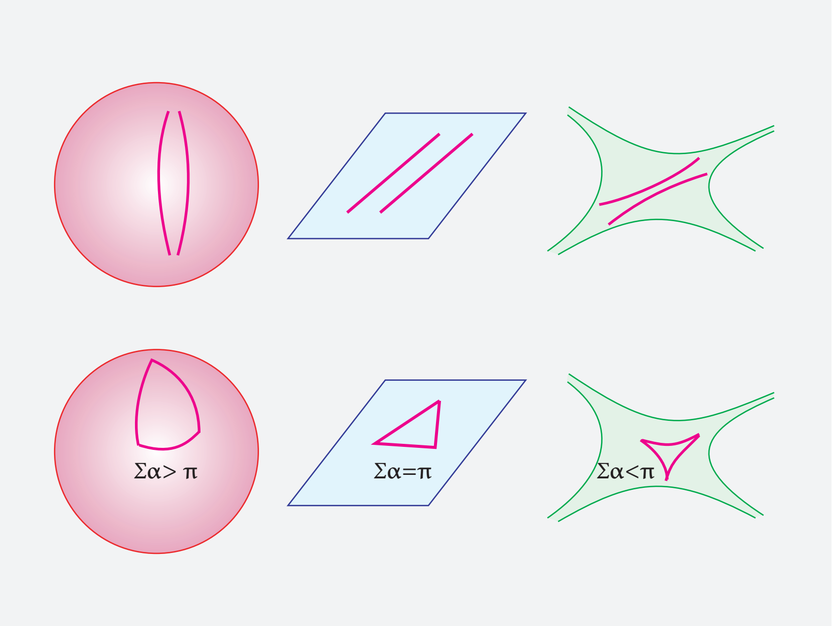


FIGURE 85 Positive, vanishing and negative curvature (in two dimensions) illustrated with the corresponding geodesic behaviour and the sum of angles in a triangle.

that straight tubes, cones and folded paper are particularly stiff and light structures. For the same reason, the best way to hold a pizza slice is to fold it along the central radius. In this case, intrinsic curvature prevents that the tip bends down.

Also roofs in the shape of a circular hyperboloid or of a hyperbolic paraboloid are stiff and have two straight lines through every point on their surface. Are these surfaces made of a bent flat sheet? Challenge 285 s

In summary, Gaussian curvature is a measure of the intrinsic curvature of two-dimensional surfaces. Such an intrinsic measure of curvature is needed if we are forced to stay and move inside the surface or inside the space that we are exploring. Because this applies to all humans, physicists are particularly interested in intrinsic curvature, though for more than two dimensions.

Three dimensions: curvature of space

For *three*-dimensional space, describing intrinsic curvature is a bit more involved. To start with, we have difficulties imagining the situation, because we usually associate curvature with extrinsic curvature. In fact, the only way to explore three-dimensional curvature of space is to think like the ant on a surface, and to concentrate on intrinsic curvature. Therefore we will describe three-dimensional curvature using two-dimensional curvature.

In curved three-dimensional space, the Gaussian curvature of an arbitrary, small two-dimensional disc around a general point will depend on the orientation of the disc. Let us first look at the simplest case. If the Gaussian curvature at a point is the same for all orientations of the disc, the point is called *isotropic*. We can imagine a small sphere around that point. In this special case, in three dimensions, the relation between the

Challenge 286 ny

measured radius r and the measured surface area A and volume V of the sphere lead to

$$A = 4\pi r^2 \left(1 - \frac{K}{3} r^2 + ...\right) \quad \text{and} \quad V = \frac{4\pi}{3} r^3 \left(1 - \frac{K}{5} r^2 + ...\right) , \qquad (190)$$

where K is the curvature for an isotropic point. This leads to

$$K = 3 \lim_{r \to 0} \left(1 - \frac{A}{4\pi r^2}\right) \frac{1}{r^2} = 6 \lim_{r \to 0} \frac{r - \sqrt{A/4\pi}}{r^3} = 6 \lim_{r \to 0} \frac{r_{\text{excess}}}{r^3} , \qquad (191)$$

where we defined the *excess radius* as $r_{\text{excess}} = r - \sqrt{A/4\pi}$. We thus find that

▷ For a three-dimensional space, *the average curvature is six times the excess radius of a small sphere divided by the cube of the radius.*

A positive curvature is equivalent to a positive excess radius, and similarly for vanishing and negative cases. The *average* curvature at a point is the curvature calculated by applying the definition with a small sphere to an arbitrary, non-isotropic point.

For a non-isotropic point in three-dimensional space, the Gaussian curvature value determined with a two-dimensional disc will depend on the *orientation* of the disc. In fact, there is a relationship between all possible disc curvatures at a given point; taken together, they must form a tensor. (Why?) In other words, the Gaussian curvature values define an ellipsoid at each point. For a full description of curvature, we thus have to specify, as for any tensor in three dimensions, the main Gaussian curvature values in three orthogonal directions, corresponding to the thee main axes of the ellipsoid.*

Challenge 287 ny

What are the curvature values for the three-dimensional space around us? Already in 1827, the mathematician and physicist Carl-Friedrich Gauß** is said to have checked whether the three angles formed by three mountain peaks near his place of residence added up to π. Nowadays we know that the deviation δ from the angle π on the surface

---

* These three disc values are not independent however, since together, they must yield the just-mentioned average volume curvature K. In total, there are thus *three* independent scalars describing the curvature in three dimensions (at each point). Using the metric tensor g_{ab} and the Ricci tensor R_{ab} to be introduced below, one possibility is to take for the three independent numbers the values $R = -2K$, $R_{ab}R^{ab}$ and $\det R/\det g$.

** Carl-Friedrich Gauß (b. 1777 Braunschweig, d. 1855 Göttingen), together with Leonhard Euler, was the most important mathematician of all times. (His name is written 'Gauss' in English texts.) A famous child prodigy, when he was 19 years old, he constructed the regular heptadecagon with compass and ruler (see www.mathworld.wolfram.com/Heptadecagon.html). He was so proud of this result that he put a drawing of the figure on his tomb. Gauss produced many results in number theory, topology, statistics, algebra, complex numbers and differential geometry which are part of modern mathematics and bear his name. Among his many accomplishments, he produced a theory of curvature and developed non-Euclidean geometry. He also worked on electromagnetism and astronomy.

Gauss was a difficult character, worked always for himself, and did not found a school. He published little, as his motto was: pauca sed matura. As a consequence, when another mathematician published a new result, he regularly produced a notebook in which he had noted the very same result already years before. These famous notebooks are now available online at www.sub.uni-goettingen.de.

of a body of mass M and radius r is given by

$$\delta = \pi - (\alpha + \beta + \gamma) \approx -A_{\text{triangle}} K = A_{\text{triangle}} \frac{GM}{r^3 c^2} . \tag{192}$$

This expression is typical for hyperbolic geometries. For the case of mathematical negative curvature K, the first equality was deduced by Johann Lambert.* The last equation came only one and a half century later, and is due to Einstein, who made clear that the negative curvature K of the space around us is related to the mass and gravitation of a body. For the case of the Earth and typical mountain distances, the angle δ is of the order of 10^{-14} rad. Gauss had no chance to detect any deviation, and in fact he detected none. Even today, studies with lasers and high-precision apparatus have detected no deviation yet – on Earth. The proportionality factor that determines the curvature of space-time on the surface of the Earth is simply too small. But Gauss did not know, as we do today, that gravity and curvature go hand in hand.

CURVATURE IN SPACE-TIME

> “Notre tête est ronde pour permettre à la pensée de changer de direction.**”
>
> Francis Picabia

In nature, with *four* space-time dimensions, specifying curvature requires a more involved approach. First of all, the use of space-time coordinates automatically introduces the speed of light c as limit speed. Furthermore, the number of dimensions being four, we expect several types of curvature: We expect a value for an average curvature at a point, defined by comparing the 4-volume of a 4-sphere in space-time with the one deduced from the measured radius; then we expect a set of ‘almost average’ curvatures defined by 3-volumes of 3-spheres in various orientations, plus a set of ‘low-level’ curvatures defined by usual 2-areas of usual 2-discs in even more orientations. Obviously, we need to bring some order to bear on this set.

Fortunately, physics can help to make the mathematics easier. We start by defining what we mean by curvature in space-time. To achieve this, we use the definition of curvature of Figure 85. As shown in the figure, the curvature K also describes how geodesics *diverge* or *converge*.

Geodesics are the straightest paths on a surface, i.e., those paths that a tiny car or tricycle would follow if it drove on the surface keeping the steering wheel straight. Locally, nearby geodesics are parallel lines. If two nearby geodesics are in a curved space, their separation s will change along the geodesics. This happens as

Challenge 288 e

$$\frac{d^2 s}{dl^2} = -Ks + \text{higher orders} \tag{193}$$

---

* Johann Lambert (1728–1777), Swiss mathematician, physicist and philosopher. Among many achievements, he proved the irrationality of π; also several laws of optics are named after him.

** ‘Our head is round in order to allow our thougths to change direction.’ Francis Picabia (b. 1879 Paris, d. 1953 Paris) dadaist and surrealist painter.

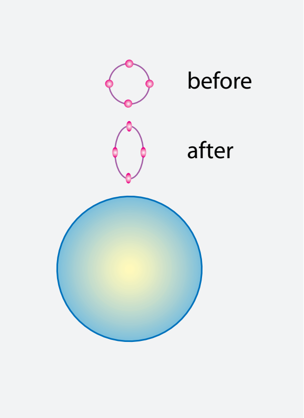


FIGURE 86 Tidal effects measure the curvature of space-time.

where l measures the length along the geodesic. Here, K is the local curvature, in other words, the inverse squared curvature radius. In the case of space-time, this relation is extended by substituting proper time τ (times the speed of light) for proper length. Thus separation and curvature are related by

$$\frac{d^2 s}{d\tau^2} = -Kc^2 s + \text{higher orders} . \qquad (194)$$

But this is the definition of an acceleration! In space-time, geodesics are the paths followed by freely falling particles. In other words, what in the purely spatial case is described by *curvature*, in the case of space-time becomes the *relative acceleration* of two nearby, freely falling particles. Indeed, we have encountered these accelerations already: Vol. I, page 197 they describe tidal effects. In short, space-time curvature and tidal effects are precisely the same.

Obviously, the magnitude of tidal effects, and thus of curvature, will depend on the orientation – more precisely on the orientation of the space-time plane formed by the two particle velocities. Figure 86 shows that the sign of tidal effects, and thus the sign of curvature, depends on the orientation: particles above each other diverge, particles side-by-side converge.

The definition of curvature also implies that K is a tensor, so that later on we will have Challenge 289 ny to add indices to it. (How many?) The fun is that we can avoid indices for a while by Ref. 188 looking at a special combination of spatial curvatures. If we take three planes in space, all orthogonal to each other and intersecting at a given point, the *sum* of these three so-called *sectional* curvatures does *not* depend on the observer. (This corresponds to the Challenge 290 ny tensor trace.) Can you confirm this, by using the definition of the curvature just given?

The sum of the three sectional curvatures defined for mutually orthogonal planes $K_{(12)}$, $K_{(23)}$ and $K_{(31)}$, is related to the excess radius defined above. Can you find out Challenge 291 ny how?

If a surface has *constant* curvature, i.e., the same curvature at all locations, geometrical Challenge 292 e objects can be moved around without deforming them. Can you picture this?

In summary, space-time curvature is an intuitive concept that describes how space-time is deformed. The local curvature of space-time is determined by following the motion of nearby, freely falling particles. If we imagine space (-time) as a mattress, a big blob of rubber inside which we live, the curvature at a point describes how this mattress is squeezed at that point. Since we live *inside* the mattress, we need to use 'insider' methods, such as excess radii and sectional curvatures, to describe the deformation.

Ref. 189

General relativity often seems difficult to learn because people do not like to think about the vacuum as a mattress, and even less to explain it in this way. We recall that for a hundred years it is an article of faith for every physicist to say that the vacuum is empty. This remains true. Nevertheless, picturing vacuum as a mattress, or as a substance, helps in many ways to understand general relativity.

Average curvature and motion in general relativity

One half of general relativity is the statement that any object moves along geodesics, i.e., along paths of *maximum* proper time. The other half is contained in a single expression: for *every* observer, the sum of all three *proper* sectional *spatial* curvatures at a point, the average curvature, is given by

$$K_{(12)} + K_{(23)} + K_{(31)} = \frac{8\pi G}{c^4} W^{(0)} \qquad (195)$$

where $W^{(0)}$ is the *proper* energy density at the point. The lower indices indicate the mixed curvatures defined by the three orthogonal directions 1, 2 and 3. This is all of general relativity in one paragraph.

We know that space-time is curved around mass and energy. Expression (195) specifies *how much* mass and energy curve space. We note that the factor on the right side is 2π divided by the maximum force.

Challenge 293 e

An equivalent description is easily found using the excess radius defined above, by introducing the mass $M = VW^{(0)}/c^2$. For the surface area A of the spherical volume V containing the mass, we get

$$r_{\text{excess}} = r - \sqrt{A/4\pi} = \frac{G}{3c^2} M \, . \qquad (196)$$

In short, general relativity affirms that for every observer, *the excess radius of a small sphere is given by the mass inside the sphere.*[*]

Note that both descriptions imply that the average space curvature at a point in empty space *vanishes*. As we will see shortly, this means that near a spherical mass the negative of the curvature *towards* the mass is equal to twice the curvature *around* the mass; the total sum is thus zero.

Curvature differs from point to point. In particular, the two descriptions imply that if

---

Ref. 190

[*] Another, equivalent formulation is that for small radii the area A is given by

$$A = 4\pi r^2 \left(1 + \frac{1}{9} r^2 R\right) \qquad (197)$$

where R is the Ricci scalar, to be introduced later on.

energy *moves*, curvature will move with it. In short, both space curvature and, as we will see shortly, space-time curvature *change* over space and time.

Challenge 294 ny

We note in passing that curvature has an annoying effect: the relative velocity of *distant* observers is undefined. Can you provide the argument? In curved space, relative velocity is defined only for *nearby* objects – in fact only for objects at no distance at all. Relative velocities of distant objects are well defined only in flat space.

The quantities appearing in expression (195) are *independent* of the observer. But often people want to use observer-dependent quantities. The relation then gets more involved; the single equation (195) must be expanded to ten equations, called *Einstein's field equations*. They will be introduced below. But before we do that, we will check that general relativity makes sense. We will skip the check that it contains special relativity as a limiting case, and go directly to the main test.

Universal gravity

> The only reason which keeps me here is gravity.
> Anonymous

For small velocities and low curvature values, the *temporal* curvatures $K_{(0j)}$ turn out to have a special property. In this case, they can be defined as the second spatial derivatives of a single scalar function φ. In other words, in everyday situations we can write

Challenge 295 e

$$K_{(0j)} = \frac{\partial^2 \varphi}{\partial (x^j)^2} \ . \qquad (198)$$

In everyday situations, this approximation is excellent, and the function φ turns out to be the gravitational potential. Indeed, low velocities and low curvature imply that we can set $W^{(0)} = \rho c^2$ and $c \to \infty$, so that we get

$$K_{(ij)} = 0 \quad \text{and} \quad K_{(01)} + K_{(02)} + K_{(03)} = \Delta\varphi = 4\pi G\rho \ . \qquad (199)$$

In other words, for small speeds, space is flat and the potential φ obeys Poisson's equation. Universal gravity is thus indeed the low speed and low curvature limit of general relativity.

Challenge 296 ny

Can you show that relation (195) between curvature and energy density indeed implies, in a more precise approximation, that time near a mass depends on the height, as mentioned before?

Page 137

The Schwarzschild metric

Ref. 188

What is the exact curvature of space-time near a spherical mass? The answer was given in 1915 by Karl Schwarzschild, who calculated the result during his military service in the First World War. Einstein then called the solution after him.

Page 145 In spherical coordinates the line element is

$$\mathrm{d}s^2 = \left(1 - \frac{2GM}{rc^2}\right) c^2 \mathrm{d}t^2 - \frac{\mathrm{d}r^2}{1 - \frac{2GM}{rc^2}} - r^2 \mathrm{d}\varphi^2 \, . \qquad (200)$$

Challenge 297 ny The curvature of the Schwarzschild metric is then by

$$\begin{aligned} K_{r\varphi} = K_{r\theta} &= -\frac{G}{c^2}\frac{M}{r^3} \quad \text{and} \quad K_{\theta\varphi} = 2\frac{G}{c^2}\frac{M}{r^3} \\ K_{t\varphi} = K_{t\theta} &= \frac{G}{c^2}\frac{M}{r^3} \quad \text{and} \quad K_{tr} = -2\frac{G}{c^2}\frac{M}{r^3} \end{aligned} \qquad (201)$$

Ref. 188 Vol. I, page 197 Challenge 298 ny everywhere. The dependence on $1/r^3$ follows from the general dependence of all tidal effects; we have already calculated them in the chapter on universal gravity. The factors G/c^2 are due to the maximum force of gravity. Only the numerical prefactors need to be calculated from general relativity. The average curvature obviously vanishes, as it does for all points in vacuum. As expected, the values of the curvatures near the surface of the Earth are exceedingly small.

Curiosities and fun challenges about curvature

> Il faut suivre sa pente, surtout si elle monte.[*]
> André Gide

Challenge 299 e A fly has landed on the outside of a cylindrical glass, 1 cm below its rim. A drop of honey is located halfway around the glass, also on the outside, 2 cm below the rim. What is the shortest distance from the fly to the drop? What is the shortest distance if the drop is on the *inside* of the glass?

* *

Challenge 300 e Where are the points of highest and lowest Gaussian curvature on an egg?

Three-dimensional curvature: the Ricci tensor[**]

> Jeder Straßenjunge in unserem mathematischen Göttingen versteht mehr von vierdimensionaler Geometrie als Einstein. Aber trotzdem hat Einstein die Sache gemacht, und nicht die großen Mathematiker.
> David Hilbert[***]

Now that we have a feeling for curvature, let us describe it in a way that allows *any* observer to talk to any *other* observer. Unfortunately, this means using formulae with

---

[*] ‘One has to follow one’s inclination, especially if it climbs upwards.’

[***] ‘Every street urchin in our mathematical Göttingen knows more about four-dimensional geometry than Einstein. Nevertheless, it was Einstein who did the work, not the great mathematicians.’

[***] The rest of this chapter might be skipped at first reading.

tensors. At fisrt, these formulae look daunting. The challenge is to see in each of the expressions the essential point (e.g. by forgetting all indices for a while) and not to be distracted by those small letters sprinkled all over them.

We mentioned above that a 4-dimensional space-time is described by 2-curvature, 3-curvature and 4-curvature. Many introductions to general relativity start with 3-curvature. 3-curvature describes the distinction between the 3-volume calculated from a radius and the actual 3-volume. The details are described by the *Ricci tensor*.* Exploring geodesic deviation, it turns out that the Ricci tensor describes how the shape of a spherical cloud of freely falling particles – a coffee cloud – is deformed along its path. More precisely, the Ricci tensor R_{ab} is (the precise formulation of) the second (proper) time derivative of the cloud volume divided by the cloud volume. In vacuum, the volume of such a falling coffee cloud always stays constant, and this despite the deformation due to tidal forces. Figure 86 illustrates that gravitation does not change coffee cloud volumes. In short, the Ricci tensor is the general-relativistic version of the Laplacian of the potential $\Delta\varphi$, or better, of the four-dimensional analogue $\Box\varphi$.

Ref. 191

Page 191

AVERAGE CURVATURE: THE RICCI SCALAR

The most global, but least detailed, definition of curvature is the one describing the distinction between the 4-volume calculated from a measured radius and the actual 4-volume. This is the *average curvature* at a space-time point and is represented by the so-called *Ricci scalar* R, defined as

$$R = -2K = \frac{-2}{r^2_{\text{curvature}}} \,. \qquad (202)$$

It turns out that the Ricci scalar can be derived from the Ricci tensor by a so-called *contraction*, which is a precise averaging procedure. For tensors of rank two, contraction is the same as taking the trace:

$$R = R^{\lambda}{}_{\lambda} = g^{\lambda\mu} R_{\lambda\mu} \,. \qquad (203)$$

The Ricci scalar describes the curvature averaged over space *and* time. In the image of a falling spherical cloud, the Ricci scalar describes the volume change of the cloud. The Ricci scalar always vanishes in vacuum. This result allows us to relate the spatial curvature to the change of time with height on the surface of the Earth.

Challenge 301 ny

THE EINSTEIN TENSOR

After two years of hard work, Einstein discovered that the best quantity for the description of curvature in nature is not the Ricci tensor R_{ab}, but a tensor built from it. This so-called *Einstein tensor* G_{ab} is defined mathematically (for vanishing cosmological constant) as

$$G_{ab} = R_{ab} - \frac{1}{2} g_{ab} R \,. \qquad (204)$$

---

* Gregorio Ricci-Cubastro (b. 1853 Lugo, d. 1925 Bologna), mathematician. He is the father of absolute differential calculus, also called 'Ricci calculus'. Tullio Levi-Civita was his pupil.

It is not difficult to understand its meaning. The value G_{00} is the sum of sectional curvatures in the planes *orthogonal* to the 0 direction and thus the sum of all spatial sectional curvatures:

$$G_{00} = K_{(12)} + K_{(23)} + K_{(31)} \ . \qquad (205)$$

Similarly, for each dimension i the diagonal element G_{ii} is the sum (taking into consideration the minus signs of the metric) of sectional curvatures in the planes *orthogonal* to the i direction. For example, we have

$$G_{11} = K_{(02)} + K_{(03)} - K_{(23)} \ . \qquad (206)$$

The distinction between the Ricci tensor and the Einstein tensor thus lies in the way in which the sectional curvatures are combined: discs *containing* the coordinate in question for the Ricci tensor, and discs *orthogonal* to the coordinate for the Einstein tensor. Both describe the curvature of space-time equally well, and fixing one means fixing the other. (What are the trace and the determinant of the Einstein tensor?) Challenge 302 d

The Einstein tensor is symmetric, which means that it has *ten* independent components. Most importantly, its divergence vanishes; it therefore describes a conserved quantity. This was the essential property which allowed Einstein to relate it to mass and energy in mathematical language.

THE DESCRIPTION OF MOMENTUM, MASS AND ENERGY

Obviously, for a complete description of gravity, the motion of momentum and energy need to be quantified in such a way that any observer can talk to any other. We have seen that momentum and energy always appear together in relativistic descriptions; the next step is thus to find out how their motions can be quantified for general observers.

First of all, the quantity describing energy, let us call it T, must be defined using the energy–momentum vector $\boldsymbol{p} = m\boldsymbol{u} = (\gamma mc, \gamma m\boldsymbol{v})$ of special relativity. Furthermore, T does not describe a single particle, but the way energy–momentum is distributed over space and time. As a consequence, it is most practical to use T to describe a *density* of energy and momentum. T will thus be a *field*, and depend on time and space, a fact usually indicated by the notation $T = T(t, x)$.

Since the energy–momentum density T describes a density over space and time, it defines, at every space-time point and for every infinitesimal surface $\mathrm{d}\boldsymbol{A}$ around that point, the flow of energy–momentum $\mathrm{d}\boldsymbol{p}$ through that surface. In other words, T is defined by the relation

$$\mathrm{d}\boldsymbol{p} = T \ \mathrm{d}\boldsymbol{A} \ . \qquad (207)$$

The surface is assumed to be characterized by its normal vector $\mathrm{d}\boldsymbol{A}$. Since the energy–momentum density is a proportionality factor between two vectors, T is a *tensor*. Of course, we are talking about 4-flows and 4-surfaces here. Therefore the energy–

momentum density tensor can be split in the following way:

$$T = \left(\begin{array}{c|ccc} w & S_1 & S_2 & S_3 \\ \hline S_1 & t_{11} & t_{12} & t_{13} \\ S_2 & t_{21} & t_{22} & t_{23} \\ S_3 & t_{31} & t_{32} & t_{33} \end{array}\right) = \left(\begin{array}{c|c} \text{energy} & \text{energy flow or} \\ \text{density} & \text{momentum density} \\ \hline \text{energy flow or} & \text{momentum} \\ \text{momentum density} & \text{flow density} \end{array}\right) \qquad (208)$$

where $w = T_{00}$ is a 3-scalar, $\boldsymbol{S}$ a 3-vector and t a 3-tensor. The total quantity T is called the *energy–momentum* (*density*) *tensor*. It has two essential properties: it is symmetric and its divergence vanishes.

The symmetry of the tensor T is a result of the conservation of angular momentum. The vanishing divergence of the tensor T, often written as

$$\partial_a T^{ab} = 0 \quad \text{or abbreviated} \quad T^{ab}{}_{,a} = 0\ , \qquad (209)$$

implies that the tensor describes a *conserved* quantity. In every volume, energy can change only via flow through its boundary surface. Can you confirm that the description of energy–momentum with this tensor satisfies the requirement that any two observers, differing in position, orientation, speed *and* acceleration, can communicate their results to each other? Challenge 303 ny

The energy–momentum density tensor gives a full description of the distribution of energy, momentum and mass over space and time. As an example, let us determine the energy–momentum density for a moving liquid. For a liquid of density ρ, a pressure p and a 4-velocity $\boldsymbol{u}$, we have

$$T^{ab} = (\rho_0 + p) u^a u^b - p g^{ab} \qquad (210)$$

where ρ_0 is the density measured in the comoving frame, the so-called *proper* density.* Obviously, ρ, ρ_0 and p depend on space and time.

Of course, for a particular material fluid, we need to know how pressure p and density ρ are related. A full material characterization thus requires the knowledge of the relation

$$p = p(\rho)\ . \qquad (212)$$

This relation is a material property and thus *cannot* be determined from relativity. It has to be derived from the constituents of matter or radiation and their interactions. The simplest possible case is *dust*, i.e., matter made of point particles** with no interactions

---

* In the *comoving* frame we thus have

$$T^{ab} = \begin{pmatrix} \rho_0 c^2 & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}\ . \qquad (211)$$

** Even though general relativity expressly forbids the existence of point particles, the approximation is useful in cases when the particle distances are large compared to their own size.

at all. Its energy–momentum tensor is given by

$$T^{ab} = \rho_0 u^a u^b \ . \tag{213}$$

Challenge 304 ny Can you explain the difference from the liquid case?

The divergence of the energy–momentum tensor vanishes for all times and positions, as you may want to check. Challenge 305 ny This property is the same as for the Einstein tensor presented above. But before we elaborate on this issue, a short remark. We did not take into account *gravitational energy*. It turns out that gravitational energy cannot be defined in general. In general, gravity does *not* have an associated energy. In certain special circumstances, such as weak fields, slow motion, or an asymptotically flat space-time, we *can* define the integral of the G^{00} component of the Einstein tensor as negative gravitational energy. Gravitational energy is thus only defined *approximately*, and only for our everyday environment.*

Einstein's field equations

> [Einstein's general theory of relativity] cloaked the ghastly appearance of atheism.
> A witch hunter from Boston, around 1935

> Do you believe in god? Prepaid reply 50 words.
> Subsequent telegram by another witch hunter to his hero Albert Einstein

> I believe in Spinoza's god, who reveals himself in the orderly harmony of what exists, not in a god who concerns himself with fates and actions of human beings.
> Albert Einstein's answer

Einstein's famous field equations were the basis of many religious worries. They contain the full description of general relativity. The equations can be deduced in many ways. The simplest way to deduce them is to start from the principle of maximum force. Page 113 Another way is to deduce the equation from the Hilbert action, as explained below. Page 201 A third way is what we are doing at present, namely to generalize the relation between curvature and energy to general observers.

Einstein's field equations are given by

$$G_{ab} = -\kappa \, T_{ab}$$

or, in more detail

$$R_{ab} - \frac{1}{2} g_{ab} R - \Lambda g_{ab} = -\kappa \, T^{ab} \quad . \tag{214}$$

---

* This approximation leads to the famous speculation that the total energy of the universe is zero. Do you agree? Challenge 306 s

The constant κ, called the *gravitational coupling constant*, has been measured to be

$$\kappa = \frac{8\pi G}{c^4} = 2.1 \cdot 10^{-43}\,/\mathrm{N} \qquad (215)$$

and its small value – the value 2π divided by the maximum force $c^4/4G$ – reflects the weakness of gravity in everyday life, or better, the difficulty of bending space-time. The constant Λ, the so-called *cosmological constant*, corresponds to a vacuum energy volume Page 243 density, or pressure Λ/κ. Its low value is quite hard to measure. The currently favoured value is

$$\Lambda \approx 10^{-52}\,/\mathrm{m}^2 \quad \text{or} \quad \Lambda/\kappa \approx 0.5\,\mathrm{nJ/m^3} = 0.5\,\mathrm{nPa}\,. \qquad (216)$$

Ref. 192 Current measurements and simulations suggest that this parameter, even though it is numerically near to the inverse square of the present radius of the universe, is a constant of nature that does not vary with time.

In summary, the field equations state that the curvature at a point is equal to the flow of energy–momentum through that point, taking into account the vacuum energy density. In other words: *Energy–momentum tells space-time how to curve, using the maximum force as proportionality factor.**

UNIVERSAL GRAVITATION – AGAIN

The field equations of general relativity can be simplified for the case in which speeds are small. In that case $T_{00} = c^2\rho$ and all other components of T vanish. Using the definition Challenge 307 ny of the constant κ and setting $\varphi = (c^2/2)h_{00}$ in $g_{ab} = \eta_{ab} + h_{ab}$, we find

$$\nabla^2\varphi = 4\pi\rho \quad \text{and} \quad \frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -\nabla\varphi \qquad (217)$$

---

* Einstein arrived at his field equations using a number of intellectual guidelines that are called *principles* in the literature. Today, many of them are not seen as central any more. Nevertheless, we give a short overview.

- *Principle of general relativity*: all observers are equivalent; this principle, even though often stated, is probably empty of any physical content.

- *Principle of general covariance*: the equations of physics must be stated in tensor form; even though it Ref. 193 is known today that all equations can be written with tensors, even universal gravity, in many cases they require unphysical 'absolute' elements, i.e., quantities which affect others but are not affected themselves. Vol. III, page 323 This unphysical idea is in contrast with the idea of *inter*action, as explained later on.

- *Principle of minimal coupling*: the field equations of gravity are found from those of special relativity by taking the simplest possible generalization. Of course, now that the equations are known and tested experimentally, this principle is only of historical interest.

- *Equivalence principle*: acceleration is locally indistinguishable from gravitation; we used it to argue that space-time is semi-Riemannian, and that gravity is its curvature.

- *Mach's principle*: inertia is due to the interaction with the rest of the universe; this principle is correct, even though it is often maintained that it is not fulfilled in general relativity. In any case, it is not the essence Page 258 of general relativity.

- *Identity of gravitational and inertial mass*: this is included in the definition of mass from the outset, but restated ad nauseam in general relativity texts; it is implicitly used in the definition of the Riemann tensor.

- *Correspondence principle*: a new, more general theory, such as general relativity, must reduce to previous theories, in this case universal gravity or special relativity, when restricted to the domains in which those are valid.

which we know well, since it can be restated as follows: a body of mass m near a body of mass M is accelerated by

$$a = G\frac{M}{r^2}, \tag{218}$$

a value which is independent of the mass m of the falling body. And indeed, as noted already by Galileo, all bodies fall with the same acceleration, independently of their size, their mass, their colour, etc. In general relativity also, gravitation is completely democratic.* The independence of free fall from the mass of the falling body follows from the description of space-time as a bent mattress. Objects moving on a mattress also move in the same way, independently of the mass value.

Understanding the field equations

To get a feeling for the complete field equations, we will take a short walk through their main properties. First of all, all motion due to space-time curvature is *reversible*, *differentiable* and thus *deterministic*. Challenge 308 e Note that only the complete motion, of space-time *and* matter *and* energy, has these properties. For particle motion only, motion is in fact *irreversible*, since some gravitational radiation is usually emitted.

By contracting the field equations we find, for vanishing cosmological constant, the following expression for the Ricci scalar:

$$R = -\kappa T\ . \tag{223}$$

This result also implies the relation between the excess radius and the mass inside a sphere. Challenge 309 ny

The field equations are *nonlinear* in the metric g, meaning that sums of solutions usually are *not* solutions. That makes the search for solutions rather difficult. For a complete solution of the field equations, initial and boundary conditions should be specified. The ways to do this form a specialized part of mathematical physics; it is not explored here. Ref. 194

Albert Einstein used to say that general relativity only provides the understanding of one side of the field equations (214), but not of the other. Can you see which side he meant? Challenge 310 ny

---

* Here is yet another way to show that general relativity fits with universal gravity. From the definition of the Riemann tensor we know that relative acceleration b_a and speed of nearby particles are related by

$$\nabla_e b_a = R_{ceda} v^c v^d\ . \tag{219}$$

From the symmetries of R we know there is a φ such that $b_a = -\nabla_a \varphi$. That means that

$$\nabla_e b^a = \nabla_e \nabla^a \varphi = R^a_{ced} v^c v^d \tag{220}$$

which implies that

$$\Delta\varphi = \nabla_a \nabla^a \varphi = R^a_{cad} v^c v^d = R_{cd} v^c v^d = \kappa(T_{cd} v^c v^d - T/2) \tag{221}$$

Introducing $T_{ab} = \rho v_a v_b$ we get

$$\Delta\varphi = 4\pi G\rho \tag{222}$$

as we wanted to show.

What can we do of interest with the field equations? In fact, to be honest, not much that we have not done already. Very few processes require the use of the full equations. Many textbooks on relativity even stop after writing them down! However, studying them is worthwhile. For example, one can show that the Schwarzschild solution is the *only* spherically symmetric solution. Similarly, in 1923, Birkhoff showed that every rotationally symmetric vacuum solution is static. This is the case even if masses themselves move, as for example during the collapse of a star.

Maybe the most beautiful applications of the field equations are the various *films* made of relativistic processes. The worldwide web hosts several of these; they allow one to see what happens when two black holes collide, what happens when an observer falls into a black hole, etc. To generate these films, the field equations usually need to be solved directly, without approximations.*

Another area of application concerns *gravitational waves*. The full field equations show that gravity waves are not harmonic, but nonlinear. Sine waves exist only approximately, for small amplitudes. Even more interestingly, if two waves collide, in many cases *singularities* of curvature are predicted to appear, i.e., points of infinite curvature. This whole theme is still a research topic and might provide new insights for the quantization of general relativity in the coming years.

We end this section with a side note. Usually, the field equations are read in one sense only, as stating that energy–momentum produces curvature. One can also read them in the other way, calculating the energy–momentum needed to produce a given curvature. When one does this, one discovers that not all curved space-times are possible, as some would lead to *negative* energy (or mass) densities. Such solutions would contradict the mentioned limit on length-to-mass ratios for physical systems.

HILBERT'S ACTION – HOW DOES SPACE BEND?

When Einstein discussed his research with David Hilbert, Hilbert found a way to do in a few weeks what had taken years for Einstein. Hilbert showed that general relativity *in empty space* could be described with the *least action principle*.

Hilbert knew that all motion minimizes action, i.e., all motion minimizes change. Hilbert set out to find the Lagrangian, i.e., the measure of change, for the *motion of space-time*, more precisely, for the *bending of space-time*. Obviously, such a measure must be observer-invariant; in particular, it must be invariant under *all* possible changes of viewpoint.

Motion due to gravity is determined by curvature. Any curvature measure independent of the observer must be a combination of the Ricci scalar R and the cosmological constant Λ. In this way both the equivalence principle and general covariance are respected. It thus makes sense to expect that the change of space-time is described by an action S given by

$$S = \frac{c^4}{16\pi G} \int (R - 2\Lambda)\, \mathrm{d}V \,. \qquad (224)$$

The volume element $\mathrm{d}V$ must be specified to use this expression in calculations. The cosmological constant Λ (added some years after Hilbert's work) appears as a mathematical

---

* See for example the www.photon.at/~werner/black-earth website.

possibility to describe the most general action that is diffeomorphism-invariant. We will see below that its value in nature, though small, seems to be different from zero.

We can also add matter to the Hilbert action; a lengthy calculation then confirms that the Hilbert action allows deducing Einstein's field equations – and vice versa. Both formulations are *equivalent*. The Hilbert action of a chunk of space-time is thus the integral of the Ricci scalar plus twice the cosmological constant over that chunk. The principle of least action states that space-time moves or bends in such a way that this integral changes as little as possible.

We note that the maximum force, with its huge value, appears as a prefactor in the action (224). A small deviation in curvature thus implies a huge observable action or change. This reflects the extreme stiffness of space-time. Can you show that the Hilbert action follows from the maximum force? Challenge 311 ny

In addition to the Hilbert action, for a full description of motion we need initial conditions. Ref. 128 The various ways to do this define a specific research field. This topic however, leads too far from our path. The same is valid for other, but equivalent, expressions of the action of general relativity.

In summary, the question 'how does space move?' is answered by the least action principle in the following way: *space evolves by minimizing scalar curvature*. The question 'how do things move?' is answered by general relativity in the same way as by special relativity: *things follow the path of maximal ageing*.

The symmetries of general relativity

The main symmetry of the Lagrangian of general relativity is called *diffeomorphism invariance* or *general covariance*. Physically speaking, the symmetry states that motion is independent of the coordinate system used. More precisely, the motion of matter, radiation and space-time does not change under arbitrary differentiable coordinate transformations, or *diffeomorphisms*. Diffeomorphism invariance is the essential symmetry of the Hilbert action: motion is independent of coordinates systems.

The field equations for empty space-time also show *scale symmetry*. This is the invariance of the equations after multiplication of all coordinates by a common numerical factor. In 1993, Torre and Anderson showed that diffeomorphism symmetry and trivial scale symmetry are the *only* symmetries of the vacuum field equations. Ref. 195

Apart from diffeomorphism invariance, full general relativity, including mass–energy, has an additional symmetry that is not yet fully elucidated. This symmetry connects the various possible initial conditions of the field equations; the symmetry is extremely complex Ref. 196 and is still a topic of research. These fascinating investigations might give new insights into the classical description of the big bang.

In summary, the symmetries of general relativity imply that also the fastest, the most distant and the most powerful motion in nature is *relative*, *continuous*, *reversible* and *mirror invariant*. The symmetries also confirm that the most violent motion *conserves energy–momentum and angular momentum*. Finally, Hilbert's action confirms that even the wildest motion in nature is *lazy*, i.e., described by the least action principle.

In short, despite adding motion of vacuum and horizons, general relativity does not change our everyday concept of motion. Relativity is a *classical* description of motion.

MASS IN GENERAL RELATIVITY

Page 285 The diffeomorphism-invariance of general relativity makes life quite interesting. We will see that it allows us to say that we live on the *inside* of a hollow sphere. We have seen that general relativity does not allow us to say where energy is actually located. If energy cannot be located, what about mass? Exploring the issue shows that mass, like energy, can be localized *only* if distant space-time is known to be flat. It is then possible to define a localized mass value by making precise an intuitive idea: the mass of an unknown body is measured by the time a probe takes to orbit the unknown body.*

Challenge 312 ny The intuitive mass definition *requires* flat space-time at infinity; it cannot be extended to other situations. In short, mass can only be localized if total mass can be defined. And *total mass* is defined only for asymptotically flat space-time. The only other notion of mass that is precise in general relativity is the *local mass density* at a point. In contrast, it is not well understood how to define the mass contained in a region larger than a point but smaller than the entirety of space-time (in the case that it is not asymptotically flat).

THE FORCE LIMIT AND THE COSMOLOGICAL CONSTANT

When the cosmological constant is taken into the picture, the maximum force principle requires a second look. In the case of a non-vanishing cosmological constant, the force limit makes sense only if the constant Λ is positive; this is the case for the currently measured value, which is $\Lambda \approx 10^{-52}/\mathrm{m}^2$. Indeed, the radius–mass relation of black holes Ref. 199 Ref. 128, Ref. 129

$$2GM = Rc^2\left(1 - \frac{\Lambda}{3}R^2\right) \qquad (227)$$

implies that a radius-*independent* maximum force is valid only for positive or zero cosmological constant. For a negative cosmological constant the force limit would only be valid for infinitely small black holes. In the following, we take a pragmatic approach and note that a maximum force limit can be seen to imply a vanishing or positive cosmological constant. Obviously, the force limit does not specify the *value* of the constant; to achieve this, a second principle needs to be added. A straightforward formulation, using the additional principle of a minimum force in nature, was proposed above. Page 134

One might ask also whether rotating or charged black holes change the argument that leads from maximum force to the derivation of general relativity. However, the deriva-

---

Ref. 197 * This definition was formalized by Arnowitt, Deser and Misner, and since then has often been called the *ADM mass*. The idea is to use the metric g_{ij} and to take the integral

$$m = \frac{c^2}{32\pi G}\int_{S_R}(g_{ij,i}\nu_j - g_{ii,j}\nu_j)\mathrm{d}A \qquad (225)$$

where S_R is the coordinate sphere of radius R, ν is the unit vector normal to the sphere and $\mathrm{d}A$ is the area element on the sphere. The limit exists for large R if space-time is asymptotically flat and if the mass distribution is sufficiently concentrated. Mathematical physicists have also shown that for any manifold whose metric changes at infinity as Ref. 198

$$g_{ij} = (1 + f/r + O(1/r^2))\delta_{ij} \qquad (226)$$

the total mass is given by $M = fc^2/G$.

tion using the Raychaudhuri equation does not change. In fact, the only change of the argument appears with the inclusion of *torsion*, which changes the Raychaudhuri equation itself. As long as torsion plays no role, the derivation given above remains valid. The inclusion of torsion is still an open research issue.

IS GRAVITY AN INTERACTION?

We tend to answer this question affirmatively, as in Galilean physics gravity was seen as an influence on the motion of bodies. In Galilean physics, we described gravity by a potential, because gravity changes motion. Indeed, a force or an interaction is what changes the motion of objects. However, we just saw that when two bodies attract each other through gravitation, both always remain in free fall. For example, the Moon circles the Earth because it continuously falls around it. Since any freely falling observer continuously remains at rest, the statement that gravity changes the motion of bodies is not correct for all observers. In fact, given that geodesics are the path of maximum straightness, we can also argue that the Moon and the Earth both follow 'straight' paths, and for all observers. But objects that follow straight paths are not under the influence of interactions, are they?

Vol. III, page 322

Let us explore this issue in another way. The most fundamental definition of 'interaction' is as the difference between the whole and the sum of its parts. In the case of gravity, an observer in free fall could indeed claim that nothing special is going on, independently of whether the other body is present or not, and could claim that gravity is not an interaction.

Page 198

Challenge 313 s

However, an interaction also transports energy between systems. Now, we have seen that gravity can be said to transport energy only approximately. The properties of gravitational energy confirm this argument. Even in its energy aspect, gravitation is an interaction only approximately.

Challenge 314 ny

A mathematical way to look at these issue is the following. Take a satellite orbiting Jupiter with energy–momentum $\boldsymbol{p} = m\boldsymbol{u}$. If we calculate the energy–momentum change along its path s, we get

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}s} = m\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}s} = m\left(\boldsymbol{e}_a\frac{\mathrm{d}\boldsymbol{u}^a}{\mathrm{d}s} + \frac{\mathrm{d}\boldsymbol{e}_a}{\mathrm{d}s}\boldsymbol{u}^a\right) = m\boldsymbol{e}_a\left(\frac{\mathrm{d}\boldsymbol{u}^a}{\mathrm{d}s} + \Gamma^a{}_{bd}\boldsymbol{u}^b\boldsymbol{u}^d\right) = 0 \qquad (228)$$

Challenge 315 ny

Ref. 200

Challenge 316 ny

where $\boldsymbol{e}$ describes the unit vector along a coordinate axis and $\Gamma^a{}_{bd}$ is the *metric connection*; it is explained below. The energy–momentum change vanishes along any geodesic, as you might check. Therefore, the energy–momentum of this motion is conserved. In other words, *no* force is acting on the satellite. We could reply that in equation (228) the second term alone is the real gravitational force. But this term can be made to vanish along the entirety of any given world line. In short, also the mathematics confirm that nothing changes between two bodies in free fall around each other: gravity could be said not to be an interaction.

Let us look at the behaviour of light. In vacuum, light is always moving freely. In a sense, we can say that radiation always is in free fall. Strangely, since we called free fall the same as rest, we should conclude that radiation always is at rest. This is not wrong! We have already seen that light cannot be accelerated.[*] We have also seen that gravitational bending is not an acceleration, since light follows straight paths in space-time in this case as well. Even though light seems to slow down near masses for distant observers, it always moves at the speed of light locally. In short, even gravitation doesn't manage to move light.

In short, if we like such intellectual games, we can argue that gravitation is not an interaction, even though it puts objects into orbits and deflects light. For all practical purposes, gravity remains an interaction.

HOW TO CALCULATE THE SHAPE OF GEODESICS

One half of general relativity states that bodies fall along geodesics. All orbits are geodesics, thus curves with the longest proper time. It is thus useful to be able to calculate these trajectories.[**] To start, one needs to know the *shape of space-time*, the notion of 'shape' being generalized from its familiar two-dimensional meaning. For a being living on the surface, it is usually described by the metric g_{ab}, which defines the distances between neighbouring points through

$$\mathrm{d}s^2 = \mathrm{d}x_a\,\mathrm{d}x^a = g_{ab}(x)\,\mathrm{d}x^a\,\mathrm{d}x^b\ . \tag{229}$$

It is a famous exercise of calculus to show from this expression that a curve $x^a(s)$ depending on a well behaved (affine) parameter s is a time-like or space-like (metric) *geodesic*, i.e., the longest possible path between the two events,[***] only if

Challenge 317 ny

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(g_{ad}\frac{\mathrm{d}x^d}{\mathrm{d}s}\right) = \frac{1}{2}\frac{\partial g_{bc}}{\partial x^a}\frac{\mathrm{d}x^b}{\mathrm{d}s}\frac{\mathrm{d}x^c}{\mathrm{d}s}\ , \tag{230}$$

---

[*] Refraction, the slowdown of light inside matter, is not a counter-example. Strictly speaking, light inside matter is constantly being absorbed and re-emitted. In between these processes, light still propagates with the speed of light in vacuum. The whole process only *looks* like a slowdown in the macroscopic limit. The same applies to diffraction and to reflection. A full list of ways to bend light can be found elsewhere.

Vol. III, page 157

[**] This is a short section for the more curious; it can be skipped at first reading.

[***] We remember that in space in everyday life, geodesics are the shortest possible paths; however, in space-time in general relativity, geodesics are the longest possible paths. In both cases, they are the 'straightest' possible paths.

as long as $\mathrm{d}s$ is different from zero along the path.* All bodies in free fall follow such geodesics. We showed above that the geodesic property implies that a stone thrown in the air falls back, unless if it is thrown with a speed larger than the escape velocity. Expression (230) thus replaces both the expression $\mathrm{d}^2x/\mathrm{d}t^2 = -\nabla\varphi$ valid for falling bodies and the expression $\mathrm{d}^2x/\mathrm{d}t^2 = 0$ valid for freely floating bodies in special relativity. Page 149

The path does not depend on the mass or on the material of the body. Therefore *antimatter* also falls along geodesics. In other words, antimatter and matter do not repel; they also attract each other. Interestingly, even experiments performed with normal matter can show this, if they are carefully evaluated. Can you find out how? Ref. 201 Challenge 318 ny

For completeness, we mention that light follows *lightlike* or *null* geodesics. In other words, there is an affine parameter u such that the geodesics follow

$$\frac{\mathrm{d}^2x^a}{\mathrm{d}u^2} + \Gamma^a{}_{bc}\frac{\mathrm{d}x^b}{\mathrm{d}u}\frac{\mathrm{d}x^c}{\mathrm{d}u} = 0 \tag{234}$$

with the different condition

$$g_{ab}\frac{\mathrm{d}x^a}{\mathrm{d}u}\frac{\mathrm{d}x^b}{\mathrm{d}u} = 0\ . \tag{235}$$

Challenge 319 ny Given all these definitions of various types of geodesics, what are the lines that are drawn in Figure 65 on page 144?

RIEMANN GYMNASTICS**

Most books introduce curvature the hard way, namely historically, using the Riemann curvature tensor. This is a short summary, so that you can understand that old stuff when you come across it.

We saw above that curvature is best described by a tensor. In 4 dimensions, this curvature tensor, usually called R, must be a quantity which allows us to calculate, among other things, the area for any orientation of a 2-disc in space-time. Now, in 4 dimensions, orientations of a disc are defined in terms of *two* 4-vectors; let us call them $\boldsymbol{p}$ and $\boldsymbol{q}$. And instead of a disc, we take the *parallelogram* spanned by $\boldsymbol{p}$ and $\boldsymbol{q}$. There are several possible definitions. Challenge 320 e

The *Riemann-Christoffel curvature tensor* R is then defined as a quantity which allows

---

* This is often written as

$$\frac{\mathrm{d}^2x^a}{\mathrm{d}s^2} + \Gamma^a_{bc}\frac{\mathrm{d}x^b}{\mathrm{d}s}\frac{\mathrm{d}x^c}{\mathrm{d}s} = 0 \tag{231}$$

where the condition

$$g_{ab}\frac{\mathrm{d}x^a}{\mathrm{d}s}\frac{\mathrm{d}x^b}{\mathrm{d}s} = 1 \tag{232}$$

must be fulfilled, thus simply requiring that all the tangent vectors are *unit* vectors, and that $\mathrm{d}s \neq 0$ all along the path. The symbols Γ appearing above are given by

$$\Gamma^a{}_{bc} = \left\{ {a \atop bc} \right\} = \frac{1}{2}g^{ad}(\partial_b g_{dc} + \partial_c g_{db} - \partial_d g_{bc})\ , \tag{233}$$

and are called *Christoffel symbols of the second kind* or simply the *metric connection*.

** This is a short section for the more curious; it can be skipped at first reading.

us to calculate the curvature $K(\boldsymbol{p},\boldsymbol{q})$ for the surface spanned by $\boldsymbol{p}$ and $\boldsymbol{q}$, with area A, through

$$K(\boldsymbol{p},\boldsymbol{q}) = \frac{R\ \boldsymbol{pqpq}}{A^2(\boldsymbol{p},\boldsymbol{q})} = \frac{R_{abcd}\,p^a q^b p^c q^d}{(g_{\alpha\delta} g_{\beta\gamma} - g_{\alpha\gamma} g_{\beta\delta})\,p^\alpha q^\beta p^\gamma q^\delta} \tag{236}$$

where, as usual, Latin indices a, b, c, d, etc. run from 0 to 3, as do Greek indices here, and a *summation* is implied when an index name appears twice. Obviously R is a tensor, of rank 4. This tensor thus describes only the *intrinsic* curvature of a space-time. In contrast, the metric g describes the complete *shape* of the surface, not only the curvature. The curvature is thus the physical quantity of relevance locally, and physical descriptions therefore use only the *Riemann** *tensor* R or quantities derived from it.**

But we can forget the just-mentioned definition of curvature. There is a second, more physical way to look at the Riemann tensor. We know that curvature means gravity. As we said above, gravity means that when two nearby particles move freely with the same velocity and the same direction, the distance between them changes. In other words, the local effect of gravity is *relative acceleration* of nearby particles.

Challenge 321 e

It turns out that the tensor R describes precisely this relative acceleration, i.e., what we called the *tidal effects* earlier on. Obviously, the relative acceleration $\boldsymbol{b}$ increases with the separation $\boldsymbol{d}$ and the square (why?) of the speed $\boldsymbol{u}$ of the two particles. Therefore we can also define R as a (generalized) proportionality factor among these quantities:

Challenge 322 ny

$$\boldsymbol{b} = R\,\boldsymbol{u}\,\boldsymbol{u}\,\boldsymbol{d} \quad \text{or, more clearly,} \quad b^a = R^a{}_{bcd}\,u^b\,u^c\,d^d\ . \tag{239}$$

The components of the Riemann curvature tensor have the dimensions of inverse square length. Since it contains all information about intrinsic curvature, we conclude that if R vanishes in a region, space-time in that region is flat. This connection is easily deduced from this second definition.***

Challenge 323 ny

---

* Bernhard Riemann (b. 1826 Breselenz, d. 1866 Selasca), important mathematician. One among his numerous important achievements is the foundation of non-Euclidean geometry.

** We showed above that space-time is curved by noting changes in clock rates, in metre bar lengths and in light propagation. Such experiments are the easiest way to determine the metric g. We know that space-time is described by a 4-dimensional manifold M with a metric g_{ab} that locally, at each space-time point, is a Minkowski metric. Such a manifold is called a *Riemannian manifold*. Only such a metric allows one to define a local inertial system, i.e., a local Minkowski space-time at every space-time point. In particular, we have

$$g_{ab} = 1/g^{ab} \quad \text{and} \quad g_a{}^b = g^a{}_b = \delta^a_b\ . \tag{237}$$

How are curvature and metric related? The solution to this question usually occupies a large number of pages in relativity books; just for information, the relation is

$$R^a{}_{bcd} = \frac{\partial \Gamma^a{}_{bd}}{\partial x^c} - \frac{\partial \Gamma^a{}_{bc}}{\partial x^d} + \Gamma^a{}_{ec}\Gamma^e{}_{bd} - \Gamma^a{}_{fd}\Gamma^f{}_{bc}\ . \tag{238}$$

The curvature tensor is built from the second derivatives of the metric. On the other hand, we can also determine the metric if the curvature is known. An approximate relation is given below.

*** This second definition is also called the definition through *geodesic deviation*. It is of course not evident that it coincides with the first. For an explicit proof, see the literature. There is also a third way to picture the tensor R, a more mathematical one, namely the original way Riemann introduced it. If one parallel-transports a vector $\boldsymbol{w}$ around a parallelogram formed by two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, each of length ε, the vector $\boldsymbol{w}$

Ref. 202

A final way to define the tensor R is the following. For a *free-falling* observer, the metric g_{ab} is given by the metric η_{ab} from special relativity. In its neighbourhood, we have

$$\begin{aligned} g_{ab} &= \eta_{ab} + \frac{1}{3} R_{acbd} x^c x^d + O(x^3) \\ &= \frac{1}{2} (\partial_c \partial_d g_{ab}) x^c x^d + O(x^3) \ , \end{aligned} \qquad (241)$$

where O denotes terms of higher order. The curvature term thus describes the departure of the space-time metric from that of flat space-time. The curvature tensor R is a large beast; it has $4^4 = 256$ components at each point of space-time; however, its symmetry properties reduce them to 20 independent numbers.* The actual number of importance in physical problems is still smaller, namely only 10. These are the components of the Ricci tensor, which can be defined with the help of the Riemann tensor by contraction, i.e., by setting

$$R_{bc} = R^a{}_{bac} \ . \qquad (244)$$

Its components, like those of the Riemann tensor, are inverse square lengths. The values of the tensor R_{bc}, or those of R^a_{bcd}, are independent of the sign convention used in the Minkowski metric, in contrast to R_{abcd}. Challenge 326 e

Challenge 327 ny Can you confirm the relation $R_{abcd} R^{abcd} = 48 m^2 / r^6$ for the Schwarzschild solution?

CURIOSITIES AND FUN CHALLENGES ABOUT GENERAL RELATIVITY

For various years, people have speculated why the Pioneer 10 and 11 artificial satellites, which are now over 70 astronomical units away from the Sun, are subject to a constant deceleration of $8 \cdot 10^{-10}\,\mathrm{m/s^2}$, directed towards, the Sun since they passed the orbit of Saturn. This deceleration is called the *Pioneer anomaly*. The origin was an intense subject of research. Several investigations have shown that the reason of the deceleration is *not* a deviation from the inverse square dependence of gravitation, as was proposed by some. Ref. 203

---

is changed to $\boldsymbol{w} + \delta \boldsymbol{w}$. One then has

$$\delta \boldsymbol{w} = -\varepsilon^2 R\, \boldsymbol{u}\, \boldsymbol{v}\, \boldsymbol{w} + \quad \text{higher-order terms} \quad . \qquad (240)$$

More can be learned about the geodesic deviation by studying the behaviour of the famous south-pointing carriage which we have encountered before. This device, used in China before the compass was discovered, only works if the world is flat. Indeed, on a curved surface, after following a large closed path, it will show a different direction than at the start of the trip. Can you explain why? Vol. I, page 244 Challenge 324 s

* The free-fall definition shows that the Riemann tensor is symmetric in certain indices and antisymmetric in others: Challenge 325 ny

$$R_{abcd} = R_{cdab} \quad , \quad R_{abcd} = -R_{bacd} = -R_{abdc} \ . \qquad (242)$$

These relations also imply that many components vanish. Of importance also is the relation

$$R_{abcd} + R_{adbc} + R_{acdb} = 0 \ . \qquad (243)$$

Note that the order of the indices is not standardized in the literature. The list of invariants which can be constructed from R is long. We mention that $\frac{1}{2} \varepsilon^{abcd} R_{cd}{}^{ef} R_{abef}$, namely the product ${}^*R\ R$ of the Riemann tensor with its dual, is the invariant characterizing the Thirring–Lense effect.

The effect is electromagnetic.

There were many hints that pointed to an asymmetry in heat radiation emission of the satellites. The on-board generators produce 2.5 kW of heat that is radiated away by the satellite. A front-to-back asymmetry of only 80 W is sufficient to explain the measured anomaly. Recent research has shown that such an asymmetry indeed exists, so that the issue is now resolved. Ref. 204

* *

Maximum power or force appearing on horizons is the basis for general relativity. Are there physical systems other than space-time that can also be described in this way?

Page 36 For special relativity, we found that all its main effects – such as a limit speed, Lorentz contraction or energy–mass equivalence – are also found for dislocations in solids. Do systems analogous to general relativity exist? So far, attempts to find such systems have only been partially successful.

Several equations and ideas of general relativity are applicable to deformations of solids, since general relativity describes the deformation of the space-time mattress. Kröner has studied this analogy in great detail. Ref. 118

Other physical systems with 'horizons', and thus with observables analogous to curvature, are found in certain liquids – where vortices play the role of black holes – and in certain quantum fluids for the propagation of light. Exploring such systems has become a research topic in its own right. Ref. 205

A full analogy of general relativity in a macroscopic system was discovered only a few years ago. This analogy will be presented in the final part of our adventure. Vol. VI, page 281

* *

Can the maximum force principle be used to eliminate competing theories of gravitation? The most frequently discussed competitors to general relativity are scalar–tensor theories of gravity, such as the proposal by Brans and Dicke and its generalizations.

Page 115 If a particular scalar-tensor theory obeys the general horizon equation (112) then it must also imply a maximum force. The general horizon equation must be obeyed both for *static* and for *dynamic* horizons. If that were the case, the specific scalar–tensor theory would be equivalent to general relativity, because it would allow one, using the argument of Jacobson, to deduce the usual field equations. This case can appear if the scalar field behaves like matter, i.e., if it has mass–energy like matter and curves space-time like matter. On the other hand, if in the particular scalar–tensor theory the general horizon equation is not obeyed for *all moving* horizons – which is the general case, as scalar–tensor theories have more defining constants than general relativity – then the maximum force does not appear and the theory is not equivalent to general relativity. This connection also shows that an experimental test of the horizon equation for *static* horizons only is not sufficient to confirm general relativity; such a test rules out only some, but not all, scalar–tensor theories.

* *

One way to test general relativity would be to send three space probes through the solar system, and measure their relative position over time, with high precision. This is best done using frequency-stabilized lasers that send light from one satellite to the other two.

Can you summarize the main technical risks involved in such a project? Can you find ways to reduce them? Challenge 328 s

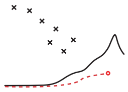
A SIMPLE SUMMARY OF THE FIELD EQUATIONS

The field equations of general relativity describe motion of space, matter and energy. They state that:

— The local curvature of space is given by the local energy density divided by the maximum force.

— Objects move along the geodesics defined by this local curvature.

This description is confirmed to full precision by all experiments performed so far.



Chapter 8

WHY CAN WE SEE THE STARS? – MOTION IN THE UNIVERSE

> Zwei Dinge erfüllen das Gemüt mit immer neuer und zunehmender Bewunderung und Ehrfurcht, je öfter und anhaltender sich das Nachdenken damit beschäftigt: der bestirnte Himmel über mir und das moralische Gesetz in mir.[**]
>
> Immanuel Kant

On clear nights, between two and five thousand stars are visible with the naked eye. Of them, several hundred have names. Why? Because in all parts of the world, the stars and the constellations they form are attached to myths. Ref. 207 In all civilisations, myths are stories told to make the incomprehensible more comprehensible. But the simple fact that we can *see* the stars is the basis for a story much more fantastic than all myths. It touches almost all aspects of modern physics and encompasses the complete history of the universe.

Which stars do we see?

> Democritus says [about the Milky Way] that it is a region of light emanating from numerous stars small and near to each other, of which the grouping produces the brightness of the whole.
>
> Ref. 208 Aetius, *Opinions*.

The stars we see on a clear night are mainly the brightest of our nearest neighbours in the surrounding region of the Milky Way. They lie at distances between four and a few thousand light years from us. Roughly speaking, in our environment there is a star about every 400 cubic light years. Our Sun is just one of the one hundred thousand million stars of the Milky Way.

At night, almost all stars visible with the naked eye are from our own galaxy. The only extragalactic object *constantly* visible to the naked eye in the northern hemisphere is the so-called Andromeda nebula, shown enlarged in Figure 91. It is a whole galaxy like our own, as Immanuel Kant had already conjectured in 1755. Several extragalactic objects are

---

[**] 'Two things fill the mind with ever new and increasing admiration and awe, the more often and persistently thought considers them: the starred sky above me and the moral law inside me.' Ref. 206 Immanuel Kant (1724–1804) was the most important philospher of the *Enlightenment*, the movement that lead to modern science and western standard of wealth and living by pushing aside the false ideas spread by religion-based governments.

FIGURE 87 A modern photograph of the visible night sky, showing a few thousand stars and the Milky Way. The image is a digital composite of many photographs of cloudless night skies taken all over the Earth. The Milky Way is positioned horizontally (© Axel Mellinger, from Ref. 209).

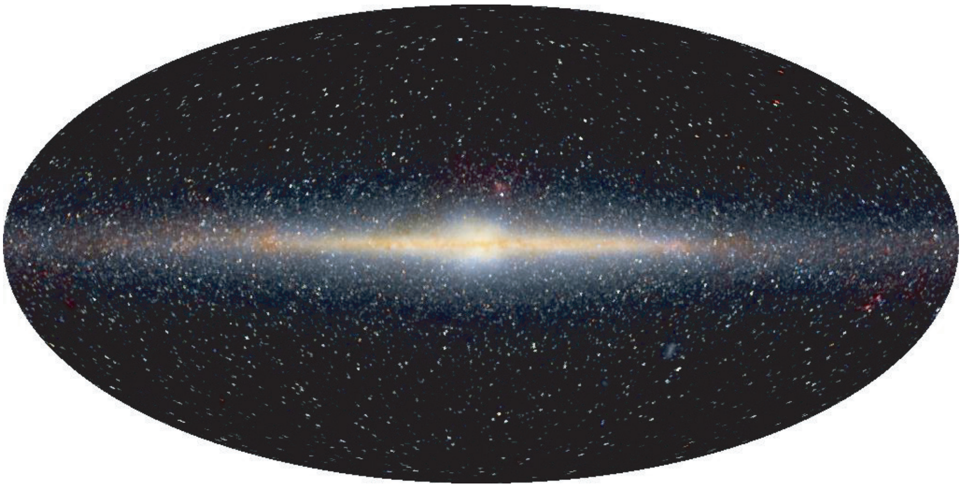


FIGURE 88 A false colour image of how the night sky, and our galaxy in particular, looks in the near infrared (courtesy NASA).

visible with the naked eye in the southern hemisphere: the Tarantula nebula, as well as the large and the small Magellanic clouds. The Magellanic clouds are neighbour galaxies to our own. Other, *temporarily* visible extragalactic objects are the rare *novae*, exploding stars which can be seen if they appear in nearby galaxies, or the still rarer *supernovae*, which can often be seen even in faraway galaxies.

In fact, the visible stars are special in other respects also. For example, telescopes show that about half of them are in fact double: they consist of two stars circling around each other, as in the case of Sirius. Measuring the orbits they follow around each other allows one to determine their masses. Can you explain how?

Challenge 329 ny

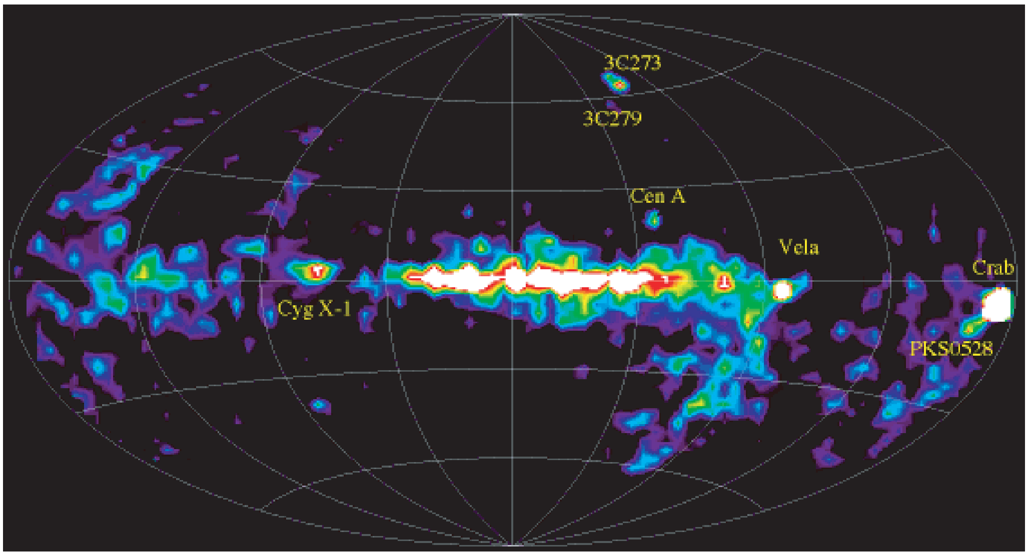


FIGURE 89 A false colour image of the X-ray sources observed in the night sky, for energies between 1 and 30 MeV (courtesy NASA).

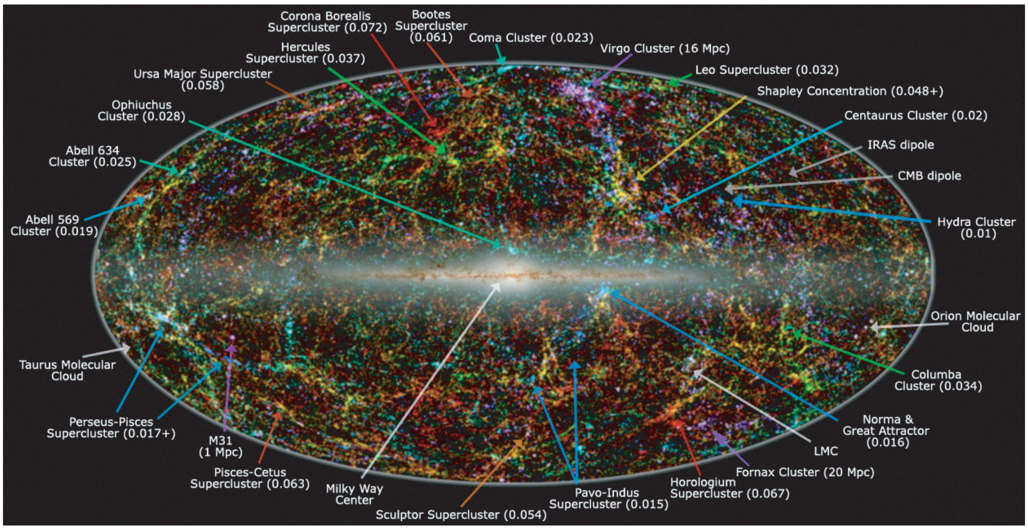


FIGURE 90 A false colour image, composed from infrared data, showing the large-scale structure of the universe around us; the colour of each galaxy represents its distance and the numbers in parentheses specify the red-shift; an infrared image of the Milky Way is superposed (courtesy Thomas Jarret/IPAC/Caltech).

Vol. III, page 163

Many more extragalactic objects are visible with telescopes. Nowadays, this is one of the main reasons to build them, and to build them as large as technically possible.

Is the universe different from our Milky Way? Yes, it is. There are several arguments to demonstrate this. First of all, our galaxy – the word *galaxy* is just the original Greek term for ‘Milky Way’ – is *flattened*, because of its rotation. If the galaxy rotates, there must be other masses which determine the background with respect to which this rotation takes

FIGURE 91 The Andromeda nebula M31, one of our neighbour galaxies (and the 31st member of the Messier object listing) (NASA).

place. In fact, there is a huge number of other galaxies – about 10^{11} – in the universe, a discovery dating only from the twentieth century. Some examples are shown in Figure 91, Figure 92 and Figure 93. The last figure shows how galaxies usually 'die': by colliding with other galaxies.

Why did our understanding of the place of our galaxy in the universe happen so late? Well, people had the same difficulty as they had when trying to determine the shape of the Earth. They had to understand that the galaxy is not only a milky strip seen on clear nights, but an actual physical system, made of about 10^{11} stars gravitating around each other.* Like the Earth, the Milky Way was found to have a three-dimensional *shape*: As shown by the infrared photograph in Figure 88, our galaxy is a flat and circular structure, with a spherical bulge at its centre. The diameter is 100 000 light years. It rotates about once every 200 to 250 million years. (Can you guess how this is measured?) The rotation is quite slow: since the Sun was formed, it has made only about 20 to 25 full turns around the centre. Challenge 330 ny

It is even possible to measure the *mass* of our galaxy. The trick is to use a binary pulsar on its outskirts. If it is observed for many years, one can deduce its acceleration around the galactic centre, as the pulsar reacts with a frequency shift which can be measured on Earth. Many decades of observation are needed and many spurious effects have to be eliminated. Nevertheless, such measurements are ongoing. Present estimates put the mass of our galaxy at 10^{42} kg or $5 \cdot 10^{11}$ solar masses. Ref. 210

How do we watch the stars?

The best images of the night sky are produced by the most sensitive telescopes. On Earth, the most sensitive telescopes are the largest ones, such as those shown in Figure 96, located in Paranal in Chile. The history and the capabilities of these telescopes are fascinating. For many wavelengths that are absorbed by the atmosphere, the most sensitive telescopes are satellite-bound, such as those shown in Figure 97. For each wavelength domain, such modern systems produce fascinating images of the night sky. Figure 87 to Ref. 211

---

* The Milky Way, or *galaxy* in Greek, was said to have originated when Zeus, the main Greek god, tried to let his son Heracles feed at Hera's breast in order to make him immortal; the young Heracles, in a sign showing his future strength, sucked so forcefully that the milk splashed all over the sky.

FIGURE 92 The elliptical galaxy NGC 205 (the 205th member of the New Galactic Catalogue) (NASA).

FIGURE 93 The colliding galaxies M51 and M51B, 65 000 al across, 31 Mal away, show how a galaxy 'dies' (NASA).

Figure 90 give some examples. A beautiful website dedicated to showing how the night sky looks at different wavelengths is www.chromoscope.net. The website allows you to slide from one wavelength to another simply by moving a cursor; watching it and exploring the beauty of the universe is worth it.

FIGURE 94 The universe is full of galaxies – this photograph shows the Perseus cluster (NASA).

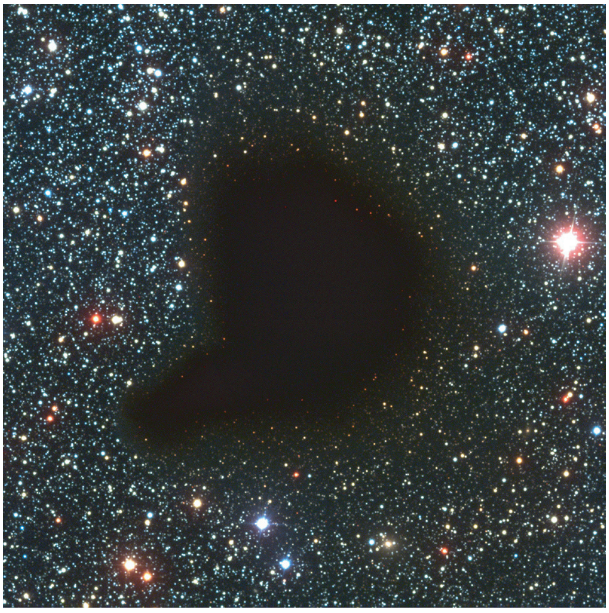


FIGURE 95 The universe contains many clouds; an example is this molecular cloud in Ophiuchus (© ESO).

What do we see at night?

Astrophysics leads to a strange conclusion about matter, quite different from how we are used to thinking in classical physics: *the matter observed in the sky is found in clouds.*

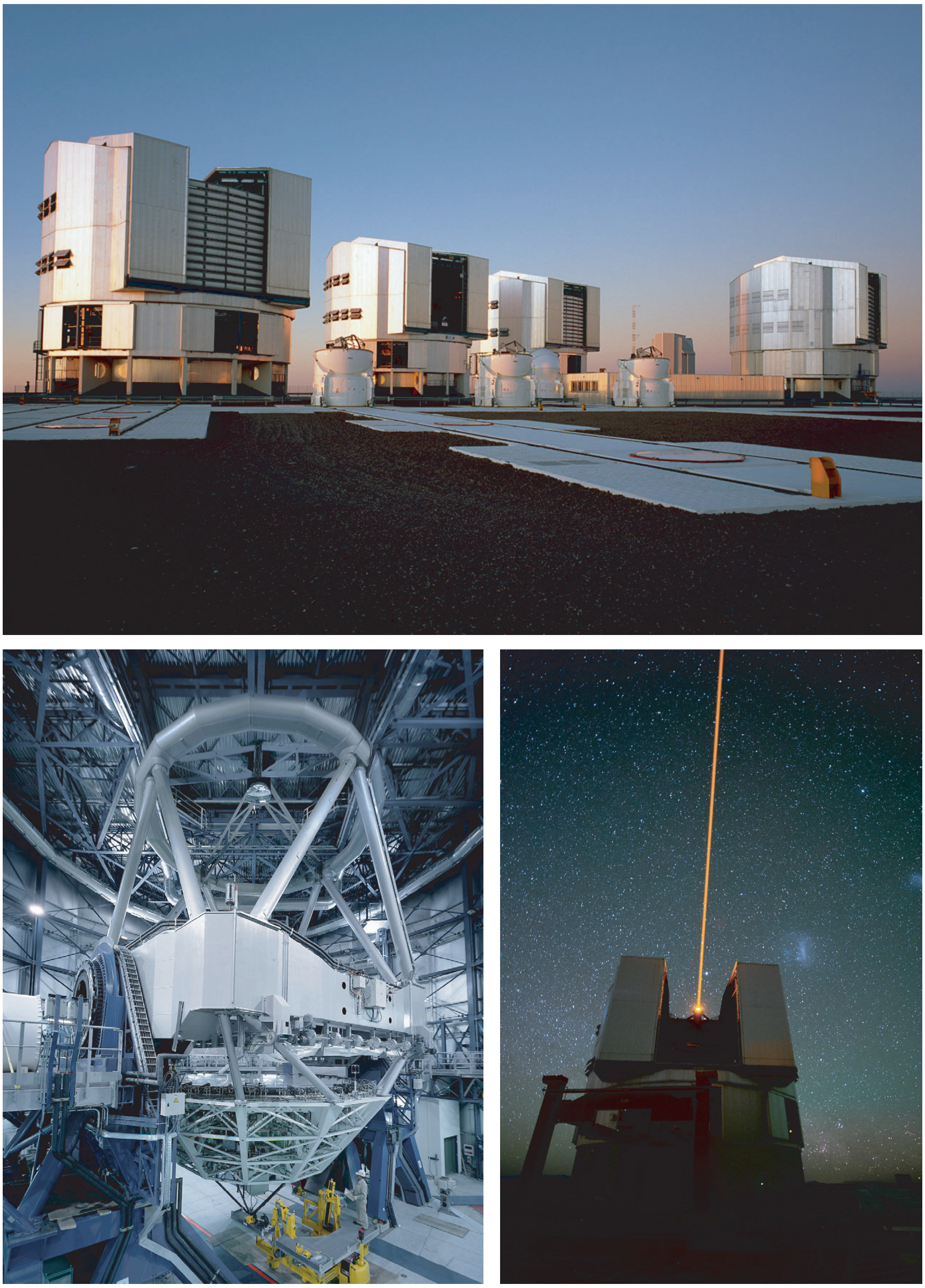


FIGURE 96 One of the four Very Large Telescopes (VLT) of the European Southern Observatory (ESO) in Paranal in Chile, the most powerful telescopes in the world, each with a diameter of 8 m (© ESO).

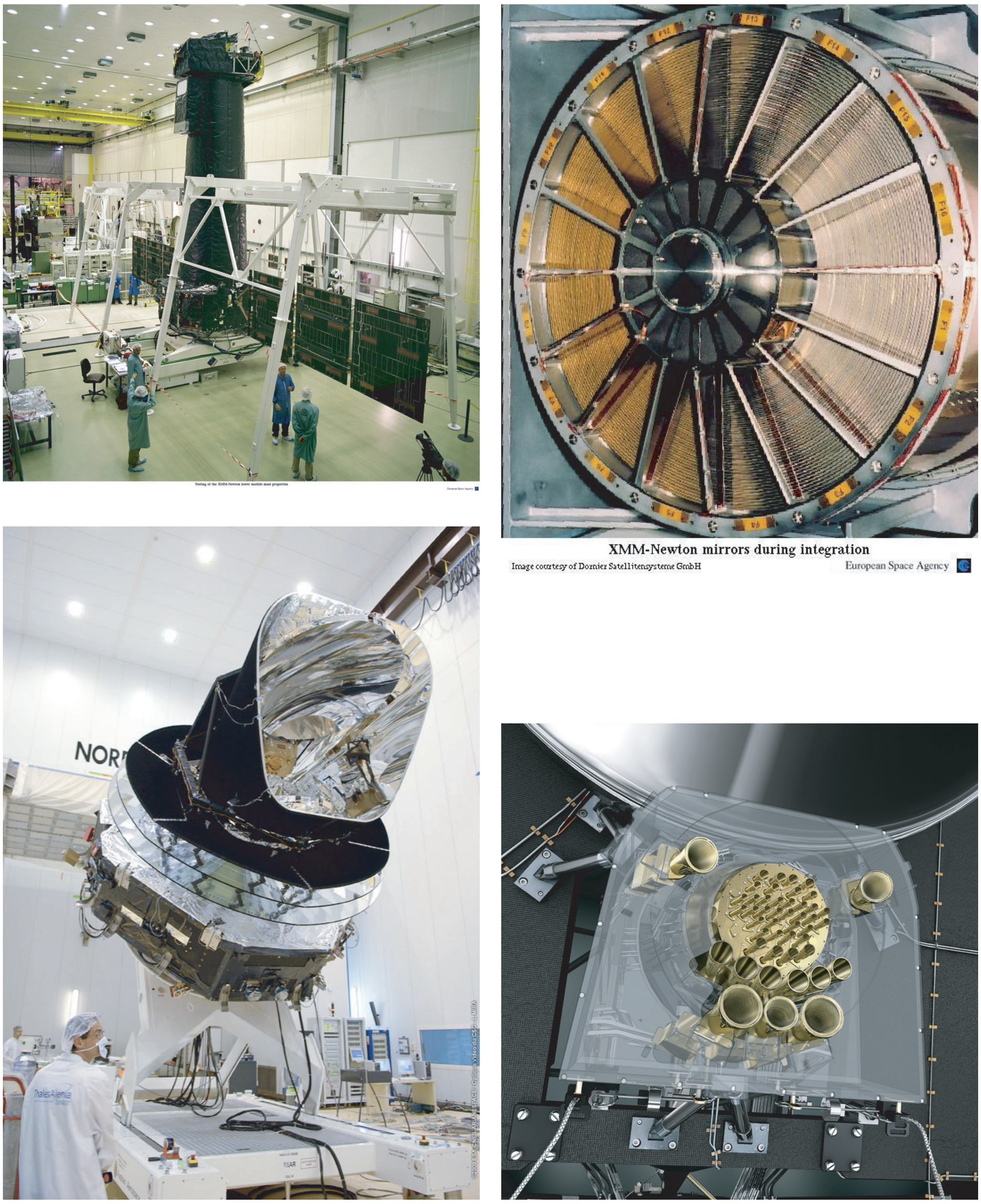


FIGURE 97 Top: the XMM-Newton satellite and its high-precision, onion-like mirrors that produced an X-ray map of the night sky. Bottom: the Planck satellite and its golden-plated microwave antennas that produced a high-resolution map of the cosmic background radiation (© ESA).

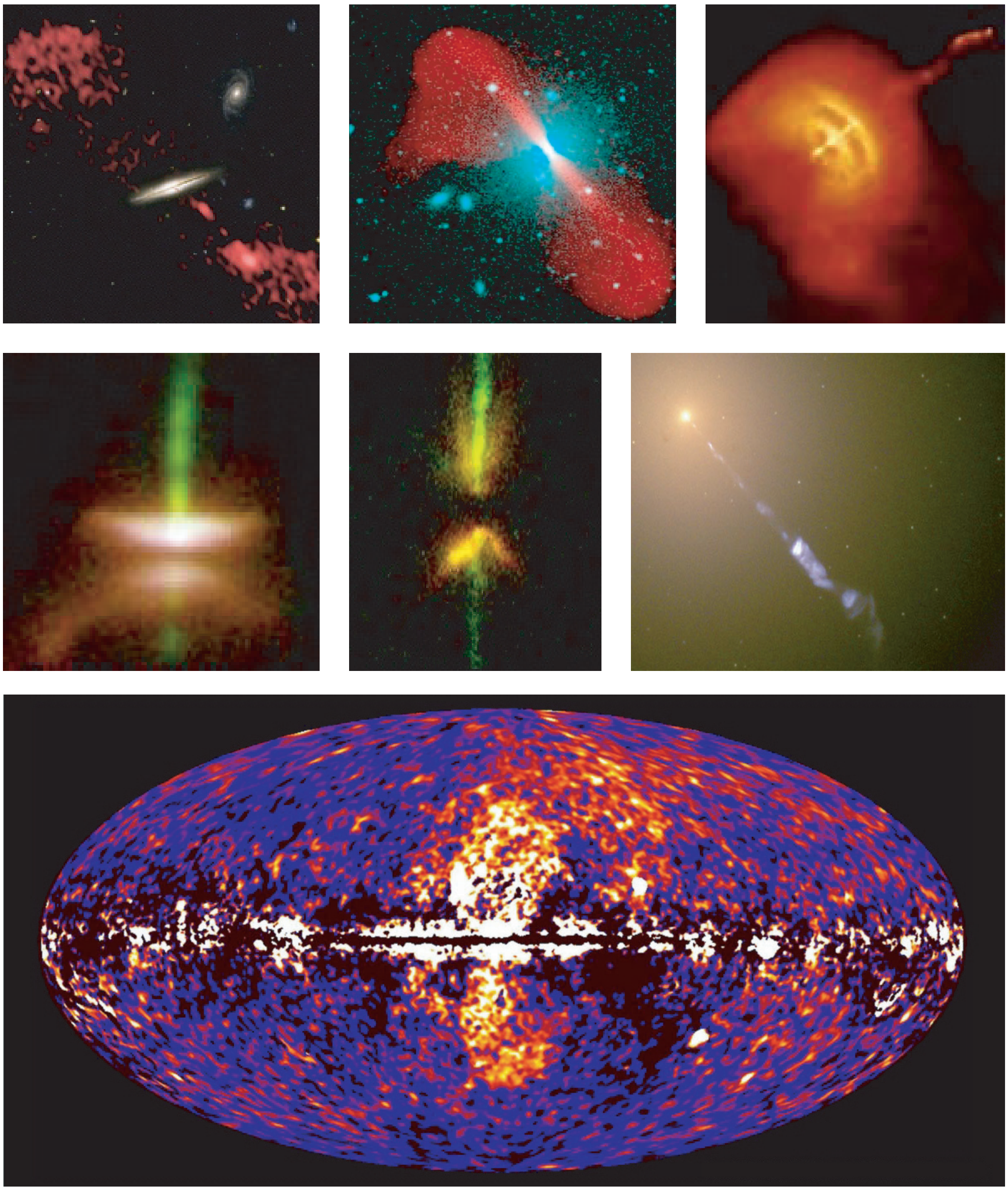


FIGURE 98 Rotating clouds emitting jets along their axis; top row: a composite image (visible and infrared) of the galaxy 0313-192, the galaxy 3C296, and the Vela pulsar; middle row: the star in formation HH30, the star in formation DG Tauri B, and a black hole jet from the galaxy M87; bottom row: the discovery of jets in our own galaxy (all NASA).

Page 216

Clouds are systems in which the matter density diminishes with the distance from the centre, with no sharp border and with no definite size. The object shown in Figure 95 is a molecular cloud. But this is not the only case. Most astrophysical objects, including planets and stars, are clouds.

The Earth is also a cloud, if we take its atmosphere, its magnetosphere and the dust ring around it as part of it. The Sun is a cloud. It is a gas ball to start with, but is even more a cloud if we take into consideration its protuberances, its heliosphere, the solar wind it generates and its magnetosphere. The solar system is a cloud if we consider its comet cloud, its asteroid belt and its local interstellar gas cloud. The galaxy is a cloud if we remember its matter distribution and the cloud of cosmic radiation it is surrounded by. In fact, even people can be seen as clouds, as every person is surrounded by gases, little dust particles from skin, vapour, etc.

Ref. 212 In the universe, almost all clouds are plasma clouds. A *plasma* is an ionized gas, such as fire, lightning, the inside of neon tubes, or the Sun. At least 99.9 % of all *matter in the universe is in the form of plasma clouds*. Only a very small percentage exists in solid or liquid form, such as toasters, toothpicks or their users.

All clouds in the universe share a number of common properties. First, all clouds seen in the universe – when undisturbed by collisions or other interactions from neighbouring objects – are *rotating*. Most clouds are therefore *flattened*: they are in shape of discs. Secondly, in many rotating clouds, matter is falling towards the centre: most clouds are *accretion discs*. Finally, undisturbed accretion discs usually emit something along the rotation axis: they possess *jets*. This basic cloud structure has been observed for young stars, for pulsars, for galaxies, for quasars and for many other systems. Figure 98 gives some examples. Finally, in 2010, jets have been found in our own galaxy, the Milky Way. (Does the Sun have jets? So far, none has been detected.) Challenge 331 r

In summary, at night we see mostly rotating, flattened plasma clouds emitting jets along their axes. But the night sky has many other phenomena. A large part of astronomy and astrophysics collects information about them. An overview about the observations is given in Table 5. Ref. 213

TABLE 5 Some observations about the universe.

Aspect	Main properties	Value
Phenomena		
Galaxy formation	observed by Hubble	several times
	trigger event	unknown
Galactic collisions	momentum	10^{45} to 10^{47} kg m/s
Star formation	cloud collapse	forms stars between 0.04 and 130 solar masses
	frequency	between 0 and 1000 solar masses per year per galaxy; around 1 solar mass per year in the Milky Way
	or by star mergers	up to 250 solar masses
Novae	new luminous stars,	$L < 10^{31}$ W
	ejecting bubble	$R \approx t \cdot c/100$
Supernovae	new bright stars,	$L < 10^{36}$ W
	rate	1 to 5 per galaxy per 1000 a
Hypernovae	optical bursts	$L > 10^{37}$ W

TABLE 5 (Continued) Some observations about the universe.

Aspect	Main properties	Value
Gamma-ray bursts	luminosity	L up to 10^{45} W, about 1 % of the whole visible universe's luminosity
	energy	*c.* 10^{46} J
	duration	*c.* 0.015 to 1000 s
	observed number	*c.* 2 per day
Radio sources	radio emission	10^{33} to 10^{38} W
X-ray sources	X-ray emission	10^{23} to 10^{34} W
Cosmic rays	energy	from 1 eV to 10^{22} eV
Gravitational lensing	light bending	angles down to $10^{-4}{}''$
Comets	recurrence, evaporation	typ. period 50 a, typ. visibility lifetime 2 ka, typ. lifetime 100 ka
Meteorites	age	up to $4.57 \cdot 10^{9}$ a
Components		
Intergalactic space	mass density	*c.* 10^{-26} kg/m^3
Quasars	red-shift	up to $z = 6$
	luminosity	$L = 10^{40}$ W, about the same as one galaxy
Galaxy superclusters	number of galaxies	*c.* 10^{8} inside our horizon
Our own local supercluster	number of galaxies	about 4000
Galaxy groups	size	100 Zm
	number of galaxies	between a dozen and 1000
Our local group	number of galaxies	30
Galaxies	size	0.5 to 2 Zm
	number	*c.* 10^{11} inside horizon
	containing	10 to 400 globular clusters
	containing	typically 10^{11} stars each
	containing	typically one supermassive and several intermediate-mass black holes
The Milky Way, our galaxy	diameter	1.0(0.1) Zm
	mass	10^{42} kg or $5 \cdot 10^{11}$ solar masses Ref. 210
	speed	600 km/s towards Hydra-Centaurus
	containing	about 30 000 pulsars Ref. 214
	containing	100 globular clusters each with 1 million stars
Globular clusters (e.g. M15)	containing	thousands of stars, one intermediate-mass black hole
	age	up to 12 Ga (oldest known objects)
Nebulae, clouds	composition	dust, oxygen, hydrogen
Our local interstellar cloud	size	20 light years

TABLE 5 (Continued) Some observations about the universe.

ASPECT	MAIN PROPERTIES	VALUE
	composition	atomic hydrogen at 7500 K
Star systems	types	orbiting double stars, over 70 stars orbited by brown dwarfs, several planetary systems
Our solar system	size	2 light years (Oort cloud)
	speed	368 km/s from Aquarius towards Leo
Stars	mass	up to 130 solar masses (more when stars merge) Ref. 215
giants and supergiants	large size	up to 1 Tm
main sequence stars		
brown dwarfs	low mass	below 0.072 solar masses
	low temperature	below 2800 K Ref. 216
L dwarfs	low temperature	1200 to 2800 K
T dwarfs	low temperature	900 to 1100 K
white dwarfs	small radius	$r \approx 5000\,\mathrm{km}$
	high temperature	cools from 100 000 to 5000 K
neutron stars	nuclear mass density	$\rho \approx 10^{17}\,\mathrm{kg/m^3}$
	small size	$r \approx 10\,\mathrm{km}$
emitters of X-ray bursts	X-ray emission	
pulsars	periodic radio emission	
	mass	up to around 25 solar masses
magnetars	high magnetic fields	up to 10^{11} T and higher Ref. 217
	some are gamma repeaters, others are anomalous X-ray pulsars	
	mass	above 25 solar masses Ref. 218
Black holes	horizon radius	$r = 2GM/c^2$, observed mass range from 3 solar masses to 10^{11} solar masses
General properties		
Cosmic horizon	distance	c. $10^{26}\,\mathrm{m} = 100\,\mathrm{Ym}$
Expansion	Hubble's constant	$71(4)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ or $2.3(2) \cdot 10^{-18}\,\mathrm{s^{-1}}$
'Age' of the universe		13.8(1) Ga
Vacuum	energy density	$0.5\,\mathrm{nJ/m^3}$ or $\Omega_\Lambda = 0.73$ for $k = 0$
		no evidence for time-dependence
Large-scale shape	space curvature	$k \approx \Omega_K = 0$ Page 236
	topology	simple at all measured scales
Dimensions	number	3 for space, 1 for time, at all measured energies and scales

TABLE 5 (Continued) Some observations about the universe.

Aspect	Main properties	Value
Matter	density	2 to 11 · 10^{-27} kg/m^3 or 1 to 6 hydrogen atoms per cubic metre $\Omega_M = 0.25$
Baryons	density	$\Omega_b = 0.04$, one sixth of the previous (included in Ω_M)
Dark matter	density	$\Omega_{DM} = 0.21$ (included in Ω_M), unknown
Dark energy	density	$\Omega_{DM} = 0.75$, unknown
Photons	number density	4 to 5 · 10^8 /m^3 = 1.7 to 2.1 · 10^{-31} kg/m^3
	energy density	$\Omega_R = 4.6 \cdot 10^{-5}$
Neutrinos	energy density	Ω_ν unknown
Average temperature	photons	2.725(2) K
	neutrinos	not measured, predicted value is 2 K
Radiation perturbations	photon anisotropy	$\Delta T/T = 1 \cdot 10^{-5}$
	density amplitude	$A = 0.8(1)$
	spectral index	$n = 0.97(3)$
	tensor-to-scalar ratio	$r < 0.53$ with 95 % confidence
Ionization optical depth		$\tau = 0.15(7)$
Decoupling		$z = 1100$

But while we are speaking of what we see in the sky, we need to clarify a general issue.

WHAT IS THE UNIVERSE?

> “I’m astounded by people who want to ‘know’ the universe when it’s hard enough to find your way around Chinatown.”
>
> Woody Allen

The term ‘universe’ implies turning. The universe is what turns around us at night. For a physicist, at least three definitions are possible for the term ‘universe’:

- The (*observable* or *visible*) *universe* is the totality of all observable mass and energy. This includes everything inside the cosmological horizon. Since the horizon is moving away from us, the amount of observable mass and energy is constantly increasing. The content of the term ‘observable universe’ is thus not fixed in time. (What is the origin of this increase? We will come back to this issue in the final leg of our adventure.) Vol. VI, page 307
- The (*believed*) *universe* is the totality of all mass and energy, *including* any that is not observable. Numerous books on general relativity state that there definitely exists matter or energy beyond the observation boundaries. We will explain the origin of this belief below. (Do you agree with it?) Challenge 332 e

FIGURE 99 The beauty of astronomy: the Cygnus Bubble, discovered in 2008, a nebula expelled from a central star (false colour image courtesy of T.A. Rector, H. Schweiker).

- The (*full*) *universe* is the sum of matter and energy *as well as* space-time itself.

These definitions are often mixed up in physical and philosophical discussions. There is *no* generally accepted consensus on the terms, so one has to be careful. In this text, when we use the term 'universe', we imply the *last* definition only. We will discover repeatedly that without clear distinction between the definitions we cannot complete our adventure. (For example: Is the amount of matter and energy in the full universe the same as in the observable universe?) Challenge 333 s

Note that the 'size' of the visible universe, or better, the distance to its horizon, is a quantity which *can* be imagined. The value of 10^{26} m, or ten thousand million light years, is not beyond imagination. If we took all the iron from the Earth's core and made it into a wire reaching to the edge of the observable universe, how thick would it be? The answer might surprise you. Challenge 334 s Also, the content of the universe is clearly finite. There are about as many visible *galaxies* in the universe as there are grains in a cubic metre of sand. To expand on the comparison, can you deduce how much space you would need to contain all the flour you would get if every little speck, with a typical size of 150 μm, represented

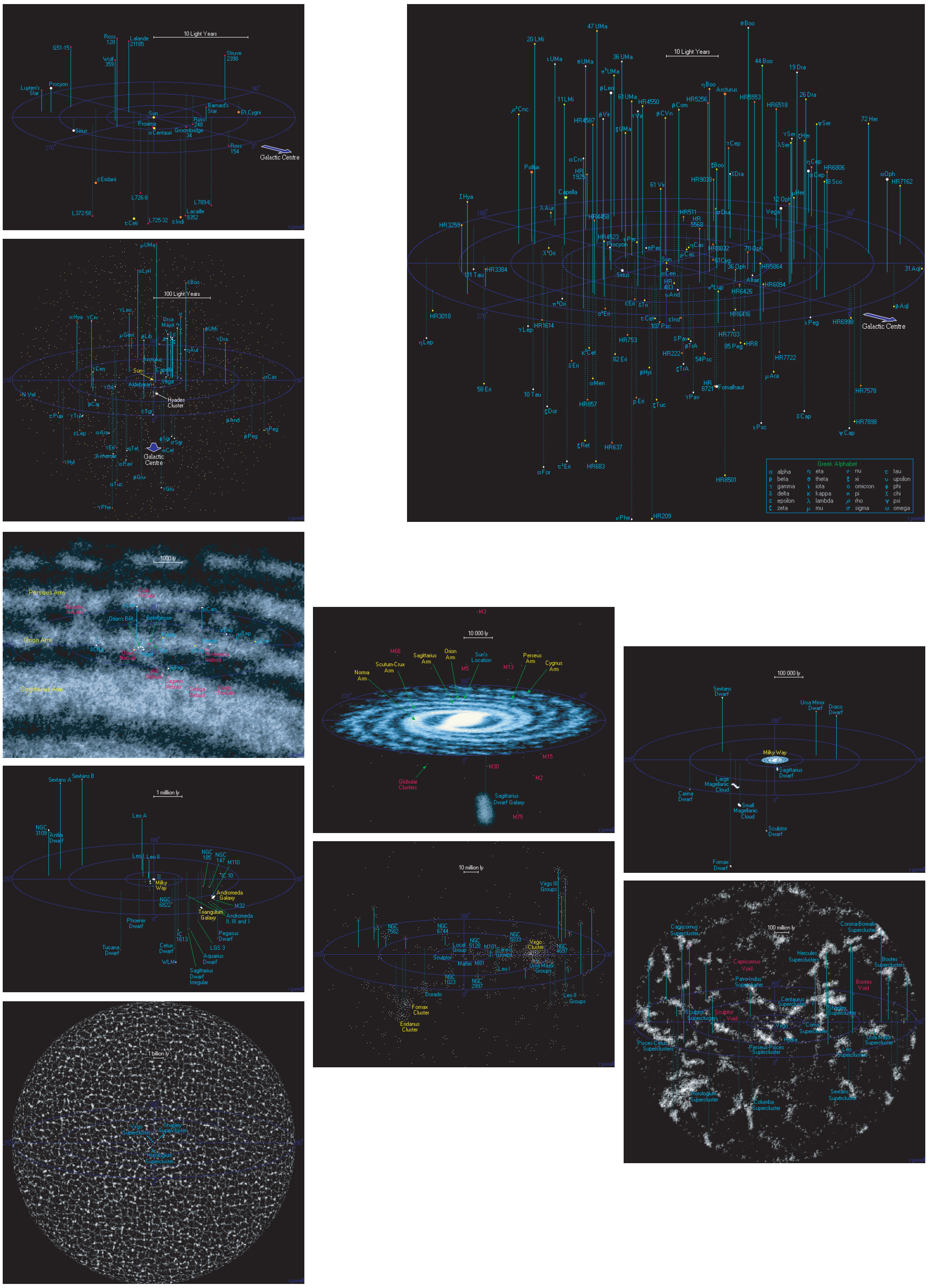


FIGURE 100 An atlas of our cosmic environment: illustrations at scales up to 12.5, 50, 250, 5 000, 50 000, 500 000, 5 million, 100 million, 1 000 million and 14 000 million light years (© Richard Powell, www.atlasoftheuniverse.com).

Challenge 335 s one star?

THE COLOUR AND THE MOTION OF THE STARS

> Ἠ τοι μὲν πρώτιστα Ξάος γένετ᾽ … *
> Hesiod, *Theogony*.

Obviously, the universe is full of motion. To get to know the universe a bit, it is useful to measure the speed and position of as many objects in it as possible. In the twentieth century, a large number of such observations were obtained from stars and galaxies. Challenge 336 s (Can you imagine how distance and velocity are determined?) This wealth of data can be summed up in two points.

First of all, on large scales, i.e., averaged over about five hundred million light years, the matter density in the universe is *homogeneous* and *isotropic*. Obviously, at smaller scales inhomogeneities exist, such as galaxies or cheesecakes. Our galaxy for example is Ref. 219 neither isotropic nor homogeneous. But at large scales the differences average out. This large-scale homogeneity of matter distribution is often called the *cosmological principle*.

The second point about the universe is even more important. In the 1920s, independently, Ref. 220 Carl Wirtz, Knut Lundmark and Gustaf Stromberg showed that on the whole, all galaxies *move away from the Earth*, and the more so, the more they were distant. There are a few exceptions for nearby galaxies, such as the Andromeda nebula itself; but in general, the speed of flight v of an object increases with distance d. In 1929, the US-American astronomer Edwin Hubble** published the first measurement of the relation between speed and distance. Despite his use of incorrect length scales he found a relation

$$v = H\,d\,, \tag{245}$$

where the proportionality constant H is today called the *Hubble constant*. A modern graph of the relation is given in Figure 101. The Hubble constant is known today to have a value around $71\,\mathrm{km\,s^{-1}Mpc^{-1}}$. (Hubble's own value was so far from this value that it is not cited any more.) For example, a star at a distance of 2 Mpc*** is moving away from Earth with a speed of around 142 km/s, and proportionally more for stars further away.

In fact, the discovery by Wirtz, Lundmark and Stromberg implies that *every* galaxy Challenge 337 s moves away from *all* the others. (Why?) In other words, the matter in the universe is *expanding*. The scale of this expansion and the enormous dimensions involved are amazing. The motion of all the thousand million galaxy groups in the sky is described by the single equation (245)! Some deviations are observed for nearby galaxies, as mentioned above, and for faraway galaxies, as we will see.

---

* 'Verily, at first Chaos came to be …' The Theogony, attributed to the probably mythical Hesiodos, was finalized around 700 BCE. It can be read in English and Greek on the www.perseus.tufts.edu website. The famous quotation here is from verse 117.

** Edwin Powell Hubble (1889–1953), important US-American astronomer. After being an athlete and taking a law degree, he returned to his childhood passion of the stars; he finally proved Immanuel Kant's 1755 conjecture that the Andromeda nebula was a galaxy like our own. He thus showed that the Milky Way is only a tiny part of the universe.

Page 307 *** A *megaparsec* or Mpc is a distance of 30.8 Zm.

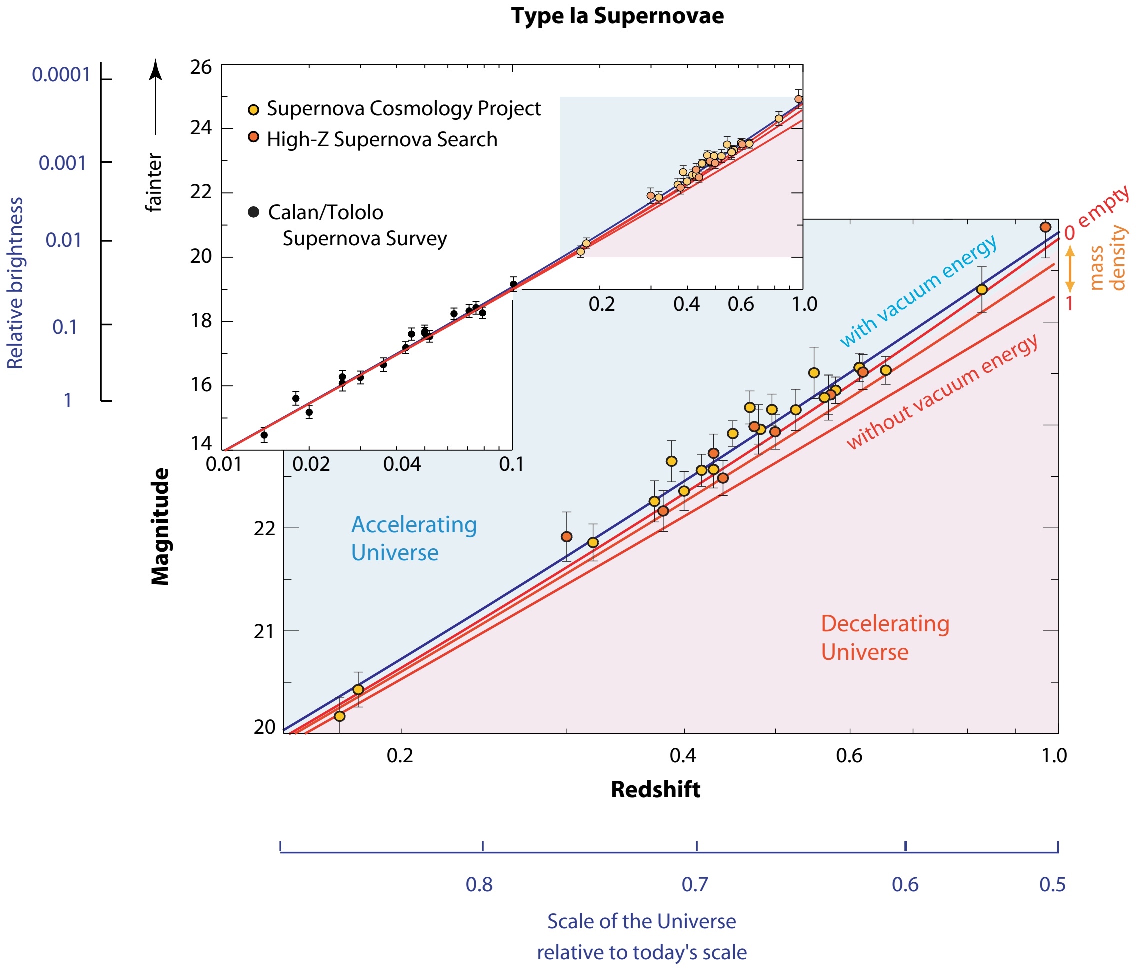


FIGURE 101 The relation between star distance and star velocity (courtesy Saul Perlmutter and the Supernova Cosmology Project).

The cosmological principle and the expansion taken together imply that the universe cannot have existed before time when it was of vanishing size; the universe thus has a *finite age*. Together with the evolution equations, as explained in more detail below, the Hubble constant points to an age value of around 13 800 million years. The expansion also means that the universe has a *horizon*, i.e., a finite maximum distance for sources whose signals can arrive on Earth. Signals from sources beyond the horizon cannot reach us.

The motion of galaxies tells something important: in the past, the night sky, and thus the universe, has been much *smaller*; matter has been much *denser* than it is now. It turns out that matter has also been much *hotter*. George Gamow* predicted in 1948 that since hot objects radiate light, the sky cannot be completely black at night, but must

Ref. 221

---

* George Gamow (b. 1904 Odessa, d. 1968 St. Boulder), physicist. He explained alpha decay as a tunnelling effect and predicted the microwave background. He wrote the first successful popular physics texts, such as *1, 2, 3, infinity* and the *Mr. Thompkins* series, which were later imitated by many other writers.

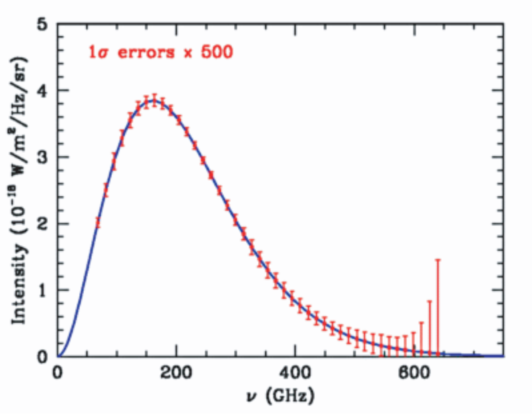


FIGURE 102 The measured spectrum of the cosmic background radiation, with the error bars multiplied by 500, compared to the calculated Planck spectrum for 2.728 K (NASA).

be filled with black-body radiation emitted when it was 'in heat'. That radiation, called the *background radiation*, must have cooled down due to the expansion of the universe. (Can you confirm this?) Despite various similar predictions by other authors, including Yakov Zel'dovich, in one of the most famous cases of missed scientific communication, the radiation was found only much later, by two researchers completely unaware of all this work. A famous paper in 1964 by Doroshkevich and Novikov had even stated that the antenna used by the (unaware) later discoverers was the best device to search for the radiation! In any case, only in 1965 did Arno Penzias and Robert Wilson discover the radiation. It was in one of the most beautiful discoveries of science, for which both later received the Nobel Prize in Physics. The radiation turns out to be described by the black-body radiation for a body with a temperature of 2.728(1) K, as illustrated in Figure 102. In fact, the spectrum follows the black-body dependence to a precision better than 1 part in 10^4.

Challenge 338 ny

Ref. 222

Ref. 223

In summary, data show that the universe started with a hot *big bang*. But apart from expansion and cooling, the past fourteen thousand million years have also produced a few other memorable events.

Do stars shine every night?

> Don't the stars shine beautifully? I am the only person in the world who knows why they do.
> Friedrich (Fritz) Houtermans (1903–1966)

Stars seem to be there for ever. In fact, every now and then a new star appears in the sky: a *nova*. The name is Latin and means 'new'. Especially bright novae are called *supernovae*. Novae and similar phenomena remind us that stars usually live much longer than humans, but that like people, stars are born, shine and die.

It turns out that one can plot all stars on the so-called *Hertzsprung–Russell diagram*. This diagram, central to every book on astronomy, is shown in Figure 103. It is a beautiful example of a standard method used by astrophysicists: collecting statistics over many examples of a type of object, one can deduce the life cycle of the object, even though their lifetime is much longer than that of a human. For example, it is possible, by clever use of

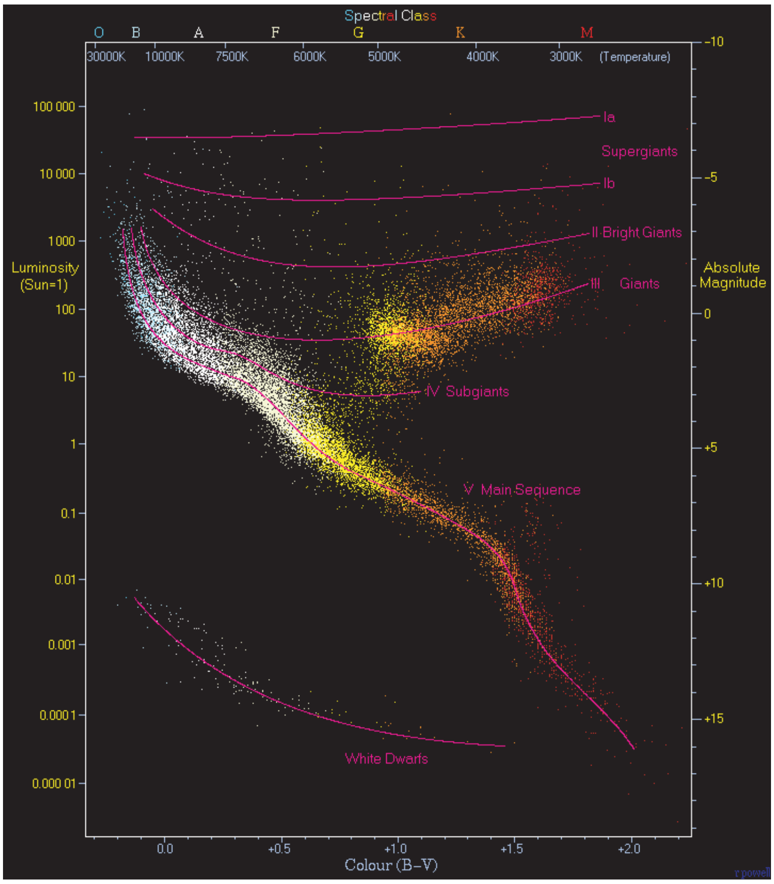


FIGURE 103 The Hertzsprung–Russell diagram (© Richard Powell).

the diagram, to estimate the age of stellar clusters, such as the M15 cluster of Figure 104, and thus arrive at a minimum age of the universe. The result is around thirteen thousand million years.

The finite lifetime of stars leads to restrictions on their visibility, especially for high red-shifts. Indeed, modern telescope can look at places (and times) so far in the past that they contained no stars yet. At those distances one only observes *quasars*; these light sources are not stars, but much more massive and bright systems. Their precise structure is still being studied by astrophysicists.

Since the stars shine, they were also *formed* somehow. Over millions of years, vast dust clouds in space can contract, due to the influence of gravity, and form a dense, hot and rotating structure: a new star. The fascinating details of their birth from dust clouds are Ref. 224 a central part of astrophysics, but we will not explore them here. Stars differ in evolution and lifetime. Above all, their evolution depends on their birth mass. Stars of the mass of the Sun live 10 to 20 Ga and die as red giants. Stars with a mass that is 20 times that of the Sun live only a few million years and die as supernovas. The most massive stars seem to have about 130 solar masses. Exceptions are those stars that form through merging of Ref. 225 several stars; they can be as massive as 250 solar masses.

Yet we do not have the full answer to our question. Why do stars shine at all? Clearly,

FIGURE 104 The Messier 15 (M15) globular star cluster, with an age of thirteen thousand million years (© ESA, NASA).

they shine because they are hot. They are hot because of nuclear reactions in their interior. We will discuss these processes in more detail in a latter volume. Vol. V, page 209

A short history of the universe

> Anima scintilla stellaris essentiae.*
> Heraclitus of Ephesus (*c.* 540 to *c.* 480 bce) Ref. 226

Not only stars are born, shine and die. Also galaxies do so. What about the universe as a whole? The most important adventures that the matter and radiation around us have experienced are summarized in Table 6. The steps not yet discussed will be studied in Ref. 227

---

* 'The soul is a spark of the substance of the stars.'

the rest of our adventure. The history table is awe-inspiring. The sequence of events is so beautiful and impressive that nowadays it is used in certain psychotherapies to point out to people the story behind their existence, and to remind them of their own worth. Enjoy.

TABLE 6 A short history of the universe.

Time before now[a]	Time from big bang[b]	Event	Temperature
c. $13.8 \cdot 10^9$ a	$\approx t_{\mathrm{Pl}}$ [b]	Time, space, matter and initial conditions are indeterminate	10^{32} K $\approx T_{\mathrm{Pl}}$
$13 \cdot 10^9$ a	*c.* $1000\ t_{\mathrm{Pl}}$ $\approx 10^{-42}$ s	Distinction of space-time from matter and radiation, initial conditions are determinate	10^{30} K
	10^{-35} s to 10^{-32} s	**Inflation & GUT epoch** starts; strong and electroweak interactions diverge	$5 \cdot 10^{26}$ K
	10^{-12} s	Antiquarks annihilate; electromagnetic and weak interaction separate	10^{15} K
	$2 \cdot 10^{-6}$ s	Quarks get confined into hadrons; universe is a plasma	10^{13} K
		Positrons annihilate	
	0.3 s	Universe becomes transparent for neutrinos	10^{10} K
	a few seconds	**Nucleosynthesis**: D, ^{4}He, ^{3}He and ^{7}Li *nuclei* form; radiation still dominates	10^9 K
	2500 a	**Matter domination** starts; density perturbations magnify	75 000 K
red-shift $z = 1100$	380 000 a	**Recombination**: during these latter stages of the big bang, H, He and Li *atoms* form, and the universe becomes ‘transparent’ for light, as matter and radiation decouple, i.e., as they acquire different temperatures; the ‘night’ sky starts to get darker and darker	3000 K
		Sky is almost black except for black-body radiation	$T_\gamma = T_{o\gamma}(1+z)$
$z = 10$ to 30		**Galaxy formation**	
$z = 9.6$		Oldestobject seen so far	
$z = 5$		Galaxy clusters form	
$z = 3$	10^6 a	First generation of stars (population II) is formed, starting hydrogen fusion; helium fusion produces carbon, silicon and oxygen	
	$2 \cdot 10^9$ a	First stars explode as supernovae[c]; iron is produced	
$z = 1$	$3 \cdot 10^9$ a	Second generation of stars (population I) appears, and subsequent supernova explosions of the ageing stars form the trace elements (Fe, Se, etc.) we are made of and blow them into the galaxy	

TABLE 6 (Continued) A short history of the universe.

Time before now[a]	Time from big bang[b]	Event	Temperature
$4.7 \cdot 10^9$ a		Primitive cloud, made from such explosion remnants, collapses; **Sun forms**	
$4.5 \cdot 10^9$ a		**Earth** and other planet **formation**: Azoicum starts[d]	
$4.5 \cdot 10^9$ a		Moon forms from material ejected during the collision of a large asteroid with the still-liquid Earth	
$4.3 \cdot 10^9$ a		Craters form on the planets	
$4.0 \cdot 10^9$ a		**Archean** eon (Archaeozoicum) starts: bombardment from space stops; Earth's crust solidifies; oldest minerals form	
$3.8 \cdot 10^9$ a		end of water collection and condensation	
$3.5 \cdot 10^9$ a		Unicellular (microscopic) life appears; stromatolites form	
$2.5 \cdot 10^9$ a		**Proterozoic** eon ('age of first life') starts: atmosphere becomes rich in oxygen thanks to the activity of microorganisms Ref. 228	
$1.3 \cdot 10^9$ a		Macroscopic, multicellular life appears, fungi conquer land	
$800 \cdot 10^6$ a		Earth is completely covered with ice for the first time (reason still unknown) Ref. 229	
600 to $540 \cdot 10^6$ a		Earth is completely covered with ice for the last time	
$540(5) \cdot 10^6$ a		**Paleozoic** era (Palaeozoicum, 'age of old life') starts, after a gigantic ice age ends: animals appear, oldest fossils (with 540(5) start of Cambrian, 495(5) Ordovician, 440(5) Silurian, 417(5) Devonian, 354(5) Carboniferous and 292(5) Permian periods)	
$480 - 450 \cdot 10^6$ a		Land plants appear	
$400 - 370 \cdot 10^6$ a		Wooden trees appear, flying insects appear	
$250(5) \cdot 10^6$ a		**Mesozoic** era (Mesozoicum, 'age of middle life', formerly called Secondary) starts: most insects and other life forms are exterminated; mammals appear (with 250(5) start of Triassic, 205(4) Jurassic and 142(3) Cretaceous periods)	
$150 \cdot 10^6$ a		Continent Pangaea splits into Laurasia and Gondwana	
		The star cluster of the Pleiades forms	
$150 \cdot 10^6$ a		Birds appear	
$142(3) \cdot 10^6$ a		Golden time of dinosaurs (Cretaceous) starts	
$100 \cdot 10^6$ a		Start of formation of Alps, Andes and Rocky Mountains	

TABLE 6 (Continued) A short history of the universe.

Time before now[a]	Time from big bang[b]	Event	Temperature
$65.5 \cdot 10^6$ a		**Cenozoic** era (Caenozoicum, 'age of new life') starts: after an asteroid hits the Earth in the Yucatan, dinosaurs become extinct, and grass and primates appear, (with 65.5 start of Tertiary, consisting of Paleogene period with Paleocene, 55.0 Eocene and 33.7 Oligocene epoch, and of Neogene period, with 23.8 Miocene and 5.32 Pliocene epoch; then 1.81 Quaternary period with Pleistocene (or Diluvium) and 0.01 Holocene (or Alluvium) epoch)	
$50 \cdot 10^6$ a		Large mammals appear	
$7(1) \cdot 10^6$ a		Hominids appears	
$3 \cdot 10^6$ a		Supernova explodes, with following consequences: more intense cosmic radiation, higher formation rate of clouds, Earth cools down drastically, high evolutionary pressure on the hominids and as a result, Homo appears Ref. 230	
500 000 a		Formation of youngest stars in galaxy	
500 000 a		Homo sapiens appears	
100 000 a		Beginning of last ice age	
90 000 a		Homo sapiens sapiens appears	
11 800 a		End of last ice age, start of Holocene	
6 000 a		First written texts	
2 500 a		Physics starts	
500 a		Use of coffee, pencil and modern physics starts	
200 a		Electricity use begins	
100 a		Einstein publishes	
10 to 120 a		You were a unicellular being	
Present	c. $14 \cdot 10^9$ a	You are reading this	$T_{o\gamma} = 2.73\,\mathrm{K}$ $T_{o\nu} \approx 1.6\,\mathrm{K}$ $T_{ob} \approx 0\,\mathrm{K}$
Future		You enjoy life; for details and reasons, see the following volumes.	

---

[a] The time coordinate used here is the one given by the coordinate system defined by the microwave background radiation, as explained on page 237. A year is abbreviated 'a' (Latin 'annus'). Errors in the last digits are given between parentheses. Sometimes the red-shift z is given instead of the time coordinate.

[b] This quantity is not exactly defined since the big bang is not a space-time event. This issue will be explored later on. Vol. VI, page 92

[c] The history of the atoms on Earth shows that we are made from the leftovers of a supernova. We truly are made of *stardust*.

[d] Apart from the term Azoicum, all other names and dates from the geological time scale are those of the International Commission on Stratigraphy; the dates are measured with the help of radioactive dating. Vol. V, page 182

Despite its length and its interest, the history table has its limitations: what happened elsewhere in the last few thousand million years? There is still a story to be written of which next to nothing is known. For obvious reasons, investigations have been rather Earth-centred.

Discovering and understanding all phenomena observed in the skies is the aim of astrophysics research. In our adventure we have to skip most of this fascinating topic, because we want to focus on motion. Interestingly, general relativity allows us to explain many of the general observations about motion across the universe in a simple manner.

THE HISTORY OF SPACE-TIME

> “A number of rabbits run away from a central point in various directions, all with the same speed. While running, one rabbit turns its head, and makes a startling observation. Which one?”

Challenge 339 s

Page 225

The data showing that the universe is sprinkled with stars all over lead to a simple conclusion: *the universe cannot be static.* Gravity always changes the distances between bodies; the only exceptions are circular orbits. Gravity also changes the *average* distances between bodies: gravity always tries to collapse clouds. The biggest cloud of all, the one formed by all the matter in the universe, must therefore be changing: either it is collapsing, or it is still expanding.

Ref. 231

The first to dare to draw this conclusion was Aleksander Friedmann.* In 1922 he deduced the possible evolutions of the universe in the case of homogeneous, isotropic mass distribution. His calculation is a classic example of simple but powerful reasoning. For a universe which is homogeneous and isotropic for every point, the line element of space-time is given by

Challenge 340 ny

$$ds^2 = c^2 dt^2 - a^2(t)(dx^2 + dy^2 + dz^2) \,. \qquad (246)$$

The quantity $a(t)$ is called the *scale factor*. The scale factor is often called, sloppily, the ‘radius’ or the ‘size’ of the universe. Matter is described by a density ρ_M and a pressure p_M. Inserting all this into the field equations, we get two equations that any school student can grasp; they are

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{kc^2}{a^2} = \frac{8\pi G}{3}\rho_M + \frac{\Lambda c^2}{3} \qquad (247)$$

---

* Aleksander Aleksandrowitsch Friedmann (1888–1925) was the first physicist who predicted the expansion of the universe. Following his early death from typhus, his work remained almost unknown until Georges A. Lemaître (b. 1894 Charleroi, d. 1966 Leuven), both priest and cosmologist, took it up and expanded it in 1927, focusing on solutions with an initial singularity. Lemaître was one of the propagators of the (erroneous!) idea that the big bang was an ‘event’ of ‘creation’ and convinced his whole religious organization of it. The Friedmann–Lemaître solutions are often erroneously called after two other physicists, who studied them again much later, in 1935 and 1936, namely H.P. Robertson and A.G. Walker.

Page 248, page 248

and

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{kc^2}{a^2} = -\frac{8\pi G}{c^2}p + \Lambda c^2 \,. \tag{248}$$

Together, they imply the two equations

$$\ddot{a} = -\frac{4\pi G}{3}(\rho_M + 3p_M/c^2)\; a + \frac{\Lambda c^2}{3}\; a \tag{249}$$

and

$$\dot{\rho}_M = -3\frac{\dot{a}}{a}(\rho_M + p_M/c^2)\;, \tag{250}$$

where the dot indicates the derivative with respect to time. Equations (249) and (250) depend on only three constants of nature: the gravitational constant G, related to the maximum force or power in nature, the speed of light c, and the cosmological constant Λ, describing the energy density of the vacuum, or, if one prefers, the smallest force in nature. Equation (249) expresses, in unusual form, the conservation of energy, i.e., the first law of thermodynamics. Energy conservation is already implied in the definition of the metric used by Friedmann. Equation (250) expresses that the cosmological constant Λ *accelerates* the expansion $\dot{a}$ and that matter, through gravity, *decelerates* the expansion $\dot{a}$ of the universe.

Page 134

Challenge 341 e

Before we discuss the equations, first a few points of vocabulary. In the following, the index 0 refers to the present time. At the present time t_0, the pressure of matter is negligible. In this case, the expression $\rho_M a^3$ is constant in time. The present-time Hubble parameter is defined by $H_0 = \dot{a}_0/a_0$. It describes the expansion speed of the universe – if you prefer, the rabbit speed in the puzzle above. It is customary to relate all mass densities to the so-called *critical mass density* ρ_c given by

Challenge 342 e

Challenge 343 ny

$$\rho_c = \frac{3H_0^2}{8\pi G} \approx (8 \pm 2) \cdot 10^{-27}\,\text{kg/m}^3 \tag{251}$$

corresponding to about 8, give or take 2, hydrogen atoms per cubic metre. The actual density of the universe is not far from this value. On Earth, we would call this value an extremely good *vacuum*. Such are the differences between everyday life and the universe as a whole. In any case, the critical density characterizes a matter distribution leading to an evolution of the universe just between never-ending expansion and collapse. In fact, this density is the critical one, leading to a so-called *marginal* evolution, only in the case of *vanishing* cosmological constant. Despite this restriction, the term 'critical mass density' is now used in all other cases as well. We can thus speak of a dimensionless mass density Ω_M defined as

$$\Omega_M = \rho_0/\rho_c\,. \tag{252}$$

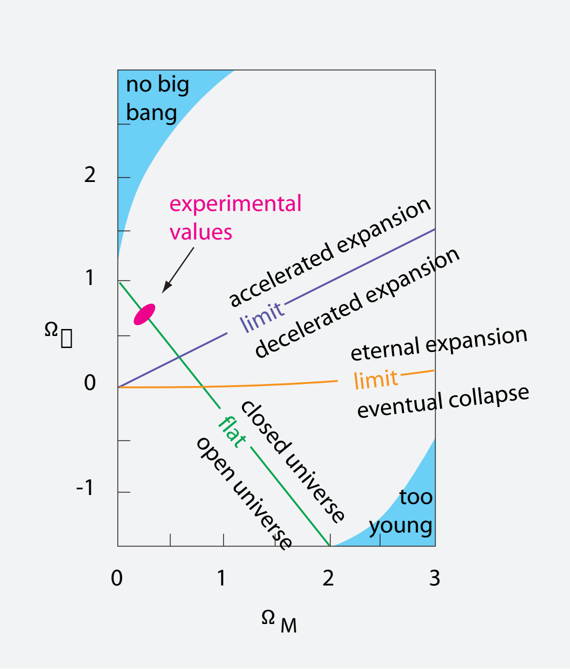


FIGURE 105 The ranges for the Ω parameters and their consequences.

The cosmological constant can also be related to this critical density by setting

$$\Omega_\Lambda = \frac{\rho_\Lambda}{\rho_c} = \frac{\Lambda c^2}{8\pi G \rho_c} = \frac{\Lambda c^2}{3H_0^2} \; . \tag{253}$$

A third dimensionless parameter Ω_K describes the curvature of space. It is defined in terms of the present-day radius of the universe R_0 and the curvature constant $k = \{1, -1, 0\}$ as

$$\Omega_K = \frac{-k}{R_0^2\, H_0^2} \tag{254}$$

and its sign is opposite to the one of the curvature k; Ω_K vanishes for vanishing curvature. Note that a *positively* curved universe, when homogeneous and isotropic, is necessarily closed and of finite volume. A *flat* or *negatively* curved universe with the same matter distribution can be open, i.e., of infinite volume, but does not need to be so. It could even be simply or multiply connected. In these cases the topology is not completely fixed by the curvature.

Challenge 344 ny

As already mentioned, the present-time Hubble parameter is defined by $H_0 = \dot{a}_0/a_0$. From equation (247) we then get the central relation

$$\Omega_M + \Omega_\Lambda + \Omega_K = 1 \; . \tag{255}$$

In the past, when data were lacking, cosmologists were divided into two camps: the *claustrophobics* believing that $\Omega_K > 0$ and the *agoraphobics* believing that $\Omega_K < 0$. More details about the measured values of these parameters will be given shortly. The diagram of Figure 105 shows the most interesting ranges of parameters together with the corresponding behaviours of the universe. Modern measurements are consistent with a flat universe, thus with $\Omega_K = 0$.

For the Hubble parameter, the most modern measurements give a value of

$$H_0 = 71 \pm 4\,\mathrm{km/sMpc} = 2.3 \pm 2 \cdot 10^{-18}\,/\mathrm{s} \qquad (256)$$

which corresponds to an *age of the universe* of 13.8 ± 1 thousand million years. In other words, the age deduced from the history of space-time agrees with the age, given above, deduced from the history of stars.

To get a feeling of how the universe evolves, it is customary to use the so-called *deceleration parameter* q_0. It is defined as

$$q_0 = -\frac{\ddot{a}_0}{a_0 H_0^2} = \frac{1}{2}\Omega_M - \Omega_\Lambda\ . \qquad (257)$$

The parameter q_0 is positive if the expansion is slowing down, and negative if the expansion is accelerating. These possibilities are also shown in the diagram of Figure 105.

An even clearer way to picture the expansion of the universe for vanishing pressure is to rewrite equation (247) using $\tau = t\,H_0$ and $x(\tau) = a(t)/a(t_0)$, yielding

$$\left(\frac{\mathrm{d}x}{\mathrm{d}\tau}\right)^2 + U(x) = \Omega_K$$
$$\text{where}\quad U(x) = -\Omega_\Lambda x - \Omega_\Lambda x^2\ . \qquad (258)$$

This looks like the evolution equation for the motion of a particle with mass 1, with total energy Ω_K in a potential $U(x)$. The resulting evolutions are easily deduced.

For *vanishing* Ω_Λ, the universe either expands for ever, or recollapses, depending on the value of the mass–energy density. For *non-vanishing*, positive Ω_Λ, the potential has exactly one maximum; if the particle has enough energy to get over the maximum, it will accelerate continuously. Data shows that this is the situation the universe seems to be in today. Either case tells:

▷ General relativity and the black night sky imply that the universe is expanding.

In other words, the universe is *not* static. This was Friedmann's daring conclusion. For a certain time range, the resulting expansion is shown in Figure 106. We note that due to its isotropic expansion, the universe has a preferred reference frame: the frame defined by average matter. The time measured in that frame is the time listed in Table 6 and in Figure 106, and it is time we assume when we talk about the *age* of the universe.

▷ General relativity and the black night sky imply that the universe once was extremely small and then expanded rapidly. The very early evolution

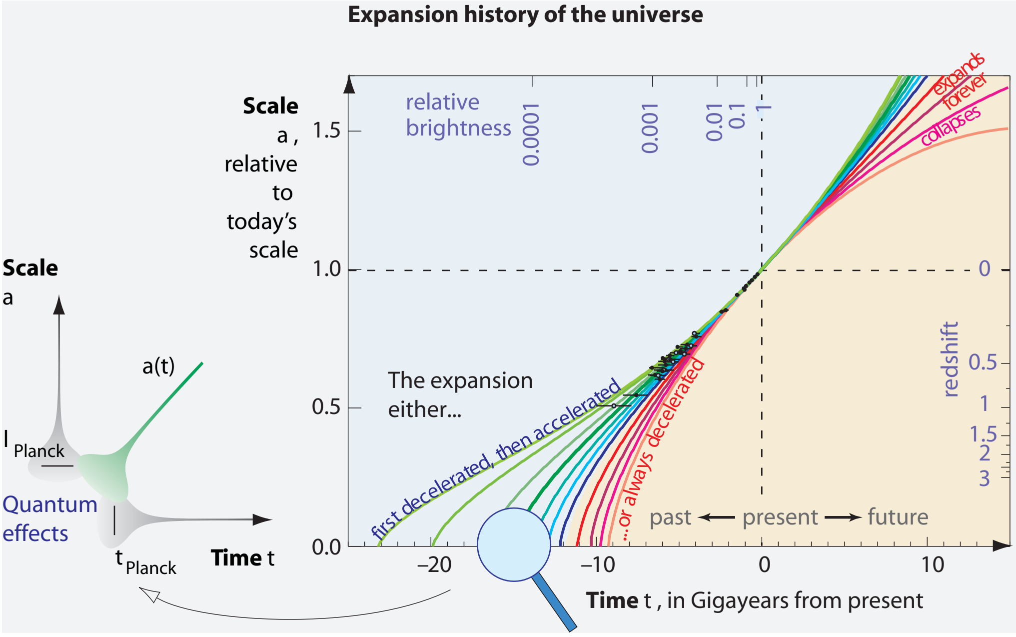


FIGURE 106 The evolution of the universe's scale a for different values of its mass density, as well as the measured data (the graph on the right is courtesy of Saul Perlmutter and the Supernova Cosmology Project).

is called the *big bang*.

There are two points to be noted: first the set of possible evolution curves is described by *two* parameters, not one. In addition, lines *cannot* be drawn down to *zero* size, but only to very small sizes. There are two main reasons: we do not yet understand the behaviour of matter at very high energy, and we do not understand the behaviour of space-time at very high energy. We return to this important issue later on.

In summary, the main conclusion from Friedmann's work is that a homogeneous and isotropic universe is *not static*: it either expands or contracts. In either case, the universe has a *finite age*. These profound ideas took many years to spread around the cosmology community; even Einstein took a long time to get accustomed to them.

An overview of the possibilities for the *long-time evolution* is given in Figure 107. The evolution can have various outcomes. In the early twentieth century, people decided among them by personal preference. Albert Einstein first preferred the solution $k = 1$ and $\Lambda = a^{-2} = 4\pi G\rho_M$. It is the unstable solution found when $x(\tau)$ remains at the top of the potential $U(x)$.

Willem de Sitter had found in 1917, much to Einstein's personal dismay, that an *empty* universe with $\rho_M = p_M = 0$ and $k = 1$ is also possible. This type of universe expands for large times. The De Sitter universe shows that in special cases, matter is not needed for space-time to exist! Challenge 345 ny

Lemaître had found expanding universes for positive mass, and his results were also contested by Einstein at first. When later the first measurements confirmed the calculations, the idea of a massive and expanding universe became popular. It then became the

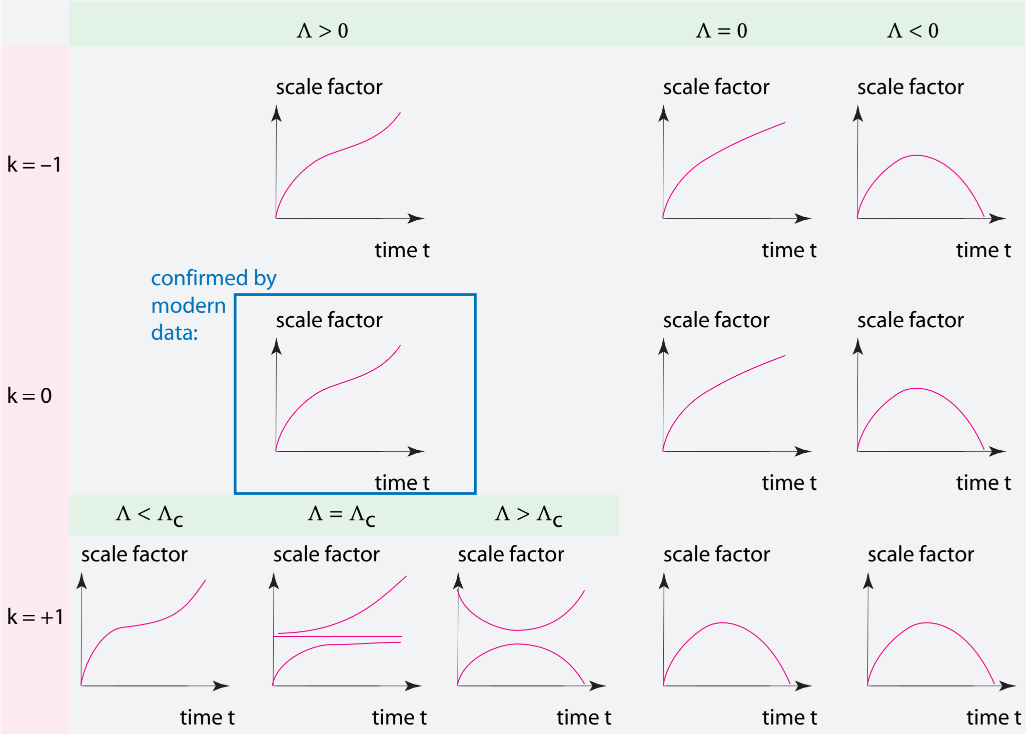


FIGURE 107 The long-term evolution of the universe's scale factor a for various parameters.

concordance model in textbooks.

In a sort of collective blindness that lasted from around 1950 to 1990, almost all cosmologists believed that $\Lambda = 0$.* Only towards the end of the twentieth century did experimental progress allow cosmologists to make statements based on evidence rather than beliefs or personal preferences, as we will find out shortly. But first of all we will settle an old issue.

Why is the sky dark at night?

> In der Nacht hat ein Mensch nur ein Nachthemd an, und darunter kommt gleich der Charakter.**
>
> Rober Musil

First of all, the sky is not black at night – it is dark blue. Seen from the surface of the Earth, it has the same blue colour as during the day, as any long-exposure photograph, such as Figure 108, shows. The blue colour of the night sky, like the colour of the sky during the day, is due to light from the stars that is scattered by the atmosphere. If we want to know the real colour of the sky, we need to go *above* the atmosphere. There, to the

Challenge 346 ny

---

* In this case, for $\Omega_M \geqslant 1$, the age of the universe follows $t_0 \leqslant 2/(3H_0)$, where the limits correspond. For vanishing mass density we have $t_0 = 1/H_o$.

** 'At night, a person is dressed only with a nightgown, and directly under it there is the character.' Robert Musil (b. 1880 Klagenfurt, d. 1942 Geneva), writer.

FIGURE 108 All colours, such as the blue of the sky, are present also at night, as this long-time exposure shows. On the top left, the bright object is Mars; the lower half shows a rare coloured fog bow created by moonlight (© Wally Pacholka).

eye, the sky is pitch black. But precise measurements show that even the empty sky is not completely black at night; it is filled with radiation of around 200 GHz; more precisely, it is filled with radiation that corresponds to the thermal emission of a body at 2.73 K. This *cosmic background radiation* is the thermal radiation left over from the big bang.

Ref. 232 Thus the universe is indeed colder than the stars. But why is this so? If the universe were homogeneous on large scales and also infinitely large, it would have an infinite number of stars. Looking in any direction, we would see the surface of a star. The night sky would be as bright as the surface of the Sun! Can you convince your grandmother about this? Challenge 347 s

In a deep forest, we see a tree in every direction, as shown in Figure 109. Similarly, in a 'deep' universe, we would see a star in every direction. Now, the average star has a surface temperature of about 6000 K. If we lived in a deep and old universe, we would effectively live inside an oven with a temperature of around 6000 K! Such a climate would make it difficult to enjoy ice cream.

So why is the sky *black* at night, despite being filled with radiation from stars at 6000 K, i.e., with *white* light? This paradox was most clearly formulated in 1823 by the astronomer Wilhelm Olbers.* Because he extensively discussed the question, it is also called *Olbers'*

---

* Heinrich Wilhelm Matthäus Olbers (b. 1758 Arbergen, d. 1840 Bremen) was an important astronomer. He discovered two planetoids, Pallas and Vesta, and five comets; he developed the method of calculating parabolic orbits for comets which is still in use today. Olbers also actively supported the mathematician and astronomer Friedrich Wilhelm Bessel in his career choice. The paradox is named after Olbers, though Vol. I, page 150

FIGURE 109 Top: in a deep, or even infinite forest, only trees are visible, and nothing behind them. Bottom: at night, we can see the stars but also what is behind, namely the black sky. The universe is thus of finite size. (© Aleks G, NASA/ESA)

paradox.

Today we know that two main effects explain the darkness of the night. First, since the universe is finite in age, distant stars are shining for less time. We see them in a younger stage or even during their formation, when they were darker. As a result, the share of brightness of distant stars is smaller than that of nearby stars, so that the average

---

others had made similar points before, such as the Swiss astronomer Jean Philippe Loÿs de Cheseaux in 1744 and Johannes Kepler in 1610.

temperature of the sky is reduced.* Today we know that even if all matter in the universe were converted into radiation, the universe would still not be as bright as just calculated. In other words, the power and lifetime of stars are much too low to produce the oven brightness just mentioned. Secondly, we can argue that the radiation of distant stars is red-shifted and that the volume that the radiation must fill is increasing continuously, so that the effective average temperature of the sky is also reduced. Ref. 233

Calculations are necessary to decide which reason for the darkness at night is the most important one. This issue has been studied in great detail by Paul Wesson; he explains that the first effect, darkness due to a maximum finite star lifetime, is larger than the second, darkness due to red-shift, by a factor of about three. However, both effects are themselves due to the finite age of the universe. We may thus correctly state that *the sky is dark at night because the universe has a finite age.* Ref. 234

We note that the darkness of the sky arises only because the speed of light is finite. Can you confirm this? Ref. 232 Challenge 349 e

The darkness of the sky also tells us that the universe has a finite age that is *large*. Indeed, the 2.7 K background radiation is that cold, despite having been emitted at 3000 K, because it is red-shifted, thanks to the Doppler effect. Under reasonable assumptions, the temperature T of this radiation changes with the scale factor $a(t)$ of the universe as Ref. 235

$$T \sim \frac{1}{a(t)} . \tag{259}$$

In a young universe, we would thus not be able to see the stars, even if they existed.

From the brightness of the sky at night, measured to be about $3 \cdot 10^{-13}$ times that of an average star like the Sun, we can deduce something interesting: the density of stars in the universe must be much smaller than in our galaxy. The density of stars in the galaxy can be deduced by counting the stars we see at night. But the average star density in the galaxy would lead to much higher values for the night brightness if it were constant throughout the universe. We can thus deduce that the galaxy is much *smaller* than the universe simply by measuring the brightness of the night sky and by counting the stars in the sky. Can you make the explicit calculation? Ref. 233 Challenge 350 e

In summary, the sky is black, or better, very dark at night because space-time and matter are of *finite, but old age*. As a side issue, here is a quiz: is there an Olbers' paradox also for gravitation? Challenge 351 ny

THE COLOUR VARIATIONS OF THE NIGHT SKY

Not only is the night sky not black; the darkness of the night sky even depends on the direction one is looking.

Since the Earth is moving when compared to the average stars, the dark colour of the sky shows a Doppler shift. But even when this motion is compensated some colour variations remain. The variations are tiny, but they can be measured with special satellites. The most precise results are those taken in 2013 by the European Planck satellite; they

---

* Can you explain that the sky is not black just because it is painted black or made of black chocolate? Or more generally, that the sky is not made of and does not contain any dark and cold substance, as Olbers himself suggested, and as John Herschel refuted in 1848? Challenge 348 ny

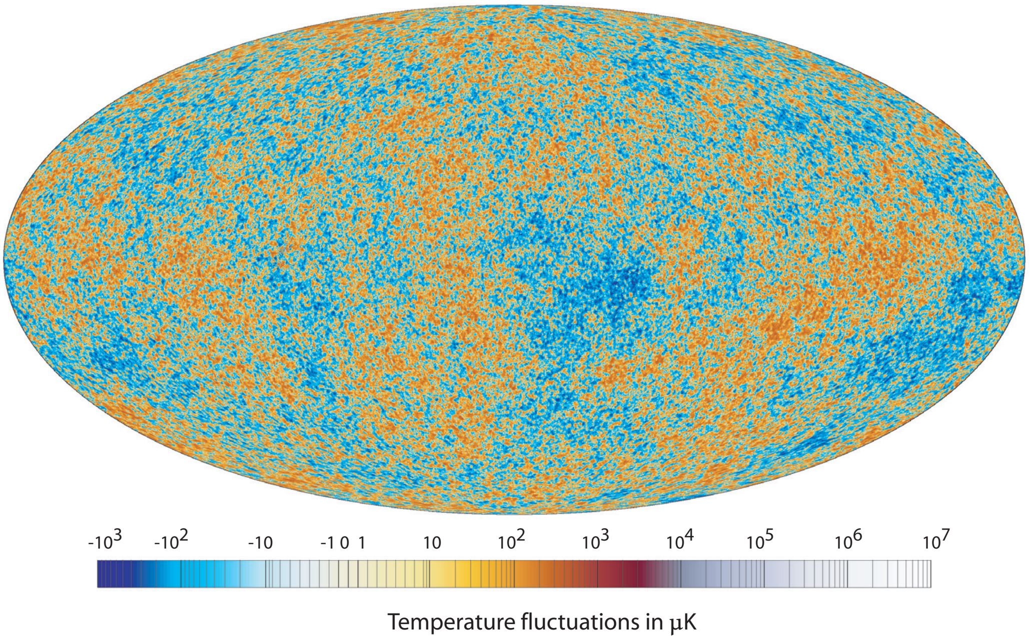


FIGURE 110 A false colour image of the *fluctuations* of the cosmic background radiation, after the Doppler shift from our local motion and the signals from the Milky Way have been subtracted (© Planck/ESA).

are shown in Figure 110. These temperature variations are in the microkelvin range; they show that the universe had already some inhomogeneities when the detected light was emitted. Figure 110 thus gives an impression of the universe when it was barely 380 000 years 'young'.

The data of Figure 110 is still being studied in great detail. It allows researchers to deduce the precise age of the universe – 13.8 Ga – its composition, and many other aspects. These studies are still ongoing.

Is the universe open, closed or marginal?

> “ – Doesn't the vastness of the universe make you feel small?
> – I can feel small without any help from the universe. ”
>
> Anonymous

Sometimes the history of the universe is summed up in two words: *bang!...crunch*. But will the universe indeed recollapse, or will it expand for ever? Or is it in an intermediate, marginal situation? The parameters deciding its fate are the mass density and cosmological constant.

The main news of the last decade of twentieth-century astrophysics are the experimental results allowing one to determine all these parameters. Several methods are being used. The first method is obvious: determine the speed and distance of distant stars. For large distances, this is difficult, since the stars are so faint. But it has now become possible to search the sky for supernovae, the bright exploding stars, and to determine their distance from their brightness. This is presently being done with the help of computerized searches of the sky, using the largest available telescopes. Ref. 236

A second method is the measurement of the anisotropy of the cosmic microwave background. From the observed power spectrum as a function of the angle, the curvature of space-time can be deduced.

A third method is the determination of the mass density using the gravitational lensing effect for the light of distant quasars bent around galaxies or galaxy clusters. Page 252

A fourth method is the determination of the mass density using galaxy clusters. All these measurements are expected to improve greatly in the years to come.

At present, these four completely independent sets of measurements provide the values Ref. 237

$$\Omega_{\mathrm{M}} \approx 0.3 \quad , \quad \Omega_{\Lambda} \approx 0.7 \quad , \quad \Omega_{\mathrm{K}} \approx 0.0 \tag{260}$$

where the errors are of the order of 0.1 or less. The values imply that

▷ The universe is spatially flat, its expansion is accelerating and there will be no big crunch.

However, no definite statement on the topology is possible. We will return to this last issue shortly. Page 254

In particular, the data show that the density of matter, including all dark matter, is only about one third of the critical value.* Over two thirds are given by the cosmological term. For the cosmological constant Λ the present measurements yield

$$\Lambda = \Omega_{\Lambda} \frac{3H_0^2}{c^2} \approx 10^{-52}\,/\mathrm{m}^2 \,. \tag{261}$$

This value has important implications for quantum theory, since it corresponds to a vacuum energy density

$$\rho_{\Lambda} c^2 = \frac{\Lambda c^4}{8\pi G} \approx 0.5\,\mathrm{nJ/m}^3 \approx \frac{10^{-46}\,(\mathrm{GeV})^4}{(\hbar c)^3} \,. \tag{262}$$

But the cosmological term also implies a negative vacuum pressure $p_{\Lambda} = -\rho_{\Lambda} c^2$. Inserting this result into the relation for the potential of universal gravity deduced from relativity Page 199

$$\Delta\varphi = 4\pi G(\rho + 3p/c^2) \tag{263}$$

---

* The difference between the total matter density and the separately measurable baryonic matter density, only about one sixth of the former value, is also not explained yet. It might even be that the universe contains matter of a type unknown so far. We can say that the universe is not WYSIWYG; there is *invisible*, or *dark* matter. This issue, the *dark matter problem*, is one of the important unsolved questions of cosmology.

Ref. 238 we get

$$\Delta\varphi = 4\pi G(\rho_{\mathrm{M}} - 2\rho_{\Lambda}) . \tag{264}$$

Challenge 352 ny Thus the gravitational acceleration around a mass M is

$$a = \frac{GM}{r^2} - \frac{\Lambda}{3}c^2 r = \frac{GM}{r^2} - \Omega_{\Lambda} H_0^2\, r , \tag{265}$$

which shows that a *positive* vacuum energy indeed leads to a *repulsive* gravitational effect. Inserting the mentioned value (261) for the cosmological constant Λ we find that the repulsive effect is negligibly small even for the distance between the Earth and the Sun. In fact, the order of magnitude of the repulsive effect is so much smaller than that of attraction that one cannot hope for a direct experimental confirmation of this deviation Challenge 353 ny from universal gravity at all. Probably astrophysical determinations will remain the only possible ones. In particular, a positive gravitational constant manifests itself through a positive component in the expansion rate.

But the situation is puzzling. The origin of the cosmological constant is *not* explained by general relativity. This mystery will be solved only with the help of quantum theory. In fact, the cosmological constant is the first and so far the only local and quantum aspect of nature detected by astrophysical means.

WHY IS THE UNIVERSE TRANSPARENT?

Could the universe be filled with water, which is transparent, as maintained by some Ref. 239 popular books in order to explain rain? No. Even if the universe were filled with air, the total mass would never have allowed the universe to reach the present size; it would have Challenge 354 ny recollapsed much earlier and we would not exist.

The universe is thus transparent because it is mostly empty. But *why* is it so empty? First of all, in the times when the size of the universe was small, all antimatter annihilated with the corresponding amount of matter. Only a tiny fraction of matter, which originally was slightly more abundant than antimatter, was left over. This 10^{-9} fraction is the matter Vol. V, page 255 we see now. As a consequence, there are 10^9 as many photons in the universe as electrons or quarks.

In addition, 380 000 years after antimatter annihilation, all available nuclei and electrons recombined, forming atoms, and their aggregates, like stars and people. No free charges interacting with photons were lurking around any more, so that from that period onwards light could travel through space as it does today, being affected only when it hits a star or a dust particle or some other atom. The observation of this cosmic background radiation shows that light can travel for over 13 000 million years without problems or disturbance. Indeed, if we recall that the average density of the universe is $10^{-26}\,\mathrm{kg/m^3}$ and that most of the matter is lumped by gravity in galaxies, we can imagine what an excellent vacuum lies in between. As a result, light can travel along large distances without noticeable hindrance.

But why is the vacuum transparent? That is a deeper question. Vacuum is transparent because it contains no electric charges and no horizons: charges or horizons are indispensable in order to absorb light. In fact, quantum theory shows that vacuum does

Vol. V, page 122 contain so-called *virtual* charges. However, these virtual charges have no effects on the transparency of vacuum.

THE BIG BANG AND ITS CONSEQUENCES

> “Μελέτη θανάτου. Learn to die.
> Plato, *Phaedo*, 81a.”

Above all, the hot big bang model, which is deduced from the colour of the stars and galaxies, states that about fourteen thousand million years ago the whole universe was extremely small. This fact gave the big bang its name. The term was created (with a sarcastic undertone) in 1950 by Fred Hoyle, who by the way never believed that it applies to nature. Nevertheless, the term caught on. Since the past smallness of the universe cannot be checked directly, we need to look for other, verifiable consequences. The main consequences are the following:

Page 227 Vol. III, page 337 Ref. 240

- All matter moves away from all other matter. This point was observed before the model was proposed.
- The maximal age for any system in the universe is finite. Recently, it was found that the maximal age is 13.8(1) Ga, around fourteen thousand million years.
- There is thermal background radiation. The observed temperature T_γ of about 2.7 K was found independently of the big bang model; it agrees with deductions from the maximal age value.
- The mass of the universe is made up of about 75 % hydrogen and 23 % helium. These values agree with the expectations.
- For non-vanishing cosmological constant Λ, the expansion of the universe accelerates. The acceleration has been observed, though its value cannot be predicted.
- For non-vanishing cosmological constant, universal gravity is slightly reduced. This point has yet to be confirmed.
- There are background neutrinos with a temperature T_ν of about 2 K; the precise prediction is $T_\nu/T_\gamma \approx (4/11)^{1/3}$ and that these neutrinos appeared about 0.3 s after the big bang. This point has yet to be confirmed.

It must be stressed that these consequences *confirm* the hot big bang model, but that historically, only the value of the background temperature was *predicted* from model. The last two points, on the temperature of neutrinos and on the deviation from universal gravity, are also true predictions, but they have not been confirmed yet. Technology will probably not allow us to check these two predictions in the foreseeable future. On the other hand, there is also no evidence against them.

Competing descriptions of the universe that avoid a hot early phase have not been too successful in matching observations. It could always be, however, that this might change in the future. Ref. 240

In addition, mathematical arguments state that with matter distributions such as the one observed in the universe, together with some rather weak general assumptions, there is no way to avoid a period in the *finite* past in which the universe was extremely small and hot. Therefore it is worth having a closer look at the situation. Ref. 241

WAS THE BIG BANG A BIG BANG?

First of all, was the big bang a kind of explosion? This description implies that some material transforms internal energy into motion of its parts. However, there was no such process in the early history of the universe. In fact, a better description is that space-time is expanding, rather than matter moving apart. The mechanism and the origin of the expansion is *unknown* at this point of our adventure. Because of the importance of spatial expansion, the whole phenomenon *cannot* be called an explosion. And obviously there neither was nor is any sound carrying medium in interstellar space, so that one cannot speak of a 'bang' in any sense of the term.

Was the big bang big? About fourteen thousand million years ago, the visible universe was rather small; much smaller than an atom. In summary, the big bang was neither big nor a bang; but the rest is correct.

WAS THE BIG BANG AN EVENT?

> Quid faciebat deus, antequam faceret caelum et terram? ...Non faciebat aliquid.*
> Augustine of Hippo, *Confessiones*, XI, 12.

The big bang theory is a description of what happened in the *whole* of space-time. Despite what is often written in careless newspaper articles, at every moment of the expansion space has been of non-vanishing size: space was *never* a single point. People who pretend it was are making ostensibly plausible, but false statements. The big bang theory is a description of the *expansion* of space-time, not of its beginning. Following the motion of matter back in time – even neglecting the issue of measurement errors – general relativity can deduce the existence of an initial singularity only if point-like matter is assumed to exist. However, this assumption is wrong. In addition, the effect of the non-linearities in general relativity at situations of high energy densities is not even completely clarified yet. Above all, the big bang occurred across the whole universe. (This is the reason that researchers ponder 'inflation' to explain various aspects of the universe.) In short, *the big bang was no event.*

Most importantly, quantum theory shows that the big bang was *not* a true singularity, as no physical observable, neither density nor temperature, ever reaches an infinitely large (or infinitely small) value. Such values cannot exist in nature.** In any case, there is a general agreement that arguments based on *pure* general relativity alone cannot make correct statements about the big bang. Nevertheless, most statements in newspaper articles are of this sort.

Vol. VI, page 102

WAS THE BIG BANG A BEGINNING?

> In the beginning there was nothing, which exploded.
> Terry Pratchett, *Lords and Ladies.*

---

* 'What was god doing before he made heaven and earth? ...He didn't do anything.' Augustine of Hippo (b. 354 Tagaste, d. 430 Hippo Regius) was an reactionary and influential theologian.

** Many physicists are still wary of making such strong statements on this point. The final part of our adventure gives the precise arguments leading to the conclusion.

Vol. VI, page 57

Asking what was before the big bang is like asking what is north of the North Pole. Just as nothing is north of the North Pole, so nothing 'was' before the big bang. This analogy could be misinterpreted to imply that the big bang took its start at a single point in time, which of course is incorrect, as just explained. But the analogy is better than it looks: in fact, there is *no* precise North Pole, since quantum theory shows that there is a fundamental indeterminacy as to its position. There is also a corresponding indeterminacy for the big bang.

In fact, it does not take more than three lines to show with quantum theory that time and space are *not* defined either at or near the big bang. We will give this simple argument in the first chapter of the final part of our adventure. The big bang therefore cannot be called a 'beginning' of the universe. There never was a time when the scale factor $a(t)$ of the universe was zero. Vol. VI, page 65

The conceptual mistake of stating that time and space exist from a 'beginning' onwards is frequently encountered. In fact, quantum theory shows that near the big bang, events can *neither* be ordered *nor* even be defined. More bluntly, there is *no* beginning; there has never been an initial event or singularity.

Obviously the concept of time is not defined 'outside' or 'before' the existence of the universe; this fact was already clear to thinkers over a thousand years ago. It is then tempting to conclude that time must have *started*. But as we saw, that is a logical mistake as well: first of all, there is no starting event, and secondly, time does not flow, as clarified already in the beginning of our walk. Ref. 242 Vol. I, page 48

A similar mistake lies behind the idea that the universe had certain 'initial conditions.' Initial conditions *by definition* make sense only for objects or fields, i.e., for entities which can be observed from the outside, i.e., for entities which have an environment. The universe does not comply with this requirement; it thus cannot have initial conditions. Nevertheless, many people still insist on thinking about this issue; interestingly, Stephen Hawking sold millions of copies of a book explaining that a description of the universe without initial conditions is the most appealing, without mentioning anywhere that there is no other possibility anyway.* Vol. I, page 237 Ref. 243

In summary, the big bang is not a beginning, nor does it imply one. We will uncover the correct way to think about it in the final part of our adventure. Vol. VI, page 306

DOES THE BIG BANG IMPLY CREATION?

> “[The general theory of relativity produces] universal doubt about god and his creation.”
> A witch hunter

Creation, i.e., the appearance of something out of nothing, needs an existing concept of space and time to make sense. The concept of 'appearance' makes no sense otherwise. But whatever the description of the big bang, be it classical, as in this chapter, or quantum mechanical, as in later ones, this condition is never fulfilled. Even in the present, classical description of the big bang, which gave rise to its name, there is *no* appearance of Vol. III, page 330

---

* This statement will still provoke strong reactions among physicists; it will be discussed in more detail in the section on quantum theory.

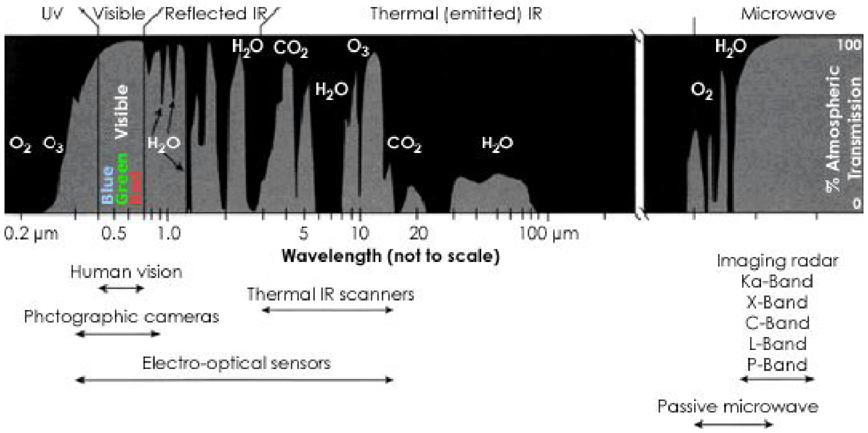


FIGURE 111 The transmittance of the atmosphere (NASA).

matter, nor of energy, nor of anything else. And this situation does not change in any later, improved description, as time or space are never defined *before* the appearance of matter.

Vol. VI, page 148

In fact, all properties of a creation are missing: there is no 'moment' of creation, no appearance from nothing, no possible choice of any 'initial' conditions out of some set of possibilities, and, as we will see in more detail in the last volume of this adventure, not even any choice of particular physical 'laws' from any set of possibilities.

Challenge 355 ny

In summary, the big bang does not imply nor harbour a creation process. The big bang was *not* an event, *not* a beginning and *not* a case of creation. It is impossible to continue our adventure if we do not accept each of these three conclusions. To deny them is to continue in the domain of beliefs and prejudices, thus effectively giving up on the mountain ascent.

Why can we see the Sun?

First of all, the Sun is visible because air is transparent. It is not self-evident that air is transparent; in fact it is transparent only to *visible* light and to a few selected other frequencies. Infrared and ultraviolet radiation are mostly absorbed. The reasons lie in the behaviour of the molecules the air consists of, namely mainly nitrogen, oxygen and a few other transparent gases. Several moons and planets in the solar system have opaque atmospheres: we are indeed lucky to be able to see the stars at all.

In fact, even air is not completely transparent; air molecules *scatter* light a little bit. That is why the sky and distant mountains appear blue and sunsets red. However, our eyes are not able to perceive this, and stars are invisible during daylight. At many wavelengths far from the visible spectrum the atmosphere is even opaque, as Figure 111 shows. (It is also opaque for all wavelengths shorter than 200 nm, up to gamma rays. On the long wavelength range, it remains transparent up to wavelength of around 10 to 20 m, depending on solar activity, when the extinction by the ionosphere sets in.)

Vol. III, page 239

Secondly, we can see the Sun because the Sun, like all hot bodies, *emits* light. We describe the details of *incandescence*, as this effect is called, later on.

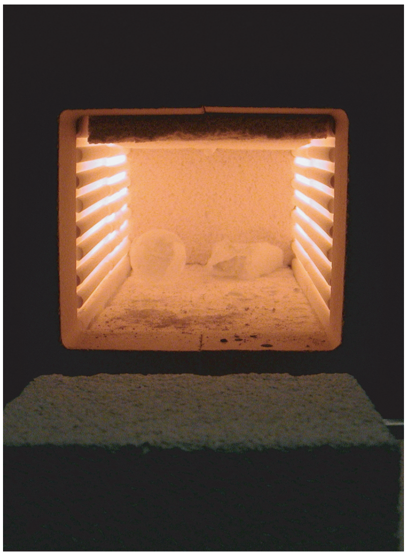


FIGURE 112 A hot red oven shows that at high temperature, objects and their environment cannot be distinguished from each other (© Wikimedia).

Thirdly, we can see the Sun because we and our environment and the Sun's environment are *colder* than the Sun. In fact, incandescent bodies can be distinguished from their background only if the background is colder. This is a consequence of the properties of incandescent light emission, usually called *black-body radiation*. The radiation is material-independent, so that for an environment with the same temperature as the body, nothing can be seen at all. Any oven, such as the shown in Figure 112 provides a proof.

Finally, we can see the Sun because it is not a black hole. If it were, it would emit (almost) no light.

Obviously, each of these conditions applies to stars as well. For example, we can only see them because the night sky is black. But then, how to explain the multicoloured sky?

WHY DO THE COLOURS OF THE STARS DIFFER?

Stars are visible because they emit visible light. We have encountered several important effects which determine colours: the diverse temperatures among the stars, the Doppler shift due to a relative speed with respect to the observer, and the gravitational red-shift.

Not all stars are good approximations to black bodies, so that the black-body radiation Vol. III, page 148 law does not always accurately describe their colour. However, most stars are reasonable approximations of black bodies. The temperature of a star depends mainly on its size, Ref. 244 its mass, its composition and its age, as astrophysicists are happy to explain. Orion is a good example of a coloured constellation: each star has a different colour. Long-exposure Vol. I, page 87 photographs beautifully show this.

The basic colour determined by temperature is changed by two effects. The first, the Challenge 356 ny *Doppler red-shift* z, depends on the speed v between source and observer as

$$z = \frac{\Delta\lambda}{\lambda} = \frac{f_S}{f_O} - 1 = \sqrt{\frac{c+v}{c-v}} - 1 \,. \qquad (266)$$

TABLE 7 The colour of the stars.

Class	Temperature	Example	Location	Colour
O	30 kK	Mintaka	δ Orionis	blue-violet
O	31(10) kK	Alnitak	ζ Orionis	blue-violet
B	22(6) kK	Bellatrix	γ Orionis	blue
B	26 kK	Saiph	κ Orionis	blue-white
B	12 kK	Rigel	β Orionis	blue-white
B	25 kK	Alnilam	ε Orionis	blue-white
B	17(5) kK	Regulus	α Leonis	blue-white
A	9.9 kK	Sirius	α Canis Majoris	blue-white
A	8.6 kK	Megrez	δ Ursae Majoris	white
A	7.6(2) kK	Altair	α Aquilae	yellow-white
F	7.4(7) kK	Canopus	α Carinae	yellow-white
F	6.6 kK	Procyon	α Canis Minoris	yellow-white
G	5.8 kK	Sun	ecliptic	yellow
K	3.5(4) kK	Aldebaran	α Tauri	orange
M	2.8(5) kK	Betelgeuse	α Orionis	red
D	<80 kK	–	–	any

Note. White dwarfs, or class-D stars, are remnants of imploded stars, with a size of only a few tens of kilometres. Not all are white; they can be yellow or red. They comprise 5 % of all stars. None is visible with the naked eye. Temperature uncertainties in the last digit are given between parentheses.

The size of the stars is an independent variable and is sometimes added as roman numerals at the end of the spectral type. (Sirius is an A1V star, Arcturus a K2III star.) Giants and supergiants exist in all classes from O to M.

To accommodate brown dwarfs, two new star classes, L and T, have been proposed.

Such shifts play a significant role only for remote, and thus faint, stars visible through the telescope. With the naked eye, Doppler shifts cannot be seen. But Doppler shifts can make distant stars shine in the infrared instead of in the visible domain. Indeed, the highest Doppler shifts observed for luminous objects are larger than 5.0, corresponding to a recessional speed of more than 94 % of the speed of light. In the universe, the red-shift is related to the scale factor $R(t)$ by

Challenge 357 ny

$$z = \frac{R(t_0)}{R(t_{\text{emission}})} - 1 \; . \tag{267}$$

Light at a red-shift of 5.0 was thus emitted when the universe was one sixth of its present age.

The other colour-changing effect, the *gravitational red-shift* z_g, depends on the matter density of the source and the light emission radius R; it is given by

$$z_g = \frac{\Delta\lambda}{\lambda} = \frac{f_S}{f_0} - 1 = \frac{1}{\sqrt{1 - \frac{2GM}{c^2 R}}} - 1 . \qquad (268)$$

Challenge 358 e It is usually quite a bit smaller than the Doppler shift. Can you confirm this?

No other red-shift processes are known; moreover, such processes would contradict Page 261 all the known properties of nature. But the colour issue leads to the next question.

ARE THERE DARK STARS?

It could be that some stars are not seen because they are dark. This could be one explanation for the large amount of dark matter seen in the recent measurements of the background radiation. This issue is currently of great interest and hotly debated. It is known that objects more massive than Jupiter but less massive than the Sun can exist in states which emit hardly any light. Any star with a mass below 7.2 % of the mass of the Sun cannot start fusion and is called a *brown dwarf*. It is unclear at present how many such objects exist. Many of the so-called extrasolar 'planets' are probably brown dwarfs. The issue is not yet settled.

Page 262 Another possibility for dark stars are black holes. These are discussed in detail below.

ARE ALL STARS DIFFERENT? – GRAVITATIONAL LENSES

> “Per aspera ad astra.[*] ”

Are we sure that at night, two stars are really different? The answer is no. Recently, it was shown that two 'stars' were actually two images of the same object. This was found by comparing the flicker of the two images. It was found that the flicker of one image was exactly the same as the other, just shifted by 423 days. This result was found by the Estonian astrophysicist Jaan Pelt and his research group while observing two images of Ref. 245 quasars in the system Q0957+561.

The two images are the result of *gravitational lensing*, an effect illustrated in Figure 113. Indeed, a large galaxy can be seen between the two images observed by Pelt, and much nearer to the Earth than the star. This effect had been already considered by Einstein; however he did not believe that it was observable. The real father of gravitational lensing Ref. 246 is Fritz Zwicky, who predicted in 1937 that the effect would be quite common and easy to observe, if lined-up galaxies instead of lined-up stars were considered, as indeed turned out to be the case.

Interestingly, when the time delay is known, astronomers are able to determine the Challenge 359 ny size of the universe from this observation. Can you imagine how?

If the two observed massive objects are lined up exactly behind each other, the more distant one is seen as *ring* around the nearer one. Such rings have indeed been observed, and the galaxy image around a central foreground galaxy at B1938+666, shown in Figure 114, is one of the most beautiful examples. In 2005, several cases of gravitational lens-

---

[*] 'Through hardship to the stars.' A famous Latin motto. Often incorrectly given as 'per ardua ad astra'.

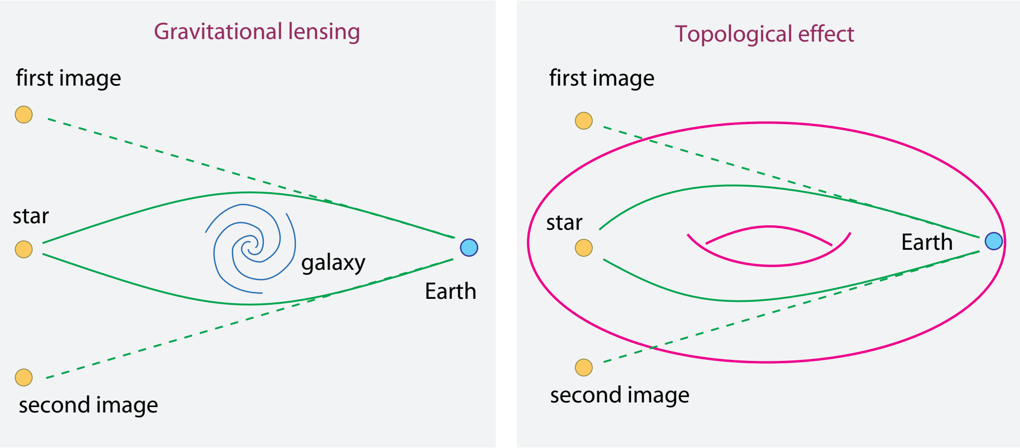


FIGURE 113 Two ways in which a single star can lead to several images.

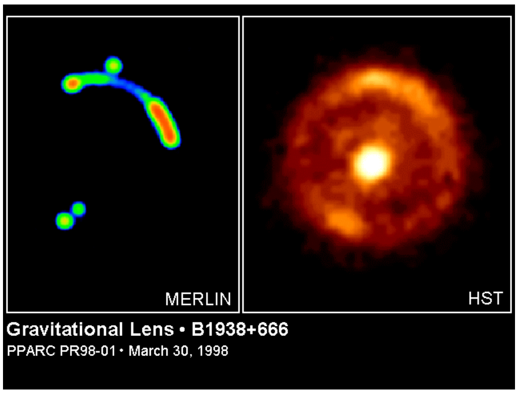


FIGURE 114 The Zwicky–Einstein ring B1938+666, seen in the radio spectrum (left) and in the optical domain (right) (NASA).

ing by stars were also discovered. More interestingly, three events where one of the two stars has a Earth-mass planet have also been observed. The coming years will surely lead to many additional observations, helped by the sky observation programme in the southern hemisphere that checks the brightness of about 100 million stars every night.

Generally speaking, images of nearby stars are truly unique, but for the distant stars the problem is tricky. For single stars, the issue is not so important, seen overall. Reassuringly, only about 80 multiple star images have been identified so far. But when whole galaxies are seen as several images at once (and several dozens are known so far) we might start to get nervous. In the case of the galaxy cluster CL0024+1654, shown in Figure 115, seven thin, elongated, blue images of the same distant galaxy are seen around the yellow, nearer, elliptical galaxies.

But multiple images can be created not only by gravitational lenses; the *shape* of the universe could also play some tricks.

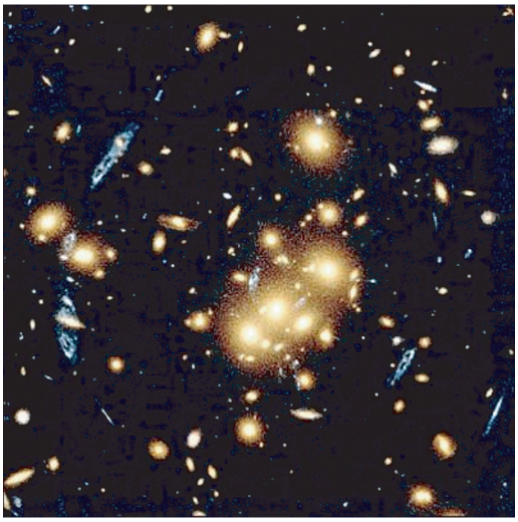


FIGURE 115 Multiple blue images of a galaxy formed by the yellow cluster CL0024+1654 (NASA).

What is the shape of the universe?

A popular analogy for the expansion of the universe is the comparison to a rubber balloon that increase in diameter by blowing air into it. The surface of the balloon is assumed to correspond to the volume of the universe. The dots on the balloon correspond to the galaxies; their distance continuously increases. The surface of the balloon is finite and has no boundary. By analogy, this suggests that the volume of the universe has a finite volume, but no boundary. This analogy presupposes that the universe has the same topology, the same 'shape' as that of a sphere with an additional dimension.

Ref. 247 But what is the experimental evidence for this analogy? Not much. Nothing definite is known about the shape of the universe. It is extremely hard to determine it, simply because of its sheer size. Experiments show that in the nearby region of the universe, say within a few million light years, the topology is simply connected. But for large distances, almost nothing is certain. Maybe research into gamma-ray bursts will tell us something about the topology, as these bursts often originate from the dawn of time.* Maybe even the study of fluctuations of the cosmic background radiation can tell us something. All this research is still in its infancy.

Since little is known, we can ask about the range of possible answers. As just mentioned, in the concordance model of cosmology, there are three options. For $k = 0$, compatible with experiments, the simplest topology of space is three-dimensional Euclidean space $\mathbb{R}^3$. For $k = 1$, space-time is usually assumed to be a product of linear time, with the topology R of the real line, and a sphere S^3 for space. That is the simplest possible shape, corresponding to a *simply-connected* universe. For $k = -1$, the simplest option for space is a hyperbolic manifold $\boldsymbol{H}^3$.

Page 236 In addition, Figure 105 showed that depending on the value of the cosmological constant, space could be finite and bounded, or infinite and unbounded. In most Friedmann–Lemaître calculations, simple-connectedness is usually tacitly assumed,

---

* The story is told from the mathematical point of view by Bob Osserman, *Poetry of the Universe*, 1996.

even though it is not at all required.

It could well be that space-time is *multiply* connected, like a higher-dimensional version of a torus, as illustrated on the right-hand side of Figure 113. A torus still has $k = 0$ everywhere, but a non-trivial global topology. For $k \neq 0$, space-time could also have even more complex topologies.* If the topology is non-trivial, it could even be that the actual number of galaxies is much smaller than the observed number. This situation would correspond to a kaleidoscope, where a few beads produce a large number of images.

In fact, the range of possibilities is not limited to the simply and multiply connected cases suggested by classical physics. If quantum effects are included, additional and much more complex options appear; they will be discussed in the last part of our walk. Vol. VI, page 101

WHAT IS BEHIND THE HORIZON?

> “If I arrived at the outermost edge of the heaven, could I extend my hand or staff into what is outside or not? It would be paradoxical not to be able to extend it.”
> Archytas of Tarentum (428–347 BCE)

> “The universe is a big place; perhaps the biggest.”
> Kilgore Trout, *Venus on the Half Shell.*

The horizon of the night sky is a tricky entity. In fact, all cosmological models show that it moves rapidly away from us. A detailed investigation shows that for a matter-dominated universe the horizon moves away from us with a velocity Ref. 249 Challenge 360 ny

$$v_{\text{horizon}} = 3c \,. \qquad (269)$$

A pretty result, isn't it? Obviously, since the horizon does not transport any signal, this is not a contradiction of relativity. Now, measurements of Ω_K show that space is essentially flat. Thus we can ask: What is behind the horizon? Page 243

If the universe is *open* or *marginal*, the matter we see at night is predicted by naively applied general relativity to be a – literally – infinitely small part of all matter existing. Indeed, applying the field equations to an open or marginal universe implies that there is an infinite amount of matter behind the horizon. Is such a statement testable? Challenge 361 s

In a *closed* universe, matter is still predicted to exist behind the horizon; however, in this case it is only a finite amount. Is this statement testable? Challenge 362 s

In short, the *concordance model of cosmology* states that there is a lot of matter behind the horizon. Like most cosmologists, we sweep the issue under the rug and take it up only later in our walk. A precise description of the topic is provided by the hypothesis of inflation.

---

* The Friedmann–Lemaître metric is also valid for any quotient of the just-mentioned simple topologies by a group of isometries, leading to dihedral spaces and lens spaces in the case $k = 1$, to tori in the case $k = 0$, and to *any* hyperbolic manifold in the case $k = -1$. Ref. 248

WHY ARE THERE STARS ALL OVER THE PLACE? – INFLATION

What were the initial conditions of matter? Matter was distributed in a constant density over space expanding with great speed. How could this happen? The researcher who has explored this question most thoroughly is Alan Guth. So far, we have based our studies of the night sky, cosmology, on two observational principles: the isotropy and the homogeneity of the universe. In addition, the universe is (almost) flat. The conjecture of *inflation* is an attempt to understand the origin of these observations.

Flatness at the present instant of time is strange: the flat state is an unstable solution of the Friedmann equations. Since the universe is still flat after fourteen thousand million years, it must have been even flatter near the big bang.

Ref. 250 Guth argued that the precise flatness, the homogeneity and the isotropy of the universe could follow if in the first second of its history, the universe had gone through a short phase of exponential size increase, which he called *inflation*. This exponential size increase, by a factor of about 10^{26}, would homogenize the universe. This extremely short evolution would be driven by a still-unknown field, the *inflaton field*. Inflation also seems to describe correctly the growth of inhomogeneities in the cosmic background radiation.

However, so far, inflation poses as many questions as it solves. Twenty years after his initial proposal, Guth himself is sceptical on whether it is a conceptual step forward. The final word on the issue has not been said yet.

WHY ARE THERE SO FEW STARS? – THE ENERGY AND ENTROPY CONTENT OF THE UNIVERSE

> “Die Energie der Welt ist constant. Die Entropie der Welt strebt einem Maximum zu.*”
> Rudolph Clausius

The matter–energy density of the universe is near the critical one. Inflation, described in the previous section, is the favourite explanation for this connection. This implies that the actual number of stars is given by the behaviour of matter at extremely high temperatures, and by the energy density left over at lower temperature. The precise connection is still the topic of intense research. But this issue also raises a question about the quotation above. Was the creator of the term ‘entropy’, Rudolph Clausius, right when he made this famous statement? Let us have a look at what general relativity has to say about all this.

In general relativity, a *total* energy can indeed be defined, in contrast to *localized* energy, which cannot. The total energy of all matter and radiation is indeed a constant of motion. It is given by the sum of the baryonic, luminous and neutrino parts:

$$E = E_\mathrm{b} + E_\gamma + E_\nu \approx \frac{c^2 M_0}{T_0} + \ldots \approx \frac{c^2}{G} + \ldots \,. \qquad (270)$$

This value is constant only when integrated over the whole universe, not when just the inside of the horizon is taken.**

---

* ‘The energy of the universe is constant. Its entropy tends towards a maximum.’

** Except for the case when pressure can be neglected.

Many people also add a gravitational energy term. If one tries to do so, one is obliged to define it in such a way that it is exactly the negative of the previous term. This value for the gravitational energy leads to the popular speculation that the *total* energy of the universe might be zero. In other words, the number of stars could also be limited by this relation.

However, the discussion of *entropy* puts a strong question mark behind all these seemingly obvious statements. Many people have tried to give values for the entropy of the universe. Some have checked whether the relation Ref. 251

$$S = \frac{kc^3}{G\hbar}\frac{A}{4} = \frac{kG}{\hbar c}4\pi M^2 \ , \qquad (271)$$

Challenge 363 ny which is correct for black holes, also applies to the universe. This assumes that all the matter and all the radiation of the universe can be described by some average temperature. They argue that the entropy of the universe is surprisingly low, so that there must be some ordering principle behind it. Others even speculate over where the entropy of the universe comes from, and whether the horizon is the source for it.

But let us be careful. Clausius assumes, without the slightest doubt, that the universe is a *closed system*, and thus deduces the statement quoted above. Let us check this assumption. Entropy describes the maximum energy that can be extracted from a hot object. After the discovery of the particle structure of matter, it became clear that entropy is also given by the number of microstates that can make up a specific macrostate. But neither definition makes any sense if applied to the universe as a whole. There is no way to extract energy from it, and no way to say how many microstates of the universe would look like the macrostate.

The basic reason is the impossibility of applying the concept of *state* to the universe. Vol. I, page 27 We first defined the state as all those properties of a system which allow one to distinguish it from other systems with the same intrinsic properties, or which differ from one observer to another. You might want to check for yourself that for the universe, such state Challenge 364 s properties do not exist at all.

We can speak of the state of space-time and we can speak of the state of matter and energy. But we cannot speak of the state of the universe, because the concept makes no sense.If there is no state of the universe, there is no entropy for it. And neither is there an energy value. This is in fact the only correct conclusion one can draw about the issue.

WHY IS MATTER LUMPED?

We are able to see the stars because the universe consists mainly of empty space, in other words, because stars are small and far apart. But why is this the case? Cosmic expansion was deduced and calculated using a homogeneous mass distribution. So why did matter lump together?

It turns out that homogeneous mass distributions are *unstable*. If for any reason the density fluctuates, regions of higher density will attract more matter than regions of lower density. Gravitation will thus cause the denser regions to increase in density and the regions of lower density to be depleted. Can you confirm the instability, simply by assuming Challenge 365 ny a space filled with dust and $a = GM/r^2$? In summary, even a tiny quantum fluctuation in the mass density will lead, after a certain time, to lumped matter.

But how did the first inhomogeneities form? That is one of the big problems of modern physics and astrophysics, and there is no accepted answer yet. Several modern experiments are measuring the variations of the cosmic background radiation spectrum with angular position and with polarization; these results, which will be available in the coming years, might provide some information on the way to settle the issue. Ref. 252

Why are stars so small compared with the universe?

Given that the matter density is around the critical one, the size of stars, which contain most of the matter, is a result of the interaction of the elementary particles composing them. Below we will show that general relativity (alone) cannot explain any size appearing in nature. The discussion of this issue is a theme of quantum theory. Page 282

Are stars and galaxies moving apart or is the universe expanding?

Can we distinguish between space expanding and galaxies moving apart? Yes, we can. Can you find an argument or devise an experiment to do so? Challenge 366 ny

The expansion of the universe does not apply to the space on the Earth. The expansion is calculated for a homogeneous and isotropic mass distribution. Matter is neither homogeneous nor isotropic inside the galaxy; the approximation of the cosmological principle is not valid down here. It has even been checked experimentally, by studying atomic spectra in various places in the solar system, that there is *no* Hubble expansion taking place around us. Ref. 253

Is there more than one universe?

The existence of 'several' universes might be an option when we study the question whether we see all the stars. But you can check that neither definition of 'universe' given above, be it 'all matter-energy' or 'all matter–energy and all space-time', allows us to speak of several universes. Challenge 367 e

There is *no* way to define a plural for universe: either the universe is everything, and then it is unique, or it is not everything, and then it is not the universe. We will discover that also quantum theory does not change this conclusion, despite recurring reports to the contrary. Vol. IV, page 166

Whoever speaks of many universes is talking gibberish.

Why are the stars fixed? – Arms, stars and Mach's principle

> “Si les astres étaient immobiles, le temps et l'espace n'existeraient plus.[*]”
>
> Maurice Maeterlink.

The two arms possessed by humans have played an important role in discussions about motion, and especially in the development of relativity. Looking at the stars at night, we

---

[*] 'If the stars were immobile, time and space would not exist any more.' Maurice Maeterlink (1862–1949) is a famous Belgian dramatist.

can make a simple observation, if we keep our arms relaxed. Standing still, our arms hang down. Then we turn rapidly. Our arms lift up. In fact they do so whenever we see the stars turning. Some people have spent a large part of their lives studying this phenomenon. Why?

Ref. 254 Stars and arms prove that motion is relative, not absolute.* This observation leads to two possible formulations of what Einstein called *Mach's principle*.

— *Inertial frames are determined by the rest of the matter in the universe.*

This idea is indeed realized in general relativity. No question about it.

— *Inertia is due to the* interaction *with the rest of the universe.*

This formulation is more controversial. Many interpret it as meaning that the *mass* of an object depends on the distribution of mass in the rest of the universe. That would mean that one needs to investigate whether mass is anisotropic when a large body is nearby. Of course, this question has been studied experimentally; one simply needs to measure whether a particle has the same mass values when accelerated in different directions. Ref. 255 Unsurprisingly, to a high degree of precision, no such anisotropy has been found. Many therefore conclude that Mach's principle is wrong. Others conclude with some pain in Ref. 256 their stomach that the whole topic is not yet settled.

But in fact it is easy to see that Mach *cannot* have meant a mass variation at all: one then would also have to conclude that mass is distance-dependent, even in Galilean physics. But this is known to be false; nobody in his right mind has ever had any doubts about Challenge 368 e it.

The whole debate is due to a misunderstanding of what is meant by 'inertia': one can interpret it as inertial *mass* or as inertial *motion* (like the moving arms under the stars). There is no evidence that Mach believed either in anisotropic mass or in distance-dependent mass; the whole discussion is an example people taking pride in not making a mistake which is incorrectly imputed to another, supposedly more stupid, person.**

Obviously, inertial effects do depend on the distribution of mass in the rest of the universe. Mach's principle is correct. Mach made some blunders in his life (he is infamous for opposing the idea of atoms until he died, against experimental evidence) but his principle is *not* one of them. Unfortunately it is to be expected that the myth about the Ref. 256 incorrectness of Mach's principle will persist, like that of the derision of Columbus.

In fact, Mach's principle is valuable. As an example, take our galaxy. Experiments show that it is flattened and rotating. The Sun turns around its centre in about 250 million years. Indeed, if the Sun did not turn around the galaxy's centre, we would fall into it in Page 211 about 20 million years. As mentioned above, from the shape of our galaxy we can draw the powerful conclusion that there must be a lot of other matter, i.e., a lot of other stars and galaxies in the universe.

---

* The original reasoning by Newton and many others used a bucket and the surface of the water in it; but the arguments are the same.

** A famous example is often learned at school. It is regularly suggested that Columbus was derided because he thought the Earth to be spherical. But he was not derided at all for this reason; there were only disagreements on the *size* of the Earth, and in fact it turned out that his critics were right, and that he was wrong in his own, much too small, estimate of the radius.

At rest in the universe

There is no preferred frame in special relativity, no absolute space. Is the same true in the actual universe? No; there *is* a preferred frame. Indeed, in the standard big-bang cosmology, the average galaxy is at rest. Even though we talk about the big bang, any average galaxy can rightly maintain that it is at rest. Each one is in free fall. An even better realization of this privileged frame of reference is provided by the background radiation.

In other words, the night sky is black because we move with almost no speed through background radiation. If the Earth had a large velocity relative to the background radiation, the sky would be bright even at night, thanks to the Doppler effect for the background radiation. In other words, the night sky is dark in all directions because of our slow motion against the background radiation.

This 'slow' motion has a speed of 368 km/s. (This is the value of the motion of the Sun; there are variations due to addition of the motion of the Earth.) The speed value is large in comparison to everyday life, but small compared to the speed of light. More detailed studies do not change this conclusion. Even the motion of the Milky Way and that of the local group against the cosmic background radiation is of the order of 600 km/s; that is still much slower than the speed of light. The reasons why the galaxy and the solar system move with these 'low' speeds across the universe have already been studied in our walk. Can you give a summary? Challenge 369 e

By the way, is the term 'universe' correct? Does the universe rotate, as its name implies? If by universe we mean the whole of experience, the question does not make sense, because rotation is only defined for bodies, i.e., for parts of the universe. However, if by universe we only mean 'all matter', the answer *can* be determined by experiments. It turns out that the rotation is extremely small, if there is any: measurements of the cosmic background radiation show that in the lifetime of the universe, its matter cannot have rotated by more than a hundredth of a millionth of a turn! In short, with a dose of humour we can say that 'universe' is a misnomer. Ref. 257

Does light attract light?

Another reason why we can see stars is that their light reaches us. But why are travelling light rays not disturbed by each other's gravitation? We know that light is energy and that any energy attracts other energy through gravitation. In particular, light is electromagnetic energy, and experiments have shown that all electromagnetic energy is subject to gravitation. Could two light beams that are advancing with a small angle between them converge, because of mutual gravitational attraction? That could have measurable and possibly interesting effects on the light observed from distant stars.

The simplest way to explore the issue is to study the following question: Do parallel light beams remain parallel? Interestingly, a precise calculation shows that mutual gravitation does *not* alter the path of two parallel light beams, even though it *does* alter the path of antiparallel light beams, i.e., parallel beams travelling in opposite directions. The reason is that for parallel beams moving at light speed, the gravitomagnetic component exactly *cancels* the gravitoelectric component. Ref. 258 Challenge 370 ny

Since light does not attract light moving along, light is not disturbed by its own gravity during the millions of years that it takes to reach us from distant stars. Light does not attract or disturb light moving alongside. So far, all known quantum-mechanical effects also confirm this conclusion.

DOES LIGHT DECAY?

In the section on quantum theory we will encounter experiments showing that light is made of particles. It is plausible that these photons might *decay* into some other particle, as yet unknown, or into lower-frequency photons. If that actually happened, we would not be able to see distant stars.

Challenge 371 e But any decay would also mean that light would change its direction (why?) and thus produce blurred images for remote objects. However, no blurring is observed. In addition, the Soviet physicist Matvey Bronshtein demonstrated in the 1930s that any light Ref. 259 decay process would have a larger rate for smaller frequencies. When people checked the shift of radio waves, in particular the famous 21 cm line, and compared it with the shift of light from the same source, no difference was found for any of the galaxies tested.

People even checked that Sommerfeld's fine-structure constant, which determines the Ref. 260 colour of objects, does not change over time. Despite an erroneous claim in recent years, no change could be detected over thousands of millions of years.

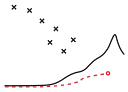
Of course, instead of decaying, light could also be *hit* by some hitherto unknown Challenge 372 ny entity. But this possibility is excluded by the same arguments. These investigations also show that there is no additional red-shift mechanism in nature apart from the Doppler Page 252 and gravitational red-shifts.

The visibility of the stars at night has indeed shed light on numerous properties of nature. We now continue our adventure with a more general issue, nearer to our quest for the fundamentals of motion.

SUMMARY ON COSMOLOGY

Asking what precisely we see at night leads to several awe-inspiring insights. First, the universe is huge – but of finite size. Secondly, the universe is extremely old – but of finite age. Thirdly, the universe is expanding.

If you ever have the chance to look through a big telescope, do so! It is wonderful.



Chapter 9

BLACK HOLES – FALLING FOREVER

> “Qui iacet in terra non habet unde cadat.[**]”
> Alanus de Insulis

Why explore black holes?

The most extreme gravitational phenomena in nature are black holes. They realize the limit of length-to-mass ratios in nature. In other words, they produce the highest force value possible in nature at their surface, the so-called *horizon*. Black holes also produce the highest space-time curvature values for a given mass value. In other terms, black holes are the most extreme general relativistic systems that are found in nature. Due to their extreme properties, the study of black holes is also a major stepping stone towards unification and the final description of motion.

Ref. 143 *Black hole* is shorthand for ‘gravitationally completely collapsed object’. Predicted over two centuries ago, it was unclear for a long time whether or not they exist. Around the year 2000, the available experimental data have now led most experts to conclude that there is a black hole at the centre of almost all galaxies, including our own (see Figure 116). Ref. 261 Black holes are also suspected at the heart of quasars, of *active galactic nuclei* and of *gamma-ray bursters*. In short, it seems that the evolution of galaxies is strongly tied to the evolution of black holes. In addition, about a dozen smaller black holes have been identified elsewhere in our galaxy. For these reasons, black holes, Ref. 262 the most impressive, the most powerful and the most relativistic systems in nature, are a fascinating subject of study.

Mass concentration and horizons

The *escape velocity* is the speed needed to launch an projectile in such a way that it never falls back down. The escape velocity depends on the mass and the size of the planet from which the launch takes place: the denser the planet is, the higher is the escape velocity. What happens when a planet or star has an escape velocity that is larger than the speed of light c? Such objects were first imagined by the British geologist John Michell in 1784, and Ref. 263 independently by the French mathematician Pierre Laplace in 1795, long before general relativity was developed. Michell and Laplace realized something fundamental: even if an object with such a high escape velocity were a hot star, to a distant observer it would appear to be completely black, as illustrated in Figure 117. The object would not allow

---

[**] ‘He who lies on the ground cannot fall down from it.’ The author’s original name is Alain de Lille (*c.* 1128 –1203).

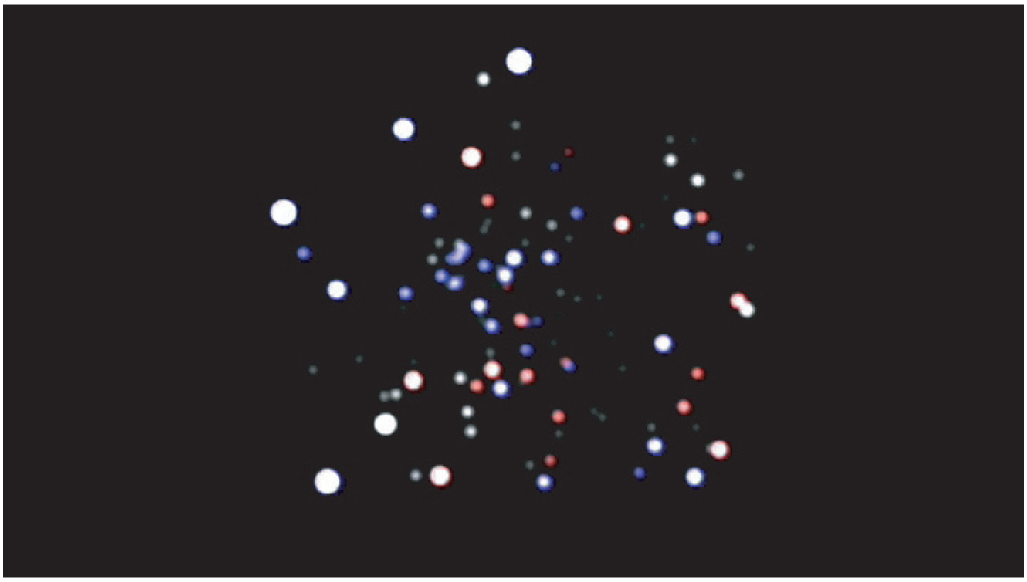


FIGURE 116 A time-lapse film, taken over a period of 16 years, of the orbits of the stars near the centre of our Galaxy. The invisible central object is so massive and small that it is almost surely a black hole (QuickTime film © ESO).

any light to leave it; in addition, it would block all light coming from behind it. In 1967, John Wheeler[*] made the now standard term *black hole*, due to Anne Ewing, popular in physics. Ref. 143

Challenge 373 e It only takes a few lines to show that light cannot escape from a body of mass M whenever the radius is smaller than a critical value given by

$$R_S = \frac{2GM}{c^2} \tag{272}$$

called the *Schwarzschild radius*. The formula is valid both in universal gravity and in general relativity, provided that in general relativity we take the radius as meaning the circumference divided by 2π. Such a body realizes the limit value for length-to-mass ratios in nature. For this and other reasons to be given shortly, we will call R_S also the *size* of the black hole of mass M. (But note that it is only half the diameter.) In principle, it is possible to imagine an object with a smaller length-to-mass ratio; however, we will discover that there is no way to observe an object smaller than the Schwarzschild radius, just as an object moving faster than the speed of light cannot be observed. However, we can observe black holes – the limit case – just as we can observe entities moving at the speed of light.

When a test mass is made to shrink and to approach the critical radius R_S, two things happen. First, the local proper acceleration for (imaginary) point masses increases

---

[*] John Archibald Wheeler (1911–2008), US-American physicist, important expert on general relativity and author of several excellent textbooks, among them the beautiful JOHN A. WHEELER, *A Journey into Gravity and Spacetime*, Scientific American Library & Freeman, 1990, in which he explains general relativity with passion and in detail, but without any mathematics.

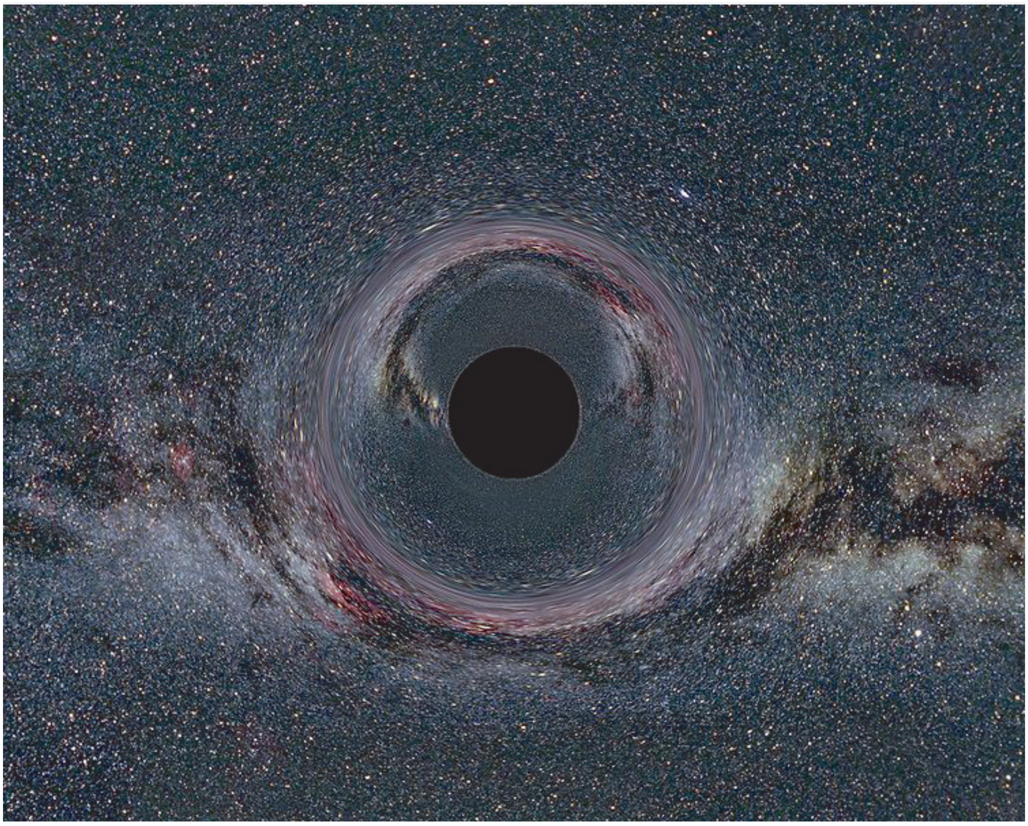


FIGURE 117 A simplified simulated image of how a black hole of ten solar masses, with Schwarzschild radius of 30 km, seen from a constant distance of 600 km, will distort an image of the Milky Way in the background. Note the Zwicky–Einstein ring formed at around twice the black hole radius and the thin bright rim (image © Ute Kraus at www.tempolimit-lichtgeschwindigkeit.de).

without bound. For realistic objects of finite size, the black hole realizes the highest force possible in nature. Something that falls into a black hole cannot be pulled back out. A black hole thus swallows all matter that falls into it. It acts like a cosmic vacuum cleaner.

At the surface of a black hole, the red-shift factor for a distant observer also increases without bound. The ratio between the two quantities is called the *surface gravity* of a black hole. It is given by

Challenge 374 ny

$$g_{\text{surf}} = \frac{GM}{R_S^2} = \frac{c^4}{4GM} = \frac{c^2}{2R_S} . \tag{273}$$

A black hole thus does not allow any light to leave it.

A surface that realizes the force limit and an infinite red-shift makes it is impossible to send light, matter, energy or signals of any kind to the outside world. A black hole is thus surrounded by a horizon. We know that a horizon is a limit surface. In fact, a horizon is a limit in two ways. First, a horizon is a limit to communication: nothing can communicate across it. Secondly, a horizon is a surface of maximum force and power. These properties are sufficient to answer all questions about the effects of horizons. For

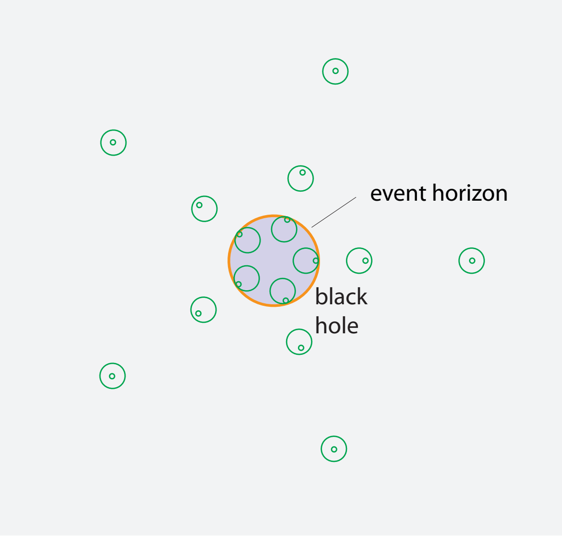


FIGURE 118 The light cones in the equatorial plane around a non-rotating black hole, seen from above the plane.

Challenge 375 s example: What happens when a light beam is sent upwards from the horizon? And from slightly above the horizon? Figure 118 provides some hints.

Black holes, regarded as astronomical objects, are thus different from planets. During the formation of planets, matter lumps together; as soon as it cannot be compressed any further, an equilibrium is reached, which determines the radius of the planet. That is the same mechanism as when a stone is thrown towards the Earth: it stops falling when it *hits* the ground. A 'ground' is formed whenever matter hits other matter. In the case of a black hole, there is no ground; everything *continues* falling. That is why, in Russian, black holes used to be called *collapsars*.

This continuous falling of a black hole takes place when the concentration of matter is so high that it overcomes all those interactions which make matter *impenetrable* in daily life. In 1939, Robert Oppenheimer* and Hartland Snyder showed theoretically that a black hole forms whenever a star of sufficient mass stops burning. When a star of sufficient mass stops burning, the interactions that form the 'floor' disappear, and everything continues falling without end. Ref. 264

A *black hole is matter in permanent free fall.* Nevertheless, its radius for an outside observer remains constant! But that is not all. Furthermore, because of this permanent free fall, black holes are the only state of matter in thermodynamic equilibrium! In a sense, floors and all other every-day states of matter are metastable: these forms are not as stable as black holes.

---

* Robert Oppenheimer (1904–1967), important US-American physicist. He can be called the father of theoretical physics in the USA. He worked on quantum theory and atomic physics. He then headed the team that developed the nuclear bomb during the Second World War. He was also the most prominent (innocent) victim of one of the greatest witch-hunts ever organized in his home country. See also the www.nap.edu/readingroom/books/biomems/joppenheimer.html website.

Black hole horizons as limit surfaces

Page 96

The characterizing property of a black hole is thus its *horizon*. The first time we encountered horizons was in special relativity, in the section on accelerated observers. The horizons due to gravitation are similar in all their properties; the section on the maximum force and power gave a first impression. The only difference we have found is due to the neglect of gravitation in special relativity. As a result, horizons in nature cannot be planar, in contrast to what is suggested by the observations of the imagined point-like observers assumed to exist in special relativity.

Page 146

Both the maximum force principle and the field equations imply that the space-time around a rotationally symmetric (thus non-rotating) and electrically neutral mass is described by

$$\mathrm{d}i^2 = \left(1 - \frac{2GM}{rc^2}\right)\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{1 - \frac{2GM}{rc^2}} - r^2\mathrm{d}\varphi^2/c^2 \; . \tag{274}$$

This is the so-called *Schwarzschild metric*. As mentioned above, r is the circumference divided by 2π; t is the time measured at infinity.

Let us now assume that the mass is strongly localized. We then find that no *outside* observer will ever receive any signal emitted from a radius value $r = 2GM/c^2$ or smaller. We have a horizon at that distance, and the situation describes a black hole. Indeed, as the proper time i of an observer at radius r is related to the time t of an observer at infinity through

$$\mathrm{d}i = \sqrt{1 - \frac{2GM}{rc^2}}\,\mathrm{d}t \; , \tag{275}$$

we find that an observer at the horizon would have vanishing proper time. In other words, at the horizon the red-shift is infinite. (More precisely, the surface of infinite red-shift and the horizon coincide only for non-rotating black holes. For rotating black holes, the two surfaces are distinct.) Everything happening at the horizon goes on infinitely slowly, as observed by a distant observer. In other words, for a distant observer observing what is going on at the horizon itself, nothing at all ever happens.

In the same way that observers cannot reach the speed of light, observers cannot reach a horizon. For a second observer, it can only happen that the first is moving almost as fast as light; in the same way, for a second observer, it can only happen that the first has almost reached the horizon. In addition, a traveller cannot feel how much he is near the speed of light for another, and experiences light speed as unattainable; in the same way, a traveller (into a large black hole) cannot feel how much he is near a horizon and experiences the horizon as unattainable.

We cannot say what happens inside the horizon.* We can take this view to the extreme and argue that the black hole metric is a type of vacuum metric. In this view, mass is a quantity that is 'built' from vacuum.

---

* Of course, mathematicians do not care about physical arguments. Therefore, Martin Kruskal and George Szekeres have defined coordinates for the inside of the black hole. However, these and similar coordinate systems are unrealistic academic curiosities, as they contradict quantum theory. Coordinate systems for the inside of a black hole horizon have the same status as coordinate systems behind the cosmological horizon: they are belief systems that are not experimentally verifiable.

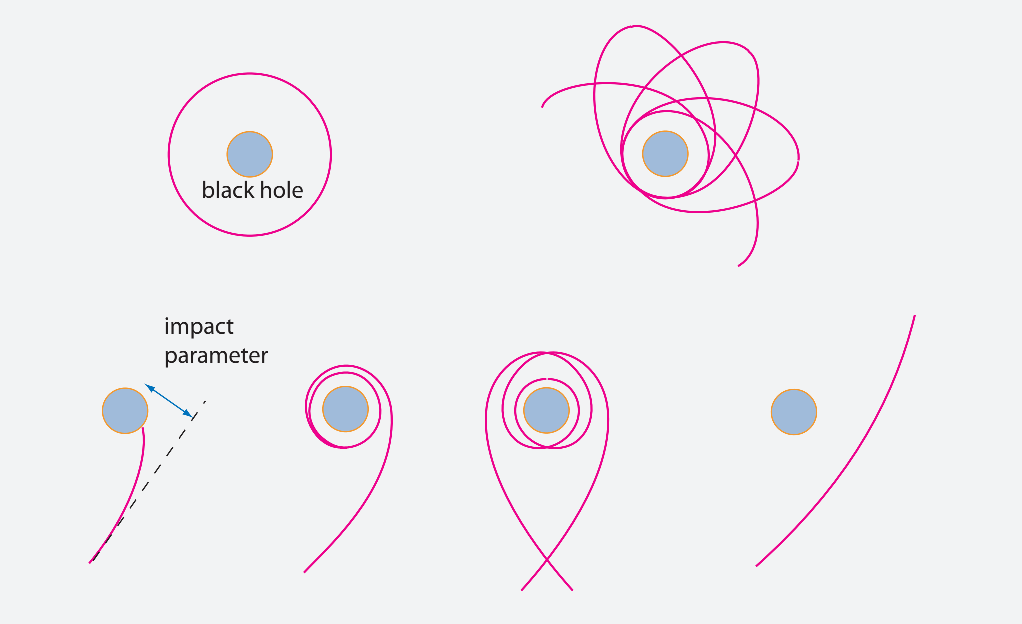


FIGURE 119 Motions of *massive* objects around a non-rotating black hole – for different impact parameters and initial velocities.

In general relativity, horizons of any kind are predicted to be *black*. Since light cannot escape from them, classical horizons are completely dark surfaces. In fact, horizons are the darkest entities imaginable: nothing in nature is darker. Nonetheless, we will discover below that physical horizons are not completely black. Page 272

Orbits around black holes

Ref. 259 Since black holes curve space-time strongly, a body moving near a black hole behaves in more complicated ways than predicted by universal gravity. In universal gravity, paths are either ellipses, parabolas, or hyperbolas; all these are plane curves. It turns out that paths lie in a plane only near *non-rotating* black holes.*

Around non-rotating black holes, also called *Schwarzschild black holes*, circular paths are impossible for radii less than $3R_S/2$ (can you show why?) and are unstable to perturbations from there up to a radius of $3R_S$. Only at larger radii are circular orbits stable. Around black holes, there are no elliptic paths; the corresponding rosetta path is shown in Figure 119. Such a path shows the famous periastron shift in all its glory. Challenge 377 ny

Note that the potential around a black hole is not appreciably different from $1/r$ for distances above about fifteen Schwarzschild radii. For a black hole of the mass of the Challenge 378 e

---

* For such paths, Kepler's rule connecting the average distance and the time of orbit

$$\frac{GMt^3}{(2\pi)^2} = r^3 \tag{276}$$

still holds, provided the proper time and the radius measured by a distant observer are used. Challenge 376 ny

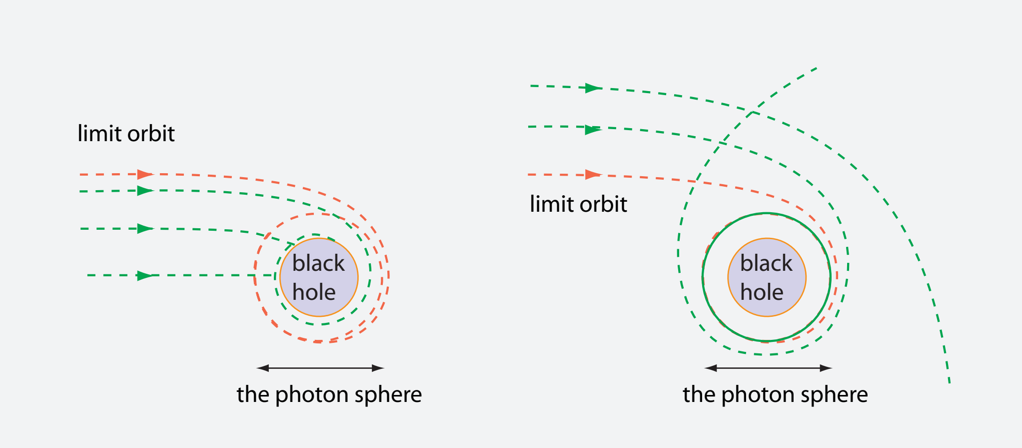


FIGURE 120 Motions of *light* passing near a non-rotating black hole.

Sun, that would be 42 km from its centre; therefore, we would not be able to note any difference for the path of the Earth around the Sun.

We have mentioned several times in our adventure that gravitation is characterized by its tidal effects. Black holes show extreme properties in this respect. If a cloud of dust falls into a black hole, the size of the cloud increases as it falls, until the cloud envelops the whole horizon. In fact, the result is valid for any extended body. This property of black holes will be of importance later on, when we will discuss the size of elementary particles.

For falling bodies coming from infinity, the situation near black holes is even more interesting. Of course there are no hyperbolic paths, only trajectories similar to hyperbolas for bodies passing far enough away. But for small, but not too small impact parameters, a body will make a number of turns around the black hole, before leaving again. The number of turns increases beyond all bounds with decreasing impact parameter, until a value is reached at which the body is captured into an orbit at a radius of $2R$, as shown in Figure 119. In other words, this orbit *captures* incoming bodies if they approach it below a certain critical angle. For comparison, remember that in universal gravity, capture is never possible. At still smaller impact parameters, the black hole swallows the incoming mass. In both cases, capture and deflection, a body can make several turns around the black hole, whereas in universal gravity it is impossible to make more than *half* a turn around a body.

Challenge 379 ny

The most absurd-looking orbits, though, are those corresponding to the parabolic case of universal gravity. (These are of purely academic interest, as they occur with probability zero.) In summary, relativity changes the motions due to gravity quite drastically.

Around *rotating* black holes, the orbits of point masses are even more complex than those shown in Figure 119; for bound motion, for example, the ellipses do not stay in one plane – thanks to the Thirring–Lense effect – leading to extremely involved orbits in three dimensions filling the space around the black hole.

For *light* passing a black hole, the paths are equally interesting, as shown in Figure 120. There are no qualitative differences with the case of rapid particles. For a non-rotating black hole, the path obviously lies in a single plane. Of course, if light passes sufficiently nearby, it can be strongly bent, as well as captured. Again, light can also make one or several turns around the black hole before leaving or being captured. The limit between the two cases is the path in which light moves in a circle around a black hole, at $3R/2$. If we were located on that orbit, we would see the back of our head by looking forward! Challenge 380 ny However, this orbit is unstable. The surface containing all orbits inside the circular one is called the *photon sphere*. The photon sphere thus divides paths leading to capture from those leading to infinity. Note that there is no *stable* orbit for light around a black hole. Challenge 381 ny Are there any rosetta paths for light around a black hole?

For light around a *rotating* black hole, paths are much more complex. Already in the equatorial plane there are two possible circular light paths: a smaller one in the direction of the rotation, and a larger one in the opposite direction. Challenge 382 ny

For *charged* black holes, the orbits for falling charged particles are even more complex. The electrical field lines need to be taken into account. Several fascinating effects appear which have no correspondence in usual electromagnetism, such as effects similar to electrical versions of the Meissner effect. The behaviour of such orbits is still an active area of research in general relativity.

BLACK HOLES HAVE NO HAIR

How is a black hole characterized? It turns out that all properties of black holes follow from a few basic quantities characterizing them, namely their mass M, their angular momentum J, and their electric charge Q.[*] All other properties – such as size, shape, colour, magnetic field – are uniquely determined by these.[**] It is as though, to use Wheeler's colourful analogy, one could deduce every characteristic of a woman from her size, her waist and her height. Physicists also say that black holes 'have no hair,' meaning that (classical) black holes have no other degrees of freedom. This expression was also introduced by Wheeler.[***] Ref. 267 This fact was proved by Israel, Carter, Robinson and Mazur; they showed that for a given mass, angular momentum and charge, there is only *one* possible black hole. Ref. 268 (However, the uniqueness theorem is not valid any more if the black hole carries nuclear quantum numbers, such as weak or strong charges.)

In other words, a black hole is independent of how it has formed, and of the materials used when forming it. Black holes all have the same composition, or better, they have no composition at all.

The mass M of a black hole is not restricted by general relativity. It may be as small as that of a microscopic particle and as large as many million solar masses. But for their angular momentum J and electric charge Q, the situation is different. A rotating black

---

[*] The existence of three basic characteristics is reminiscent of particles. We will find out more about the relation between black holes and particles in the final part of our adventure. Vol. VI, page 148

[**] Mainly for marketing reasons, non-rotating and electrically neutral black holes are often called *Schwarzschild* black holes; uncharged and rotating ones are often called *Kerr* black holes, after Roy Kerr, who discovered the corresponding solution of Einstein's field equations in 1963. Ref. 265 Electrically charged but non-rotating black holes are often called *Reissner–Nordström black holes*, after the German physicist Hans Reissner and the Finnish physicist Gunnar Nordström. The general case, charged and rotating, is sometimes named after Kerr and Newman. Ref. 266

[***] Ref. 143 Wheeler claims that he was inspired by the difficulty of distinguishing between bald men; however, it is not a secret that Feynman, Ruffini and others had a clear anatomical image in mind when they stated that 'black holes, in contrast to their surroundings, have no hair.'

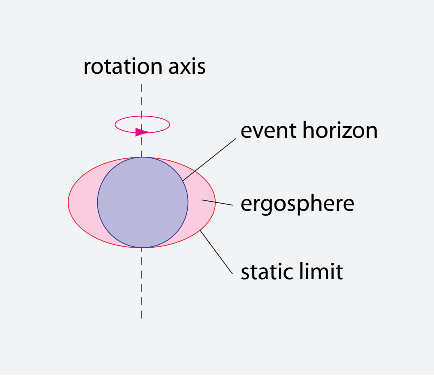


FIGURE 121 The ergosphere of a rotating black hole.

hole has a maximum possible angular momentum and a maximum possible electric (and magnetic) charge.* The limit on the angular momentum appears because its perimeter may not move faster than light. The electric charge is also limited. The two limits are not independent: they are related by Challenge 383 ny

$$\left(\frac{J}{cM}\right)^2 + \frac{GQ^2}{4\pi\varepsilon_0 c^4} \leqslant \left(\frac{GM}{c^2}\right)^2 . \tag{277}$$

This follows from the limit on length-to-mass ratios at the basis of general relativity. Rotating black holes realizing the limit (277) are called *extremal* black holes. The limit (277) implies that the horizon radius of a general black hole is given by Challenge 384 ny

$$r_\mathrm{h} = \frac{GM}{c^2}\left(1 + \sqrt{1 - \frac{J^2c^2}{M^4G^2} - \frac{Q^2}{4\pi\varepsilon_0 GM^2}}\right) \tag{278}$$

For example, for a black hole with the mass and half the angular momentum of the Sun, namely $2 \cdot 10^{30}$ kg and $0.45 \cdot 10^{42}$ kg m^2/s, the charge limit is about $1.4 \cdot 10^{20}$ C.

How does one distinguish rotating from non-rotating black holes? First of all by the *shape*. Non-rotating black holes must be spherical (any non-sphericity is radiated away as gravitational waves) and rotating black holes have a slightly flattened shape, uniquely determined by their angular momentum. Ref. 269 Because of their rotation, their surface of infinite gravity or infinite red-shift, called the *static limit*, is different from their (outer) horizon, as illustrated in Figure 121. The region in between is called the *ergosphere*; this is a misnomer as it is *not* a sphere. (It is so called because, as we will see shortly, it can be used to extract energy from the black hole.) The motion of bodies within the ergosphere can be quite complex. It suffices to mention that rotating black holes drag any in-falling body into an orbit around them; this is in contrast to non-rotating black holes, which swallow in-falling bodies. In other words, rotating black holes are not really 'holes' at all, but rather vortices.

---

[*] More about the conjectured magnetic charge later on. In black holes, it enters like an additional type of charge into all expressions in which electric charge appears. Vol. III, page 55

The distinction between rotating and non-rotating black holes also appears in the horizon surface area. The (horizon) surface area A of a non-rotating and uncharged black hole is obviously related to its mass M by Challenge 385 e

$$A = \frac{16\pi G^2}{c^4} M^2 \,. \tag{279}$$

The relation between surface area and mass for a rotating and charged black hole is more complex: it is given by

$$A = \frac{8\pi G^2}{c^4} M^2 \left(1 + \sqrt{1 - \frac{J^2 c^2}{M^4 G^2} - \frac{Q^2}{4\pi\varepsilon_0 G M^2}} \right) \tag{280}$$

where J is the angular momentum and Q the charge. In fact, the relation

$$A = \frac{8\pi G}{c^2} M r_{\mathrm{h}} \tag{281}$$

is valid for *all* black holes. Obviously, in the case of an electrically charged black hole, the rotation also produces a magnetic field around it. This is in contrast with non-rotating black holes, which cannot have a magnetic field.

BLACK HOLES AS ENERGY SOURCES

Can one extract energy from a black hole? Roger Penrose has discovered that this is possible for *rotating* black holes. Ref. 270 A rocket orbiting a rotating black hole in its ergosphere could switch its engines on and would then get hurled into outer space at tremendous velocity, much greater than what the engines could have produced by themselves. In fact, the same effect is used by rockets on the Earth, and is the reason why all satellites orbit the Earth in the same direction; it would require much more fuel to make them turn the other way.*

The energy gained by the rocket would be lost by the black hole, which would thus slow down and lose some mass; on the other hand, there is a mass increases due to the exhaust gases falling into the black hole. This increase always is larger than, or at best equal to, the loss due to rotation slowdown. The best one can do is to turn the engines on exactly at the horizon; then the horizon area of the black hole stays constant, and only its rotation is slowed down.**

As a result, for a neutral black hole *rotating* with its maximum possible angular momentum, $1 - 1/\sqrt{2} = 29.3\,\%$ of its total energy can be extracted through the Penrose process. For black holes rotating more slowly, the percentage is obviously smaller. Challenge 387 ny

---

[*] And it would be much more dangerous, since any small object would hit such an against-the-stream satellite at about 15.8 km/s, thus transforming the object into a dangerous projectile. Challenge 386 ny In fact, any power wanting to destroy satellites of the enemy would simply have to load a satellite with nuts or bolts, send it into space the wrong way, and distribute the bolts into a cloud. It would make satellites impossible for many decades to come.

[**] It is also possible to extract energy from rotational black holes through gravitational radiation.

For *charged* black holes, such irreversible energy extraction processes are also possible. Can you think of a way? Using expression (277), we find that up to 50 % of the mass of a non-rotating black hole can be due to its charge. In fact, in the quantum part of our adventure we will encounter an energy extraction process which nature seems to use quite frequently.

Challenge 388 ny
Challenge 389 ny
Vol. V, page 153

The Penrose process allows one to determine how angular momentum and charge increase the mass of a black hole. The result is the famous mass–energy relation

Ref. 271

$$M^2 = \frac{E^2}{c^4} = \left(m_{\text{irr}} + \frac{Q^2}{16\pi\varepsilon_0 G m_{\text{irr}}}\right)^2 + \frac{J^2}{4m_{\text{irr}}^2}\frac{c^2}{G^2} = \left(m_{\text{irr}} + \frac{Q^2}{8\pi\varepsilon_0 \rho_{\text{irr}}}\right)^2 + \frac{J^2}{\rho_{\text{irr}}^2}\frac{1}{c^2} \qquad (282)$$

which shows how the electrostatic and the rotational energy enter the mass of a black hole. In the expression, m_{irr} is the *irreducible mass* defined as

$$m_{\text{irr}}^2 = \frac{A(M, Q=0, J=0)}{16\pi}\frac{c^4}{G^2} = \left(\rho_{\text{irr}}\frac{c^2}{2G}\right)^2 \qquad (283)$$

and ρ_{irr} is the *irreducible radius*.

Detailed investigations show that there is no process which *decreases* the horizon area, and thus the irreducible mass or radius, of the black hole. People have checked this in all ways possible and imaginable. For example, when two black holes merge, the total area increases. One calls processes which keep the area and energy of the black hole constant *reversible*, and all others *irreversible*. In fact, the area of black holes behaves like the *entropy* of a closed system: it never decreases. That the area in fact *is* an entropy was first stated in 1970 by Jacob Bekenstein. He deduced that only when an entropy is ascribed to a black hole, is it possible to understand where the entropy of all the material falling into it is collected.

Ref. 272

The black hole entropy is a function only of the mass, the angular momentum and the charge of the black hole. You might want to confirm Bekenstein's deduction that the entropy S is proportional to the horizon area. Later it was found, using quantum theory, that

Challenge 390 ny

$$S = \frac{A}{4}\frac{kc^3}{\hbar G} = \frac{A\,k}{4\,l_{\text{Pl}}^2}\,. \qquad (284)$$

This famous relation cannot be deduced without quantum theory, as the absolute value of entropy, as for any other observable, is never fixed by classical physics alone. We will discuss this expression later on in our adventure.

Vol. V, page 154

If black holes have an entropy, they also must have a temperature. If they have a temperature, they must shine. Black holes thus cannot be black! This was proven by Stephen Hawking in 1974 with extremely involved calculations. However, it could have been deduced in the 1930s, with a simple Gedanken experiment which we will present later on. You might want to think about the issue, asking and investigating what strange consequences would appear if black holes had no entropy. Black hole radiation is a further, though tiny (quantum) mechanism for energy extraction, and is applicable even to non-rotating, uncharged black holes. The interesting connections between black holes, ther-

Vol. V, page 147

TABLE 8 Types of black holes.

Black hole type	Mass	Charge	Angular momentum	Experimental evidence
Supermassive black holes	10^5 to 10^{11} $m_\odot$	unknown	unknown	orbits of nearby stars, light emission from accretion
Intermediate black holes	50 to 10^5 $m_\odot$	unknown	unknown	X-ray emission of accreting matter
Stellar black holes	1 to 50 $m_\odot$	unknown	unknown	X-ray emission from double star companion
Primordial black holes	below 1 $m_\odot$	unknown	unknown	undetected so far; research ongoing
Micro black holes	below 1 g	n.a.	n.a.	none; appear only in science fiction and in the mind of cranks

Vol. V, page 147 Challenge 391 ny modynamics, and quantum theory will be presented in the upcoming parts of our adventure. Can you imagine other mechanisms that make black holes shine?

FORMATION OF AND SEARCH FOR BLACK HOLES

How might black holes form? At present, at least three possible mechanisms have been distinguished; the question is still a hot subject of research. First of all, black holes could have formed during the early stages of the universe. These *primordial black holes* might grow through *accretion*, i.e., through the swallowing of nearby matter and radiation, or disappear through one of the mechanisms to be studied later on. Ref. 273 Vol. V, page 150

Of the *observed* black holes, the so-called *supermassive* black holes are found at the centre of every galaxy studied so far. They have typical masses in the range from 10^6 to 10^9 solar masses and contain about 0.5 % of the mass of a galaxy. For example, the black hole at the centre of the Milky Way has about 2.6 million solar masses, while the central black hole of the galaxy M87 has 6400 million solar masses. Supermassive black holes seem to exist at the centre of almost all galaxies, and seem to be related to the formation of galaxies themselves. Supermassive black holes are supposed to have formed through the collapse of large dust clouds, and to have grown through subsequent accretion of matter. The latest ideas imply that these black holes accrete a lot of matter in their early stage; the matter falling in emits lots of radiation, which would explain the brightness of quasars. Later on, the rate of accretion slows, and the less spectacular Seyfert galaxies form. It may even be that the supermassive black hole at the centre of the galaxy triggers the formation of stars. Still later, these supermassive black holes become almost dormant, or quiescent, like the one at the centre of the Milky Way. Ref. 261

Ref. 274 On the other hand, black holes can form when old massive stars *collapse*. It is estimated that when stars with at least three solar masses burn out their fuel, part of the matter remaining will collapse into a black hole. Such *stellar* black holes have a mass between one and a hundred solar masses; they can also continue growing through subsequent accretion. This situation provided the first ever candidate for a black hole, Cygnus X-1, Ref. 261 which was discovered in 1971. Over a dozen stellar black holes of between 4 and 20 solar masses are known to be scattered around our own galaxy; all have been discovered after 1971.

Recent measurements suggest also the existence of *intermediate* black holes, with typical masses around a thousand solar masses; the mechanisms and conditions for their formation are still unknown. The first candidates were found in the year 2000. Astronomers are also studying how large numbers of black holes in star clusters behave, and how often they collide. Under certain circumstances, the two black holes merge. Whatever the outcome, black hole collisions emit strong gravitational waves. In fact, this Page 181 signal is being looked for at the gravitational wave detectors that are in operation around the globe.

The search for black holes is a popular sport among astrophysicists. Conceptually, the simplest way to search for them is to look for strong gravitational fields. But only double stars allow one to measure gravitational fields directly, and the strongest ever measured Ref. 275 is 30 % of the theoretical maximum value. Another obvious way is to look for strong gravitational lenses, and try to get a mass-to-size ratio pointing to a black hole; however, no black holes was found in this way yet. Still another method is to look at the dynamics of stars near the centre of galaxies. Measuring their motion, one can deduce the mass of the body they orbit. The most favoured method to search for black holes is to look for extremely intense X-ray emission from point sources, using space-based satellites or balloon-based detectors. If the distance to the object is known, its absolute brightness can be deduced; if it is above a certain limit, it must be a black hole, since normal matter cannot produce an unlimited amount of light. This method is being perfected with the aim of directly observing of energy disappearing into a horizon. This disappearance may Ref. 276 in fact have been observed recently.

Finally, there is the suspicion that small black holes might be found in the halos of galaxies, and make up a substantial fraction of the so-called dark matter.

In summary, the list of discoveries about black holes is expected to expand dramatically in the coming years.

Singularities

Solving the equations of general relativity for various initial conditions, one finds that a cloud of dust usually collapses to a *singularity*, i.e., to a point of infinite density. The same conclusion appears when one follows the evolution of the universe backwards in time. In fact, Roger Penrose and Stephen Hawking have proved several mathematical theorems on the necessity of singularities for many classical matter distributions. These theorems assume only the continuity of space-time and a few rather weak conditions on Ref. 277 the matter in it. The theorems state that in expanding systems such as the universe itself, or in collapsing systems such as black holes in formation, events with infinite matter density should exist somewhere in the past, or in the future, respectively. This result is usually summarized by saying that there is a mathematical proof that the universe started in a singularity.

In fact, the derivation of the initial singularities makes a hidden, but strong assumption about matter: that dust particles have no proper size, i.e., that they are point-like. In other words, it is assumed that dust particles are singularities. Only with this assumption can one deduce the existence of initial or final singularities. However, we have seen that the maximum force principle can be reformulated as a minimum size principle for matter. The argument that there must have been an initial singularity of the universe is thus flawed! The experimental situation is clear: there is overwhelming evidence for an early state of the universe that was extremely hot and dense; but there is *no* evidence for *infinite* temperature or density.

Mathematically inclined researchers distinguish two types of singularities: those with a horizon – also called *dressed* singularities – and those without a horizon, the so-called *naked* singularities. Naked singularities are especially strange: for example, a toothbrush could fall into a naked singularity and disappear without leaving any trace. Since the field equations are time invariant, we could thus expect that every now and then, naked singularities emit toothbrushes. (Can you explain why dressed singularities are less dangerous?) Challenge 392 ny

To avoid the spontaneous appearance of toothbrushes, over the years many people have tried to discover some theoretical principles forbidding the existence of naked singularities. It turns out that there are two such principles. The first is the maximum force or maximum power principle we encountered above. The maximum force implies that no infinite force values appear in nature; in other words, there are no naked singularities in nature. This statement is often called *cosmic censorship*. Obviously, if general relativity were not the correct description of nature, naked singularities *could* still appear. Cosmic censorship is thus still discussed in research articles. The experimental search for naked singularities has not yielded any success; in fact, there is not even a candidate observation for the – less abstruse – *dressed* singularities. But the theoretical case for 'dressed' singularities is also weak. Since there is no way to interact with anything behind a horizon, it is futile to discuss what happens there. There is no way to prove that behind a horizon a singularity exists. Dressed singularities are articles of faith, not of physics. Ref. 278

In fact, there is another principle preventing singularities, namely *quantum theory*. Whenever we encounter a prediction of an infinite value, we have extended our description of nature to a domain for which it was not conceived. To speak about singularities, one must assume the applicability of pure general relativity to very small distances and very high energies. As will become clear in the last volume, nature does not allow this: the combination of general relativity and quantum theory shows that it makes no sense to talk about 'singularities', nor about what happens 'inside' a black hole horizon. The reason is that arbitrary small time and space values do not exist in nature. Vol. VI, page 102 Vol. VI, page 65

Curiosities and fun challenges about black holes

> “Tiens, les trous noirs. C’est troublant.*”
> Anonymous

Black holes have many counter-intuitive properties. We will first have a look at the classical effects, leaving the quantum effects for later on. Vol. V, page 156

* *

Following universal gravity, light could climb upwards from the surface of a black hole

---

* No translation possible.

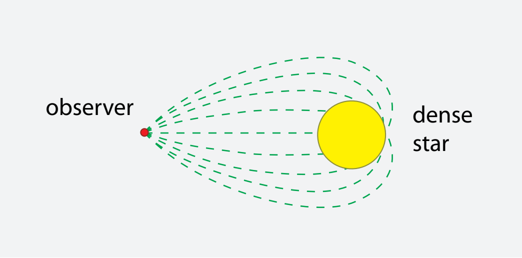


FIGURE 122 Motion of some light rays from a dense body to an observer.

and then fall back down. In general relativity, a black hole does not allow light to climb up at all; it can only fall. Can you confirm this? Challenge 393 ny

* *

What happens to a person falling into a black hole? An outside observer gives a clear answer: the falling person *never* arrives there since she needs an infinite time to reach the horizon. Can you confirm this result? Challenge 394 ny The falling person, however, reaches the horizon in a *finite* amount of her own time. Can you calculate it? Challenge 395 ny

This result is surprising, as it means that for an outside observer in a universe with *finite* age, black holes cannot have formed yet! At best, we can only observe systems that are busy forming black holes. In a sense, it might be correct to say that black holes do not exist. Black holes could have existed right from the start in the fabric of space-time. On the other hand, we will find out later why this is impossible. In short, it is important to keep in mind that the idea of black hole is a limit concept but that usually, limit concepts (like baths or temperature) are useful descriptions of nature. Independently of this last issue, we can confirm that in nature, the length-to-mass ratio always satisfies

$$\frac{L}{M} \geqslant \frac{4G}{c^2} \; . \qquad (285)$$

No exception has ever been observed.

* *

Interestingly, the *size* of a person falling into a black hole is experienced in vastly different ways by the falling person and a person staying outside. If the black hole is large, the in-falling observer feels almost nothing, as the tidal effects are small. The outside observer makes a startling observation: he sees the falling person *spread* all over the horizon of the black hole. *In-falling, extended bodies cover the whole horizon.* Can you explain this fact, for example by using the limit on length-to-mass ratios? Challenge 396 ny

This strange result will be of importance later on in our exploration, and lead to important results about the size of point particles.

* *

An observer near a (non-rotating) black hole, or in fact near any object smaller than 7/4 times its gravitational radius, can even see the complete *back* side of the object, as shown in Figure 122. Can you imagine what the image looks like? Challenge 397 ny Note that in addition to the paths shown in Figure 122, light can also turn several times around the black hole before reaching the observer! Therefore, such an observer sees an infinite number of images of the black hole. The resulting formula for the angular size of the innermost image was given above. Page 157

In fact, the effect of gravity means that it is possible to observe more than half the surface of *any* spherical object. In everyday life, however, the effect is small: for example, light bending allows us to see about 50.0002 % of the surface of the Sun.

* *

A mass point inside the smallest circular path of light around a black hole, at $3R/2$, cannot stay in a circle, because in that region, something strange happens. A body which circles another in everyday life always feels a tendency to be pushed outwards; this centrifugal effect is due to the inertia of the body. But at values below $3R/2$, a circulating body is pushed *inwards* by its inertia. There are several ways to explain this paradoxical effect. Ref. 279 The simplest is to note that near a black hole, the weight increases faster than the centrifugal force, as you may want to check yourself. Challenge 398 ny Only a rocket with engines switched on and pushing towards the sky can orbit a black hole at $3R/2$.

* *

By the way, how can gravity, or an electrical field, come out of a black hole, if no signal and no energy can leave it? Challenge 399 s

* *

Do *white holes* exist, i.e., time-inverted black holes, in which everything flows out of, instead of into, some bounded region? Challenge 400 ny

* *

Challenge 401 ny Show that a cosmological constant Λ leads to the following metric for a black hole:

$$\mathrm{d}\tau^2 = \frac{\mathrm{d}s^2}{c^2} = \left(1 - \frac{2GM}{rc^2} - \frac{\Lambda}{3} r^2\right) \mathrm{d}t^2 - \frac{\mathrm{d}r^2}{c^2 - \frac{2GM}{r} - \frac{\Lambda c^2}{3} r^2} - \frac{r^2}{c^2} \mathrm{d}\varphi^2 \ . \qquad (286)$$

Note that this metric does not turn into the Minkowski metric for large values of r. However, in the case that Λ is small, the metric is almost flat for values of r that satisfy $1/\sqrt{\Lambda} \gg r \gg 2GM/c^2$.

As a result, the inverse square law is also modified:

$$a = -\frac{GM}{r^2} + \frac{\Lambda c^2}{3} r \ . \qquad (287)$$

With the known values of the cosmological constant, the second term is negligible inside the solar system.

* *

In quantum theory, the *gyromagnetic ratio* is an important quantity for any rotating

Challenge 402 ny charged system. What is the gyromagnetic ratio for rotating black holes?

* *

A large black hole is, as the name implies, black. Still, it can be seen. If we were to travel towards it in a spaceship, we would note that the black hole is surrounded by a bright rim, like a thin halo, as shown in Figure 117. The ring at the radial distance of the photon sphere is due to those photons which come from other luminous objects, then circle the hole, and finally, after one or several turns, end up in our eye. Can you confirm this result? Challenge 403 s

* *

Challenge 404 ny Do moving black holes Lorentz-contract? Black holes do shine a little bit. It is true that the images they form are complex, as light can turn around them a few times before reaching the observer. In addition, the observer has to be far away, so that the effects of curvature are small. All these effects can be taken into account; nevertheless, the question remains subtle. The reason is that the concept of Lorentz contraction makes no sense in general relativity, as the comparison with the uncontracted situation is difficult to define precisely.

* *

Are black holes made of space or of matter? Both answers are correct! Can you confirm this? Challenge 405 s

* *

Power is energy change over time. General relativity limits power values to $P \leqslant c^5/4G$. In other words, no engine in nature can provide more than $0.92 \cdot 10^{52}$ W or $1.2 \cdot 10^{49}$ horsepower. Can you confirm that black holes support this limit? Challenge 406 e

* *

Black holes produce problems in the microscopic domain, where quantum theory holds, as was pointed out by Jürgen Ehlers. Ref. 280 Quantum theory is built on point particles, and point particles move on *time-like* world lines. But following general relativity, point particles have a singularity inside their black hole horizon; and singularities always move on *space-like* world lines. Microscopic black holes, in contrast to macroscopic black holes, thus contradict quantum theory.

Summary on black holes

A black hole is matter in permanent free fall. Equivalently, a black hole is a strongly curved type of space. Since black holes are defined through their horizon, they can be seen either as limiting cases of matter systems or as limiting cases of curved empty space.

Black holes realize the maximum force. For a given mass value, black holes also realize maximum density, maximum blackness and maximum entropy. Black holes deflect, capture and emit matter and light in peculiar ways.

A QUIZ – IS THE UNIVERSE A BLACK HOLE?

Could it be that we live inside a black hole? Both the universe and black holes have horizons. Interestingly, the horizon distance r_0 of the universe is about

$$r_0 \approx 3ct_0 \approx 4 \cdot 10^{26}\ \mathrm{m} \tag{288}$$

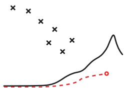
and its matter content is about

$$m_0 \approx \frac{4\pi}{3}\rho_o r_0^3 \quad \text{whence} \quad \frac{2Gm_0}{c^2} = 72\pi G\rho_0 c t_0^3 = 6 \cdot 10^{26}\ \mathrm{m} \tag{289}$$

for a density of $3 \cdot 10^{-27}\ \mathrm{kg/m^3}$. Thus we have

$$r_0 \approx \frac{2Gm_0}{c^2}\ , \tag{290}$$

which is similar to the black hole relation $r_S = 2Gm/c^2$. Is this a coincidence? No, it is not: all systems with high curvature more or less obey this relation. But are we nevertheless falling into a large black hole? You can answer that question by yourself. Challenge 407 s



CHAPTER 10

DOES SPACE DIFFER FROM TIME?

> “Tempori parce.**
>
> Seneca”

Time is our master, says a frequently heard statement. Nobody says that of space. Time and space are obviously different in everyday life. But what is the difference between them in general relativity? Do we need them at all? These questions are worth an exploration.

General relativity states that we live in a (pseudo-Riemannian) space-time of variable curvature. The curvature is an observable and is related to the distribution and motion of matter and energy. The precise relation is described by the field equations. However, there is a fundamental problem.

The equations of general relativity are invariant under numerous transformations which *mix* the coordinates x_0, x_1, x_2 and x_3. For example, the viewpoint transformation

$$\begin{aligned} x'_0 &= x_0 + x_1 \\ x'_1 &= -x_0 + x_1 \\ x'_2 &= x_2 \\ x'_3 &= x_3 \end{aligned} \qquad (291)$$

is allowed in general relativity, and leaves the field equations invariant. You might want to search for other examples of transformations that follow from diffeomorphism invariance. Challenge 408 e

Viewpoint transformations that mix space and time imply a consequence that is clearly in sharp contrast with everyday life: diffeomorphism invariance makes it *impossible* to distinguish space from time *inside* general relativity. More explicitly, the coordinate x_0 cannot simply be identified with the physical time t, as we implicitly did up to now. This identification is only possible in *special* relativity. In special relativity the invariance under Lorentz (or Poincaré) transformations of space and time singles out energy, linear momentum and angular momentum as the fundamental observables. In general relativity, there is *no* (non-trivial) metric isometry group; consequently, there are *no* basic physical observables singled out by their characteristic of being conserved. But

---

** ‘Care about time.’ Lucius Annaeus Seneca (*c.* 4 BCE–65), *Epistolae* 14, 94, 28.

invariant quantities are necessary for communication! In fact, we can *talk* to each other only because we live in an approximately *flat* space-time: if the angles of a triangle did not add up to π (two right angles), there would be *no* invariant quantities and we would not be able to communicate.

How have we managed to sweep this problem under the rug so far? We have done so in several ways. The simplest way was to always require that in some part of the situation under consideration space-time was our usual flat Minkowski space-time, where x_0 can be identified with t. We can fulfil this requirement either at infinity, as we did around spherical masses, or in zeroth approximation, as we did for gravitational radiation and for all other perturbation calculations. In this way, we eliminate the free mixing of coordinates and the otherwise missing invariant quantities appear as expected. This pragmatic approach is the usual way out of the problem. In fact, it is used in some otherwise excellent texts on general relativity that preclude any deeper questioning of the issue. Ref. 235

A common variation of this trick is to let the distinction between space and time 'sneak' into the calculations by the introduction of matter and its properties, or by the introduction of radiation, or by the introduction of measurements. The material properties of matter, for example their thermodynamic state equations, always distinguish between space and time. Radiation does the same, by its propagation. Obviously this is true also for those special combinations of matter and radiation called clocks and metre bars. Both matter and radiation distinguish between space and time simply by their presence.

In fact, if we look closely, the method of introducing matter to distinguish space and time is the same as the method of introducing Minkowski space-time in some limit: all properties of matter are defined using flat space-time descriptions.*

Another variation of the pragmatic approach is the use of the cosmological time coordinate. An isotropic and homogeneous universe does have a preferred time coordinate, namely the one time coordinate that is used in all the tables on the past and the future of the universe. This method is in fact a combination of the previous two. Page 231 Vol. III, page 347

But we are on a special quest here. We want to *understand* motion in principle, not only to calculate it in practice. We want a *fundamental* answer, not a pragmatic one. And for this we need to know how the positions x_i and time t are connected, and how we can define invariant quantities. The question also prepares us for the task of combining gravity with quantum theory, which is the aim of the final part of our adventure.

A fundamental solution to the problem requires a description of clocks together with the system under consideration, and a deduction of how the reading t of a clock relates to the behaviour of the system in space-time. But we know that any description of a system requires measurements: for example, in order to determine the initial conditions. And initial conditions require space and time. We thus enter a vicious circle: that is precisely what we wanted to avoid in the first place.

A suspicion arises. Is there in fact a fundamental difference between space and time? Let us take a tour of various ways to investigate this question.

---

* We note something astonishing here: the inclusion of some condition at small distances (the description of matter) has the same effect as the inclusion of some condition at infinity (the asymptotic Minkowski space). Is this just coincidence? We will come back to this issue in the last part of our adventure. Challenge 409 ny Vol. VI, page 113

Can space and time be measured?

Vol. I, page 439

In order to distinguish between space and time in general relativity, we must be able to measure them. But already in the section on universal gravity we have mentioned the impossibility of measuring lengths, times and masses with gravitational effects alone. Does this situation change in general relativity? Lengths and times are connected by the speed of light, and in addition lengths and masses are connected by the gravitational constant. Despite this additional connection, it takes only a moment to convince oneself that the problem persists.

Ref. 281

In fact, we need *electrodynamics* and the *granularity of matter* to perform measurements. In other words, we need the elementary charge e in order to form length scales. The simplest one is

$$l_{\text{em scale}} = \frac{e}{\sqrt{4\pi\varepsilon_0}} \frac{\sqrt{G}}{c^2} \approx 1.4 \cdot 10^{-36}\ \text{m} . \qquad (292)$$

Vol. III, page 26

Here, ε_0 is the permittivity of free space. Alternatively, we can argue that *quantum physics* provides a length scale, since we can use the quantum of action $\hbar$ to define the length scale

$$l_{\text{qt scale}} = \sqrt{\frac{\hbar G}{c^3}} \approx 1.6 \cdot 10^{-35}\ \text{m} , \qquad (293)$$

which is called the *Planck length* or *Planck's natural length unit*. However, this does not change the argument, because we need electrodynamics to measure the value of $\hbar$.

The equivalence of the two arguments is shown by rewriting the elementary charge e as a combination of nature's fundamental constants:

$$e = \sqrt{4\pi\varepsilon_0 c\hbar\alpha} . \qquad (294)$$

Here, $\alpha \approx 1/137.06$ is the fine-structure constant that characterizes the strength of electromagnetism. In terms of α, expression (292) becomes

$$l_{\text{em scale}} = \sqrt{\frac{\alpha\hbar G}{c^3}} = \sqrt{\alpha}\, l_{\text{qt scale}} . \qquad (295)$$

Summing up:

▷ Every length measurement is based on the electromagnetic coupling constant α and on the Planck length.

Challenge 410 e

Of course, the same is true for every time and every mass measurement. There is thus no way to define or measure lengths, times and masses using gravitation or general relativity only.*

Ref. 282

Challenge 411 s

[*] In the past, John Wheeler used to state that his *geometrodynamic clock*, a device which measures time by bouncing light back and forth between two parallel mirrors, was a counter-example. However, that is not correct. Can you confirm this?

Given this sobering result, we can take the opposite point of view. We ask whether in general relativity space and time are required at all.

ARE SPACE AND TIME NECESSARY?

Ref. 283 Robert Geroch answers this question in a beautiful five-page article. He explains how to formulate the general theory of relativity *without the use of space and time*, by taking as starting point the physical observables only.

Geroch starts with the set of all observables. Among them there is one, called v, which stands out. It is the only observable which allows one to say that for any two observables a_1, a_2 there is a third one a_3, for which

$$(a_3 - v) = (a_1 - v) + (a_2 - v) \,. \qquad (296)$$

Such an observable is called the *vacuum*. Geroch shows how to use such an observable to construct derivatives of observables. Then he builds the so-called Einstein algebra, which comprises the whole of general relativity.

Usually in general relativity, we describe motion in three steps: we deduce space-time from matter observables, we calculate the evolution of space-time, and then we deduce the motion of matter that follows from space-time evolution. Geroch's description shows that the middle step, and thus the use of space and time, is unnecessary.

Indirectly, the principle of maximum force makes the same statement. General relativity can be derived from the existence of limit values for force or power. Space and time are only tools needed to translate this principle into consequences for real-life observers.

In short, it is possible to formulate general relativity *without* the use of space and time. Since both are unnecessary, it seems unlikely that there should be a fundamental difference between them. Nevertheless, one difference is well-known.

DO CLOSED TIME-LIKE CURVES EXIST?

Is it possible that the time coordinate behaves, at least in some regions, like a torus? When we walk, we can return to the point of departure. Is it possible, to come *back in time* to where we have started? The question has been studied in great detail.

Ref. 241 The standard reference on closed time-like curves is the text by Hawking and Ellis; they list the required properties of space-time, explaining which are mutually compatible or exclusive. They find, for example, that space-times which are smooth, globally hyperbolic, oriented and time-oriented do not contain any such curves. It is usually assumed that the observed universe has these properties, so that the actual observation of closed time-like curves is unlikely. Indeed, no candidate has ever been suggested – even though it would be a scientific sensation. Later on, we will find that also searches for such curves at the microscopic scale have also failed to find any example in nature. Vol. V, page 159

In summary, there are *no* closed time-like curves in nature. The impossibility of such curves seems to point to a difference between space and time. But in fact, this difference is only apparent. These investigations are based on the behaviour of matter. Thus all arguments assume a specific distinction between space and time right from the start. In short, this line of enquiry cannot help us to decide whether space and time differ. Therefore we look at the issue in another way.

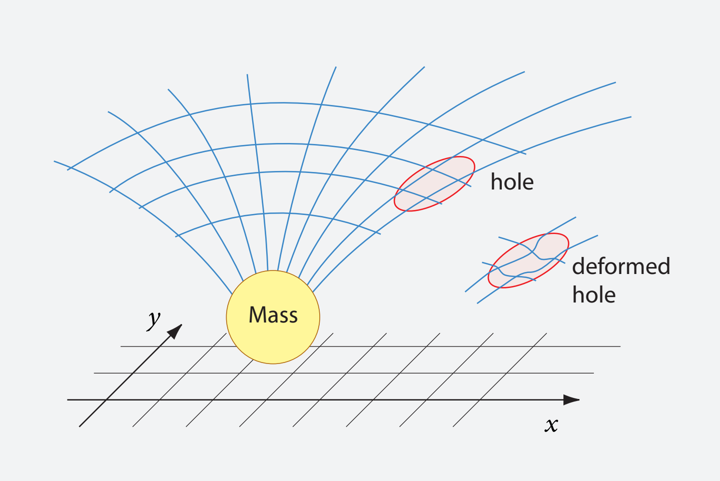


FIGURE 123 A ‘hole’ in space in a schematic view.

Is general relativity local? – The hole argument

When Albert Einstein developed general relativity, he had quite some trouble with diffeomorphism invariance. Most startling is his famous *hole argument*, better called the *hole paradox*. Take the situation shown in Figure 123, in which a mass deforms the space-time around it. Einstein imagined a small region of the vacuum, the *hole*, which is shown as a small ellipse. What happens if we somehow change the curvature inside the hole while leaving the situation outside it unchanged, as shown in the inset of the picture? Ref. 284

On the one hand, the new situation is obviously physically different from the original one, as the curvature inside the hole is different. This difference thus implies that the curvature outside a region does not determine the curvature inside it. That is extremely unsatisfactory. Worse, if we generalize this operation to the time domain, we seem to get the biggest nightmare possible in physics: determinism is lost.

On the other hand, general relativity is diffeomorphism invariant. The deformation shown in the figure is a diffeomorphism; so the new situation must be physically equivalent to the original situation.

Which argument is correct? Einstein first favoured the first point of view, and therefore dropped the whole idea of diffeomorphism invariance for about a year. Only later did he understand that the second assessment is correct, and that the first argument makes a fundamental mistake: it assumes an independent existence of the coordinate axes x and y, as shown in the figure. But during the deformation of the hole, the coordinates x and y automatically change as well, so that there is *no* physical difference between the two situations.

The moral of the story is that *there is no difference between space-time and the gravitational field*. Space-time is a quality of the field, as Einstein put it, and not an entity with a separate existence, as suggested by the graph. Coordinates have no physical meaning; only distances (intervals) in space and time have one. In particular, diffeomorphism invariance proves that *there is no flow of time*. Time, like space, is only a relational entity: time and space are relative; they are not absolute.

The relativity of space and time has practical consequences. For example, it turns out

FIGURE 124 A model of the hollow Earth theory (© Helmut Diehl).

that many problems in general relativity are equivalent to the Schwarzschild situation, even though they appear completely different at first sight. As a result, researchers have ‘discovered’ the Schwarzschild solution (of course with different coordinate systems) over twenty times, often thinking that they had found a new, unknown solution. We now discuss a startling consequence of diffeomorphism invariance.

IS THE EARTH HOLLOW?

> Any pair of shoes proves that we live on the inside of a sphere. Their soles are worn out at the ends, and hardly at all in between.
>
> Anonymous

The *hollow Earth hypothesis*, i.e., the bizarre conjecture that we live on the *inside* of a sphere, was popular in esoteric circles around the year 1900, and still remains so among certain eccentrics today, especially in Britain, Germany and the US. They maintain, as illustrated in Figure 124, that the solid Earth *encloses* the sky, together with the Moon, the Sun and the stars. Most of us are fooled by education into another description, because we are brought up to believe that light travels in straight lines. Get rid of this wrong belief, they say, and the hollow Earth appears in all its glory. Vol. I, page 60

Interestingly, the reasoning is partially correct. There is *no way* to disprove this sort of description of the universe. In fact, as the great physicist Roman Sexl used to explain, the diffeomorphism invariance of general relativity even proclaims the equivalence between the two views. The fun starts when either of the two camps wants to tell the other that *only* its own description can be correct. You might check that any such argument is wrong; it is fun to slip into the shoes of such an eccentric and to defend the hollow Earth hy- Ref. 285

Challenge 412 e pothesis against your friends. It is easy to explain the appearance of day and night, of the horizon, and of the satellite images of the Earth. It is easy to explain what happened during the flight to the Moon. You can drive many bad physicists crazy in this way! The usual description and the hollow Earth description are exactly equivalent. Can you confirm that even quantum theory, with its introduction of length scales into nature, does not change this situation? Challenge 413 s

In summary, diffeomorphism invariance is not an easy symmetry to swallow. But it is best to get used to it now, as the rest of our adventure will throw up even more surprises. Indeed, in the final part of our walk we will discover that there is an even larger symmetry of nature that is similar to the change in viewpoint from the hollow Earth view to the standard view. This symmetry, space-time duality, is valid not only for distances measured from the centre of the Earth, but for distances measured from any point in nature. Vol. VI, page 113

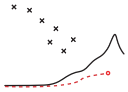
A summary: are space, time and mass independent?

We can conclude from this short discussion that there is no fundamental distinction between space and time in general relativity. The only possible distinctions are the pragmatic ones that make use of matter, radiation, or space-time at infinity.

Vol. I, page 438 In the beginning of our adventure we found that we needed matter to define space and time. Now we have found that we even need matter to distinguish *between* space *and* time. Similarly, in the beginning of our adventure we found that space and time are required to define matter; now we have found that we even need *flat* space-time to define matter. In these fundamental issues, general relativity has thus brought *no* improvement over the results of Galilean physics. Vol. I, page 203

In summary, general relativity does *not* provide a way out of the circular reasoning we discovered in Galilean physics. Indeed, general relativity makes the issue even less clear than before. Matter and radiation remain essential to define and distinguish space and time, and space and time remain essential to define and distinguish matter and radiation. Continuing our mountain ascent is the only way out.

In the next parts of our adventure, quantum physics will confirm that matter is needed to distinguish between space and time. No distinction between space and time *without matter* is possible in principle. Then, in the last part of our adventure, we will discover that mass and space are on an equal footing in nature. Because either is defined with the other, we will deduce that particles and vacuum are made of the same substance. It will turn out that distinctions between space and time are possible only at low, everyday energies; but no such distinction exists in principle. Vol. IV, page 167 Vol. VI, page 80



CHAPTER 11

GENERAL RELATIVITY IN A NUTSHELL – A SUMMARY FOR THE LAYMAN

> “Sapientia felicitas.[**]”
>
> Antiquity

General relativity is the final, correct description of *macroscopic motion*. General relativity describes, first of all, all macroscopic motion due to gravity, and in particular, describes how the observations of motion of *any* two observers are related to each other. Above all, general relativity describes the most rapid, the most powerful, the most violent and the most distant motions. For this reason, general relativity describes the motion of matter and of empty space, including the motion of horizons and the evolution of what is usually called the border of the universe.

The description of macroscopic motion with general relativity is *final* and *correct*. Calculations and predictions from general relativity match all observations where the match is possible. The match is not possible for dark matter; this issue is not settled yet.

General relativity is based on two principles deduced from observations:

- All observers agree that there is a ‘perfect’ speed in nature, namely a common *maximum* energy speed relative to (nearby) matter. The invariant speed value $c = 299\,792\,458\,\mathrm{m/s}$ is realized by massless radiation, such as light or radio signals.
- All observers agree that there is a ‘perfect’ force in nature, a common *maximum* force that can be realized relative to (nearby) matter. The invariant force value $F = c^4/4G = 3.0258(4) \cdot 10^{43}\,\mathrm{N}$ is realized on event horizons.

These two observations contain the full theory of relativity. In particular, from these two observations we deduce:

- Space-time consists of events in 3+1 *continuous dimensions*, with a variable curvature. The curvature can be deduced from distance measurements among events, for example from tidal effects. Measured times, lengths and curvatures vary from observer to observer in a predictable way. In short, we live in a *pseudo-Riemannian space-time*.
- Space-time and space are *curved near mass and energy*. The curvature at a point is determined by the energy–momentum density at that point, and described by the field equations. When matter and energy move, the space curvature moves along with them. A built-in delay in this movement renders faster-than-light transport of energy impossible. The proportionality constant between energy and curvature is so small

---

[**] ‘Wisdom is happiness.’ This old saying once was the motto of Oxford University.

that the curvature is not observed in everyday life; only its indirect manifestation, namely universal gravity, is observed.

- All macroscopic motion – that of matter, of radiation and of vacuum – is described by the field equations of general relativity.
- Space is *elastic*: it prefers being flat. Being elastic, it can oscillate independently of matter; one then speaks of gravitational radiation or of gravity waves.
- Freely falling matter moves along *geodesics*, i.e., along paths of maximal length in curved space-time. In space this means that light bends when it passes near large masses by twice the amount predicted by universal gravity.
- In order to describe gravitation we need *curved* space-time, i.e., general relativity, *at the latest* whenever distances are of the order of the Schwarzschild radius $r_S = 2Gm/c^2$. When distances are much larger than this value, the relativistic description with gravity and gravitomagnetism (frame-dragging) is sufficient. When distances are even larger and speeds much slower than those of light, the description by universal gravity, namely $a = Gm/r^2$, together with flat Minkowski space-time, will do as a good approximation.
- Space and time are not distinguished globally, but only locally. *Matter* and *radiation* are required to make the distinction.

In addition, all the matter and energy we observe in the sky lead us to the following conclusions:

- The universe has a finite size, given roughly by $r_{\text{max}} \approx 1/\sqrt{\Lambda} \approx 10^{26}$ m. The cosmological constant Λ also has the effect of an energy density. One speaks of *dark energy*. Page 199
- The universe has a *finite age*; this is the reason for the darkness of the sky at night. A horizon limits the measurable space-time intervals to about fourteen thousand million years.
- On the cosmological scale, everything moves away from everything else: the universe is *expanding*. The details of the underlying expansion of space, as well as the night-sky horizon, are described by the field equations of general relativity.

In short, experiments show that all motion of energy, including matter and radiation, is limited in speed and in momentum flow, or force. A maximum force implies that space curves and that the curvature can move. The well-known basic properties of everyday motion remain valid: also relativistic motion that includes gravity is continuous, conserves energy–momentum and angular momentum, is relative, is reversible, is mirror-invariant (except for the weak interaction, where a generalized way to predict mirror-inverse motion holds). Above all, like everyday motion, also every example of relativistic motion that includes gravity is lazy: all motion minimizes action. Vol. I, page 29 Vol. V, page 245

In summary, the principles of maximum force and of maximum speed hold for every motion in nature. They are universal truths. The theory of general relativity that follows from the two principles describes all *macroscopic* motion that is observed in the universe, including the most rapid, the most powerful and the most distant motions known – be it motion of matter, radiation, vacuum or horizons.

THE ACCURACY OF THE DESCRIPTION

Was general relativity worth the effort? The discussion of its accuracy is most conveniently split into two sets of experiments. The first set consists of measurements of how *matter moves*. Do objects really follow geodesics? As summarized in Table 9, all experiments agree with the theory to within measurement errors, i.e., at least within 1 part in 10^{12}. In short, the way matter falls is indeed well described by general relativity. Ref. 286 Ref. 287

The second set of measurements concerns the dynamics of space-time itself. Does *space-time move* following the field equations of general relativity? In other words, is space-time really bent by matter in the way the theory predicts? Many experiments have been performed, near to and far from Earth, in both weak and strong gravitational fields. All agree with the predictions to within measurement errors. However, the best measurements so far have only about 3 significant digits. Note that even though numerous experiments have been performed, there are only few *types* of tests, as Table 9 shows. The discovery of a new type of experiment almost guarantees fame and riches. Most sought after, of course, is the direct detection of gravitational waves. Ref. 286, Ref. 287 Challenge 414 ny

Another comment on Table 9 is in order. After many decades in which all measured effects were only of the order v^2/c^2, several so-called *strong field effects* in pulsars allowed us to reach the order v^4/c^4. Soon a few effects of this order should also be detected even inside the solar system, using high-precision satellite experiments. The present crown of all measurements, the gravity wave emission delay, is the only v^5/c^5 effect measured so far. Ref. 286 Page 180

The difficulty of achieving high precision for space-time curvature measurements is the reason why mass is measured with balances, always (indirectly) using the prototype kilogram in Paris, instead of defining some standard curvature and fixing the value of G. Indeed, no useful terrestrial curvature experiment has ever been carried out. A breakthrough in this domain would make the news. The terrestrial curvature methods currently available would not even allow one to define a kilogram of oranges or peaches with a precision high enough to distinguish it from the double amount!

A different way to check general relativity is to search for alternative descriptions of gravitation. Quite a number of alternative theories of gravity have been formulated and studied, but so far, only general relativity is in agreement with all experiments. Ref. 287, Ref. 288

In summary, as Thibault Damour likes to explain, general relativity is at least 99.999 999 999 9 % correct concerning the motion of matter and energy, and at least 99.9 % correct about the way matter and energy curve and move space-time. No exceptions, no anti-gravity and no unclear experimental data are known. All motion on Earth and in the skies is described by general relativity. The most violent, the most rapid and the most distant movements known behave as expected. Albert Einstein's achievement has no flaws. Ref. 286

We note that general relativity has not been tested for microscopic motion. In this context, *microscopic motion* is any motion for which the action value is near the quantum of action $\hbar$, namely 10^{-34} Js. The exploration of microscopic motion in strong gravitational fields is the topic of the last part of our adventure.

TABLE 9 Types of tests of general relativity.

Measured effect	Confirmation	Type	Reference
Equivalence principle	10^{-12}	motion of matter	Ref. 156, Ref. 286, Ref. 289
$1/r^2$ dependence (dimensionality of space-time)	10^{-10}	motion of matter	Ref. 290
Time independence of G	10^{-19} /s	motion of matter	Ref. 286
Red-shift (light and microwaves on Sun, Earth, Sirius)	10^{-4}	space-time curvature	Ref. 135, Ref. 133, Ref. 286
Perihelion shift (four planets, Icarus, pulsars)	10^{-3}	space-time curvature	Ref. 286
Light deflection (light, radio waves around Sun, stars, galaxies)	10^{-3}	space-time curvature	Ref. 286
Time delay (radio signals near Sun, near pulsars)	10^{-3}	space-time curvature	Ref. 286, Ref. 175
Gravitomagnetism (Earth, pulsar)	10^{-1}	space-time curvature	Ref. 167, Ref. 168
Geodesic effect (Moon, pulsars)	10^{-1}	space-time curvature	Ref. 164, Ref. 286
Gravity wave emission delay (pulsars)	10^{-3}	space-time curvature	Ref. 286

RESEARCH IN GENERAL RELATIVITY AND COSMOLOGY

Research in general relativity is still intense, though declining; it is declining most strongly in Switzerland and Germany, the countries where Albert Einstein developed the theory. Research in cosmology and astrophysics, however, is at a high point at present. Here is a short overview. Ref. 291

* *

The most interesting experimental studies of general relativity are the tests using double pulsars, the search for gravitational waves, and the precision measurements using satellites. Among others a special satellite will capture all possible pulsars of the galaxy. All these experiments expand the experimental tests into domains that have not been accessible before. So far, all tests completely confirm general relativity – as expected.

* *

The investigation of cosmic collisions and many-body problems, especially those involving neutron stars and black holes, helps astrophysicists to improve their understanding of the rich behaviour they observe in their telescopes. Ref. 269

* *

The study of chaos in the field equations is of fundamental interest in the study of the early universe, and may be related to the problem of galaxy formation, one of the biggest open problems in physics. Ref. 292

* *

Gathering data about galaxy formation is the main aim of several satellite systems and purpose-build telescopes. One focus is the search for localized cosmic microwave background anisotropies due to protogalaxies. Ref. 293

* *

The precise determination of the cosmological parameters, such as the matter density, the curvature and the vacuum density, is a central effort of modern astrophysics. Ref. 237 The exploration of vacuum density – also called *cosmological constant* or *dark energy* – and the clarification of the nature of *dark matter* occupy a large fraction of astrophysicists.

* *

Astronomers and astrophysicists regularly discover new phenomena in the skies. The various types of gamma-ray bursts, X-ray bursts and optical bursts are still not completely understood. Ref. 294 Gamma-ray bursts, for example, can be as bright as 10^{17} sun-like stars combined; however, they last only a few seconds. More details on this research topic are given later on. Vol. V, page 153

* *

A computer database of all known exact solutions of the field equations is being built. Among other things, researchers are checking whether they really are all different from each other. Ref. 295

* *

Ref. 296 Solutions of the field equations with non-trivial topology, such as wormholes and particle-like solutions, constitue a fascinating field of enquiry. However, such solutions are made impossible by quantum effects. Vol. V, page 159

* *

Other formulations of general relativity, describing space-time with quantities other than the metric, are continuously being developed, in the hope of clarifying the relationship between gravity and the quantum world. The so-called Ashtekar variables are such a modern description. Ref. 297

* *

The study of the early universe and its relation of elementary particle properties, with conjectures such as *inflation*, a short period of accelerated expansion during the first few seconds after the big bang, is still an important topic of investigation. Ref. 298, Ref. 299

* *

The unification of quantum physics, particle physics and general relativity is an important research field and will occupy researchers for many years to come. The aim is to find a complete description of motion. This is the topic of the final part of this adventure. Vol. VI, page 17

* *

Finally, the teaching of general relativity, which for many decades has been hidden behind Greek indices, differential forms and other antididactic approaches, will benefit greatly from future improvements that focus more on the physics and less on the formalism. Ref. 300

COULD GENERAL RELATIVITY BE DIFFERENT?

> It's a good thing we have gravity, or else when
> birds died they'd just stay right up there.
> Hunters would be all confused.
>
> Steven Wright

The constant of gravitation provides a limit for the density and the acceleration of objects, as well as for the power of engines. We based all our deductions on its invariance. Is it possible that the constant of gravitation G changes from place to place or that it changes with time? The question is tricky. At first sight, the answer is a loud: 'Yes, of course! Just see what happens when the value of G is changed in formulae.' However, this answer is *wrong*. It is as wrong as it was wrong for the speed of light c. Page 104

Since the constant of gravitation enters into our definition of gravity and acceleration, it thus enters, even if we do not notice it, into the construction of all rulers, all measurement standards and all measuring instruments. Therefore there is *no way* to detect whether its value actually varies.

▷ A change in the maximum force and thus in the gravitational constant G cannot be measured.

Yes, the invariance of the limit force and of G is counter-intuitive. Experiments are able to detect the *existence* of a maximum force. Nevertheless, no imaginable experiment could detect a *variation* of the maximum force value, neither over space nor over time. Just try! Challenge 415 e Every measurement of force is, whether we like it or not, a comparison with the limit force. There is no way, in principle, to falsify the invariance of a measurement standard. This is even more astonishing because measurements of this type are regularly reported. Page 290 In a sense, Table 9 is a list of such experiments! But the result of any such experiment is easy to predict: no change will ever be found and no deviation from general relativity will ever be found.

Are other changes possible? Could the number of space dimensions be different from 3? This issue is quite involved. For example, three is the smallest number of dimensions for which a vanishing Ricci tensor is compatible with non-vanishing curvature. On the other hand, more than three dimensions would give deviations from the inverse square 'law' of gravitation. There are no data pointing in this direction. All experiments confirm that space has exactly three dimensions.

Could the equations of general relativity be different? During the past century, theoreticians have explored many alternative equations. However, almost none of the alternatives proposed so far seem to fit experimental data – nor the existence of a maximum force. Only two candidates are still regularly discussed.

First, the inclusion of *torsion* in the field equations is one attempt to include particle

Ref. 302 spin in general relativity. The inclusion of torsion in general relativity does not require new fundamental constants; indeed, the absence of torsion was assumed in the Ref. 301 Raychaudhuri equation. The use of the extended Raychaudhuri equation, which includes torsion, might allow one to deduce the full Einstein–Cartan theory from the maximum force principle. However, all arguments so far suggest that torsion is an unnecessary complication.

Secondly, one issue remains unexplained: the question of the existence of dark matter. The rotation speed of visible matter far from the centre of galaxies *might* be due to the existence of dark matter or to some deviation from the inverse square dependence of universal gravity. The latter option would imply a modification in the field equations for Ref. 303 astronomically large distances. The dark matter option assumes that we have difficulties observing something, the modified dynamics option assumes that we missed something in the equations. Also certain experiments about light deflection seem to point to some invisible kind of matter spread around galaxies. Is this a new form of matter? At present, most researchers tend to assume the existence of dark matter, and further assume that is some unknown type of matter. But since the nature of dark matter is not understood, and since it has never been detected in the laboratory, the issue is not settled.

In summary, given the principle of maximum force, it seems extremely unlikely if not impossible that nature is not described by general relativity.

> “It was, of course, a lie what you read about my religious convictions, a lie which is being systematically repeated. I do not believe in a personal God and I have never denied this but have expressed it clearly. If something is in me which can be called religious then it is the unbounded admiration for the structure of the world so far as our science can reveal it.”
>
> Albert Einstein, 24 March 1954.

THE LIMITATIONS OF GENERAL RELATIVITY

Despite its success and its fascination, the description of motion presented so far is unsatisfactory; maybe you already have some gut feeling about certain unresolved issues. Challenge 416 e

First of all, even though the speed of light is the starting point of the whole theory, we still do not know what light actually *is*. Understanding what light is will be our next topic.

Secondly, we have seen that everything that has mass falls along geodesics. But a mountain does not fall. Somehow the matter below prevents it from falling. How? And where does mass come from anyway? What is matter? General relativity does not provide any answer; in fact, it does not describe matter *at all*. Einstein used to say that the left-hand side of the field equations, describing the curvature of space-time, was granite, while the right-hand side, describing matter, was sand. Indeed, at this point we still do not know what matter and mass *are*. (And we know even less what dark matter is.) As already remarked, to change the sand into rock we first need quantum physics and then, in a further step, its unification with relativity. This is the programme for the rest of our adventure.

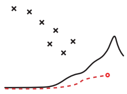
We have also seen that matter is necessary to clearly distinguish between space and time, and in particular, to understand the working of clocks, metre bars and balances. But one question remains: why are there units of mass, length and time in nature *at all*? Understanding why measurements are possible at all will be another of the topics of quantum physics.

We also know too little about the vacuum. We need to understand the magnitude of the cosmological constant, its time dependence and the number of space-time dimensions. Only then can we answer the simple question: Why is the sky so far away? General relativity does not help here. We will find out that the observed smallness of the cosmological constant contradicts the simplest version of quantum theory; this is one of the reasons why we still have quite some height to scale before we reach the top of Motion Mountain. Vol. VI, page 58

Finally, we swept another important issue under the rug. General relativity forbids the existence of point objects, and thus of point particles. But the idea of point particles is one reason that we introduced space points in the first place. What is the final fate of the idea of space point? What does this imply for the properties of horizons, for black holes and the night sky? Also these issues remain open at this stage.

In short, to describe motion well, we need a more precise description of light, of matter and of the vacuum. In other words, we need to know more about everything! Otherwise we cannot hope to answer questions about mountains, clocks and stars. In particular, we need to know more about light, matter and vacuum at *small* scales. We need to understand the *microscopic* aspects of the world. Vol. V, page 67

At small scales, the curvature of space is negligible. We therefore take a step backwards, to situations *without* gravity, and explore the microscopic motion of light, matter and vacuum. This domain is called *quantum physics*. Figure 1, shown in the preface, gives an impression of the topics that await us. And despite the simplification to flat space-time, the adventure is beautiful and intense: we will explore the motion at the basis of life. Page 8



Appendix A

UNITS, MEASUREMENTS AND CONSTANTS

Measurements are comparisons with standards. Standards are based on *units*. Many different systems of units have been used throughout the world. Most of these standards confer power to the organization in charge of them. Such power can be misused; this is the case today, for example in the computer industry, and was so in the distant past. The solution is the same in both cases: organize an independent and global standard. For measurement units, this happened in the eighteenth century: in order to avoid misuse by authoritarian institutions, to eliminate problems with differing, changing and irreproducible standards, and – this is not a joke – to simplify tax collection and to make it more just, a group of scientists, politicians and economists agreed on a set of units. It is called the *Système International d'Unités*, abbreviated *SI*, and is defined by an international treaty, the 'Convention du Mètre'. The units are maintained by an international organization, the 'Conférence Générale des Poids et Mesures', and its daughter organizations, the 'Commission Internationale des Poids et Mesures' and the 'Bureau International des Poids et Mesures' (BIPM). All originated in the times just before the French revolution. Ref. 304

SI UNITS

All SI units are built from seven *base units*. Their simplest definitions, translated from French into English, are the following ones, together with the dates of their formulation and a few comments:

- 'The *second* is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium 133 atom.' (1967) The 2019 definition is equivalent, but much less clear.*
- 'The *metre* is the length of the path travelled by light in vacuum during a time interval of 1/299 792 458 of a second.' (1983) The 2019 definition is equivalent, but much less clear.*
- 'The *kilogram*, symbol kg, is the SI unit of mass. It is defined by taking the fixed numerical value of the Planck constant h to be $6.626\,070\,15 \cdot 10^{-34}$ when expressed in the unit $\mathrm{J \cdot s}$, which is equal to $\mathrm{kg \cdot m^2 \cdot s^{-1}}$.' (2019)*
- 'The *ampere*, symbol A, is the SI unit of electric current. It is defined by taking the fixed numerical value of the elementary charge e to be $1.602\,176\,634 \cdot 10^{-19}$ when expressed in the unit C, which is equal to $\mathrm{A \cdot s}$.' (2019)* This definition is equivalent to: One ampere is $6.241\,509\,074... \cdot 10^{18}$ elementary charges per second.
- 'The *kelvin*, symbol K, is the SI unit of thermodynamic temperature. It is defined by

taking the fixed numerical value of the Boltzmann constant k to be $1.380649 \cdot 10^{-23}$ when expressed in the unit $\mathrm{J} \cdot \mathrm{K}^{-1}$.' (2019)*

- 'The mole, symbol mol, is the SI unit of amount of substance. One mole contains exactly $6.02214076 \cdot 10^{23}$ elementary entities.' (2019)*
- 'The *candela* is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency $540 \cdot 10^{12}$ hertz and has a radiant intensity in that direction of (1/683) watt per steradian.' (1979) The 2019 definition is equivalent, but much less clear.*

We note that both time and length units are defined as certain properties of a standard example of motion, namely light. In other words, also the Conférence Générale des Poids et Mesures makes the point that the observation of motion is a *prerequisite* for the definition and construction of time and space. *Motion is the fundament of every observation and of all measurement.* By the way, the use of light in the definitions had been proposed already in 1827 by Jacques Babinet.**

From these basic units, all other units are defined by multiplication and division. Thus, all SI units have the following properties:

- SI units form a system with *state-of-the-art precision*: all units are defined with a precision that is higher than the precision of commonly used measurements. Moreover, the precision of the definitions is regularly being improved. The present relative uncertainty of the definition of the second is around 10^{-14}, for the metre about 10^{-10}, for the kilogram about 10^{-9}, for the ampere 10^{-7}, for the mole less than 10^{-6}, for the kelvin 10^{-6} and for the candela 10^{-3}.
- SI units form an *absolute* system: all units are defined in such a way that they can be reproduced in every suitably equipped laboratory, independently, and with high precision. This avoids as much as possible any error or misuse by the standard-setting organization. In fact, the SI units are as now as near as possible to Planck's natural units, which are presented below. In practice, the SI is now an international standard defining the numerical values of the seven constants $\Delta\nu_{\mathrm{Cs}}$, c, $\hbar$, e, k, N_{A} and K_{cd}. After over 200 years of discussions, the CGPM has little left to do.
- SI units form a *practical* system: the base units are quantities of everyday magnitude. Frequently used units have standard names and abbreviations. The complete list includes the seven base units just given, the supplementary units, the derived units and the admitted units.

The *supplementary* SI units are two: the unit for (plane) angle, defined as the ratio of arc length to radius, is the *radian* (rad). For solid angle, defined as the ratio of the subtended area to the square of the radius, the unit is the *steradian* (sr).

The *derived* units with special names, in their official English spelling, i.e., without capital letters and accents, are:

Ref. 305

* The symbols of the seven units are s, m, kg, A, K, mol and cd. The full offical definitions are found at www.bipm.org. For more details about the levels of the caesium atom, consult a book on atomic physics. The Celsius scale of temperature θ is defined as: $\theta/{}^\circ\mathrm{C} = T/\mathrm{K} - 273.15$; note the small difference with the number appearing in the definition of the kelvin. In the definition of the candela, the frequency of the light corresponds to 555.5 nm, i.e., green colour, around the wavelength to which the eye is most sensitive.

** Jacques Babinet (1794–1874), French physicist who published important work in optics.

Name	Abbreviation
hertz	$\mathrm{Hz} = 1/\mathrm{s}$
pascal	$\mathrm{Pa} = \mathrm{N/m^2} = \mathrm{kg/m\,s^2}$
watt	$\mathrm{W} = \mathrm{kg\,m^2/s^3}$
volt	$\mathrm{V} = \mathrm{kg\,m^2/As^3}$
ohm	$\Omega = \mathrm{V/A} = \mathrm{kg\,m^2/A^2s^3}$
weber	$\mathrm{Wb} = \mathrm{Vs} = \mathrm{kg\,m^2/As^2}$
henry	$\mathrm{H} = \mathrm{Vs/A} = \mathrm{kg\,m^2/A^2s^2}$
lumen	$\mathrm{lm} = \mathrm{cd\,sr}$
becquerel	$\mathrm{Bq} = 1/\mathrm{s}$
sievert	$\mathrm{Sv} = \mathrm{J/kg} = \mathrm{m^2/s^2}$

Name	Abbreviation
newton	$\mathrm{N} = \mathrm{kg\,m/s^2}$
joule	$\mathrm{J} = \mathrm{Nm} = \mathrm{kg\,m^2/s^2}$
coulomb	$\mathrm{C} = \mathrm{As}$
farad	$\mathrm{F} = \mathrm{As/V} = \mathrm{A^2s^4/kg\,m^2}$
siemens	$\mathrm{S} = 1/\Omega$
tesla	$\mathrm{T} = \mathrm{Wb/m^2} = \mathrm{kg/As^2} = \mathrm{kg/Cs}$
degree Celsius	°C (see definition of kelvin)
lux	$\mathrm{lx} = \mathrm{lm/m^2} = \mathrm{cd\,sr/m^2}$
gray	$\mathrm{Gy} = \mathrm{J/kg} = \mathrm{m^2/s^2}$
katal	$\mathrm{kat} = \mathrm{mol/s}$

We note that in all definitions of units, the kilogram only appears to the powers of 1, 0 and −1. Can you try to formulate the reason?

Challenge 417 s

The *admitted* non-SI units are *minute*, *hour*, *day* (for time), *degree* $1° = \pi/180\,\mathrm{rad}$, *minute* $1' = \pi/10\,800\,\mathrm{rad}$, *second* $1'' = \pi/648\,000\,\mathrm{rad}$ (for angles), *litre*, and *tonne*. All other units are to be avoided.

All SI units are made more practical by the introduction of standard names and abbreviations for the powers of ten, the so-called *prefixes*:*

Power	Name	
10^{1}	deca	da
10^{2}	hecto	h
10^{3}	kilo	k
10^{6}	Mega	M
10^{9}	Giga	G
10^{12}	Tera	T
10^{15}	Peta	P

Power	Name	
10^{-1}	deci	d
10^{-2}	centi	c
10^{-3}	milli	m
10^{-6}	micro	µ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f

Power	Name	
10^{18}	Exa	E
10^{21}	Zetta	Z
10^{24}	Yotta	Y
unofficial:		
10^{27}	Xenta	X
10^{30}	Wekta	W
10^{33}	Vendekta	V
10^{36}	Udekta	U

Ref. 306

Power	Name	
10^{-18}	atto	a
10^{-21}	zepto	z
10^{-24}	yocto	y
10^{-27}	xenno	x
10^{-30}	weko	w
10^{-33}	vendeko	v
10^{-36}	udeko	u

- SI units form a *complete* system: they cover in a systematic way the full set of observables of physics. Moreover, they fix the units of measurement for all other sciences as well.

---

* Some of these names are invented (yocto to sound similar to Latin *octo* ‘eight’, zepto to sound similar to Latin *septem*, yotta and zetta to resemble them, exa and peta to sound like the Greek words ἑξάκις and πεντάκις for ‘six times’ and ‘five times’, the unofficial ones to sound similar to the Greek words for nine, ten, eleven and twelve); some are from Danish/Norwegian (atto from *atten* ‘eighteen’, femto from *femten* ‘fifteen’); some are from Latin (from *mille* ‘thousand’, from *centum* ‘hundred’, from *decem* ‘ten’, from *nanus* ‘dwarf’); some are from Italian (from *piccolo* ‘small’); some are Greek (micro is from μικρός ‘small’, deca/deka from δέκα ‘ten’, hecto from ἑκατόν ‘hundred’, kilo from χίλιοι ‘thousand’, mega from μέγας ‘large’, giga from γίγας ‘giant’, tera from τέρας ‘monster’).

Translate: I was caught in such a traffic jam that I needed a microcentury for a picoparsec and that my car’s fuel consumption was two tenths of a square millimetre.

Challenge 418 e

- SI units form a *universal* system: they can be used in trade, in industry, in commerce, at home, in education and in research. They could even be used by extraterrestrial civilizations, if they existed.
- SI units form a *self-consistent* system: the product or quotient of two SI units is also an SI unit. This means that in principle, the same abbreviation, e.g. 'SI', could be used for every unit.

The SI units are not the only possible set that could fulfil all these requirements, but they are the only existing system that does so.*

The meaning of measurement

Challenge 419 e

Every measurement is a comparison with a standard. Therefore, any measurement requires *matter* to realize the standard (even for a speed standard), and *radiation* to achieve the comparison. The concept of measurement thus assumes that matter and radiation exist and can be clearly separated from each other.

Every measurement is a comparison. Measuring thus implies that space and time exist, and that they differ from each other.

Every measurement produces a measurement result. Therefore, every measurement implies the *storage* of the result. The process of measurement thus implies that the situation before and after the measurement can be distinguished. In other terms, every measurement is an *irreversible* process.

Every measurement is a process. Thus every measurement takes a certain amount of time and a certain amount of space.

All these properties of measurements are simple but important. Beware of anybody who denies them.

Curiosities and fun challenges about units

The second does not correspond to 1/86 400th of the day any more, though it did in the year 1900; the Earth now takes about 86 400.002 s for a rotation, so that the *International Earth Rotation Service* must regularly introduce a leap second to ensure that the Sun is at the highest point in the sky at 12 o'clock sharp.** The time so defined is called *Universal Time Coordinate*. The speed of rotation of the Earth also changes irregularly from day to day due to the weather; the average rotation speed even changes from winter to summer because of the changes in the polar ice caps; and in addition that average decreases over time, because of the friction produced by the tides. The rate of insertion of leap seconds is therefore higher than once every 500 days, and not constant in time.

---

* Apart from international units, there are also *provincial* units. Most provincial units still in use are of Roman origin. The mile comes from *milia passum*, which used to be one thousand (double) strides of about 1480 mm each; today a nautical mile, once defined as minute of arc on the Earth's surface, is defined as exactly 1852 m. The inch comes from *uncia/onzia* (a twelfth – now of a foot). The pound (from *pondere* 'to weigh') is used as a translation of *libra* – balance – which is the origin of its abbreviation lb. Even the habit of counting in dozens instead of tens is Roman in origin. These and all other similarly funny units – like the system in which all units start with 'f', and which uses furlong/fortnight as its unit of velocity – are now officially defined as multiples of SI units.

** Their website at hpiers.obspm.fr gives more information on the details of these insertions, as does maia.usno.navy.mil, one of the few useful military websites. See also www.bipm.fr, the site of the BIPM.

* *

Not using SI units can be expensive. In 1999, the space organisation NASA lost a satellite on Mars because some software programmers had used provincial units instead of SI units in part of the code. As a result of using feet instead of meters, the Mars Climate Orbiter crashed into the planet, instead of orbiting it; the loss was around 100 million euro.[*]

* *

The most precisely measured quantities in nature are the frequencies of certain millisecond pulsars,[**] the frequency of certain narrow atomic transitions, and the Rydberg constant of *atomic* hydrogen, which can all be measured as precisely as the second is defined. The caesium transition that defines the second has a finite linewidth that limits the achievable precision: the limit is about 14 digits.

* *

The least precisely measured of the fundamental constants of physics are the gravitational constant G and the strong coupling constant α_s. Even less precisely known are the age of the universe and its density (see Table 14). Page 308

* *

Variations of quantities are often much easier to measure than their values. For example, in gravitational wave detectors, the sensitivity achieved in 1992 was $\Delta l/l = 3 \cdot 10^{-19}$ for lengths of the order of 1 m. Ref. 307 In other words, for a block of about a cubic metre of metal it is possible to measure length changes about 3000 times smaller than a proton radius. These set-ups are now being superseded by ring interferometers. Ring interferometers measuring frequency differences of 10^{-21} have already been built; and they are still being improved. Ref. 308

* *

The table of SI prefixes covers 72 orders of magnitude. How many additional prefixes will be needed? Even an extended list will include only a small part of the infinite range of possibilities. Will the Conférence Générale des Poids et Mesures have to go on forever, defining an infinite number of SI prefixes? Why? Challenge 420 s

* *

The French philosopher Voltaire, after meeting Newton, publicized the now famous story that the connection between the fall of objects and the motion of the Moon was discovered by Newton when he saw an apple falling from a tree. More than a century later, just before the French Revolution, a committee of scientists decided to take as the unit of force precisely the force exerted by gravity on a *standard apple*, and to name it after the English scientist. After extensive study, it was found that the mass of the standard

---

[*] This story revived an old but false urban legend claiming that only three countries in the world do not use SI units: Liberia, the USA and Myanmar.

[**] An overview of this fascinating work is given by J. H. Taylor, *Pulsar timing and relativistic gravity*, Philosophical Transactions of the Royal Society, London A **341**, pp. 117–134, 1992.

apple was 101.9716 g; its weight was called 1 newton. Since then, visitors to the museum in Sèvres near Paris have been able to admire the standard metre, the standard kilogram and the standard apple.*

Precision and accuracy of measurements

Measurements are the basis of physics. Every measurement has an *error*. Errors are due to lack of precision or to lack of accuracy. *Precision* means how well a result is reproduced when the measurement is repeated; *accuracy* is the degree to which a measurement corresponds to the actual value.

Lack of precision is due to accidental or *random errors*; they are best measured by the *standard deviation*, usually abbreviated σ; it is defined through

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 , \qquad (297)$$

Challenge 421 s

where $\bar{x}$ is the average of the measurements x_i. (Can you imagine why $n-1$ is used in the formula instead of n?)

For most experiments, the distribution of measurement values tends towards a normal distribution, also called *Gaussian distribution*, whenever the number of measurements is increased. The distribution, shown in Figure 304, is described by the expression

$$N(x) \approx \mathrm{e}^{-\frac{(x-\bar{x})^2}{2\sigma^2}} . \qquad (298)$$

Challenge 422 e

The square σ^2 of the standard deviation is also called the *variance*. For a Gaussian distribution of measurement values, 2.35σ is the full width at half maximum.

Ref. 310

Lack of accuracy is due to *systematic errors*; usually these can only be estimated. This estimate is often added to the random errors to produce a *total experimental error*, sometimes also called *total uncertainty*. The *relative* error or uncertainty is the ratio between the error and the measured value.

Challenge 423 e

For example, a professional measurement will give a result such as 0.312(6) m. The number between the parentheses is the standard deviation σ, in units of the last digits. As above, a Gaussian distribution for the measurement results is assumed. Therefore, a value of 0.312(6) m implies that the actual value is expected to lie

— within 1σ with 68.3 % probability, thus in this example within 0.312 ± 0.006 m;
— within 2σ with 95.4 % probability, thus in this example within 0.312 ± 0.012 m;
— within 3σ with 99.73 % probability, thus in this example within 0.312 ± 0.018 m;
— within 4σ with 99.9937 % probability, thus in this example within 0.312 ± 0.024 m;
— within 5σ with 99.999 943 % probability, thus in this example within 0.312 ± 0.030 m;
— within 6σ with 99.999 999 80 % probability, thus within 0.312 ± 0.036 m;
— within 7σ with 99.999 999 999 74 % probability, thus within 0.312 ± 0.041 m.

Ref. 309

* To be clear, this is a joke; no standard apple exists. It is *not* a joke however, that owners of several apple trees in Britain and in the US claim descent, by rerooting, from the original tree under which Newton had his insight. DNA tests have even been performed to decide if all these derive from the same tree. The result was, unsurprisingly, that the tree at MIT, in contrast to the British ones, is a fake.

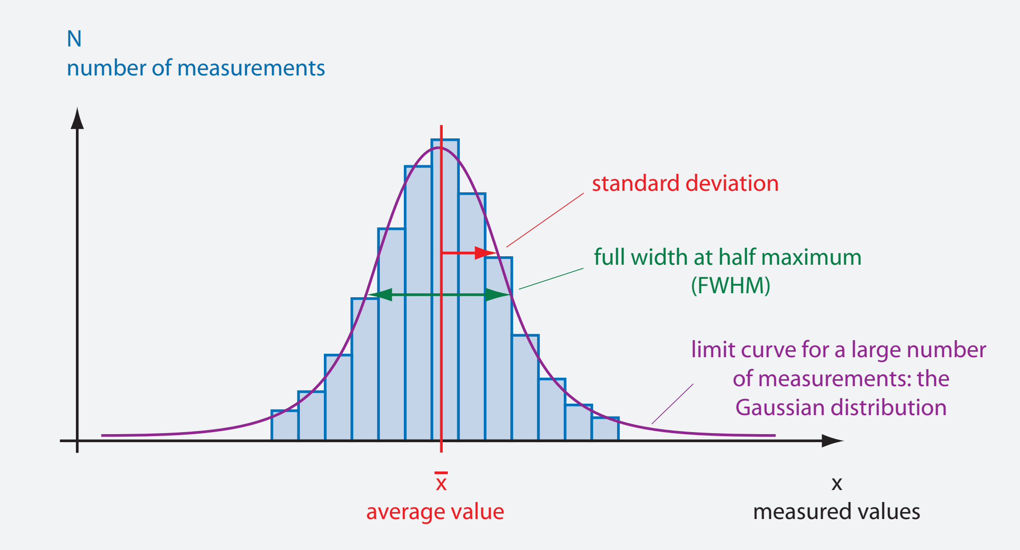


FIGURE 125 A precision experiment and its measurement distribution. The precision is high if the width of the distribution is narrow; the accuracy is high if the centre of the distribution agrees with the actual value.

Challenge 424 s (Do the latter numbers make sense?)

Note that standard deviations have one digit; you must be a world expert to use two, and a fool to use more. If no standard deviation is given, a (1) is assumed. As a result, among professionals, 1 km and 1000 m are *not* the same length!

What happens to the errors when two measured values A and B are added or subtracted? If the all measurements are independent – or uncorrelated – the standard deviation of the sum *and* that of difference is given by $\sigma = \sqrt{\sigma_A^2 + \sigma_B^2}$. For both the product or ratio of two measured and uncorrelated values C and D, the result is $\rho = \sqrt{\rho_C^2 + \rho_D^2}$, where the ρ terms are the *relative* standard deviations.

Challenge 425 s Assume you measure that an object moves 1.0 m in 3.0 s: what is the measured speed value?

Limits to precision

What are the limits to accuracy and precision? There is no way, even in principle, to measure a length x to a *precision* higher than about 61 digits, because in nature, the ratio between the largest and the smallest measurable length is $\Delta x/x > l_{\text{Pl}}/d_{\text{horizon}} = 10^{-61}$. (Is this ratio valid also for force or for volume?) In the final volume of our text, studies of clocks and metre bars strengthen this theoretical limit.

Challenge 426 e

Vol. VI, page 94

But it is not difficult to deduce more stringent practical limits. No imaginable machine can measure quantities with a higher precision than measuring the diameter of the Earth within the smallest length ever measured, about 10^{-19} m; that is about 26 digits of precision. Using a more realistic limit of a 1000 m sized machine implies a limit of 22 digits. If, as predicted above, time measurements really achieve 17 digits of precision, then they are nearing the practical limit, because apart from size, there is an additional practical restriction: cost. Indeed, an additional digit in measurement precision often means an additional digit in equipment cost.

PHYSICAL CONSTANTS

In physics, general observations are deduced from more fundamental ones. As a consequence, many measurements can be deduced from more fundamental ones. The most fundamental measurements are those of the physical constants.

The following tables give the world's best values of the most important physical constants and particle properties – in SI units and in a few other common units – as published in the standard references. The values are the world averages of the best measurements made up to the present. As usual, experimental errors, including both random and estimated systematic errors, are expressed by giving the standard deviation in the last digits. In fact, behind each of the numbers in the following tables there is a long story which is worth telling, but for which there is not enough room here.

Ref. 311

Ref. 312

In principle, *all* quantitative properties of matter can be calculated with quantum theory – more precisely, equations of the standard model of particle – and a set of *basic* physical constants that are given in the next table. For example, the colour, density and elastic properties of any material can be predicted, in principle, in this way.

Vol. V, page 261

TABLE 11 Basic physical constants.

QUANTITY	SYMBOL	VALUE IN SI UNITS	UNCERT.[a]
Constants that define the SI measurement units			
Vacuum speed of light [c]	c	299 792 458 m/s	0
Original Planck constant [c]	h	$6.626\,070\,15 \cdot 10^{-34}$ Js	0
Reduced Planck constant, quantum of action	$\hbar$	$1.054\,571\,817 \ldots \cdot 10^{-34}$ Js	0
Positron charge [c]	e	0.160 217 6634 aC	0
Boltzmann constant [c]	k	$1.380\,649 \cdot 10^{-23}$ J/K	0
Avogadro's number	N_A	$6.022\,140\,76 \cdot 10^{23}$ 1/mol	0
Constant that *should* define the SI measurement units			
Gravitational constant	G	$6.674\,30(15) \cdot 10^{-11}$ Nm²/kg²	$2.2 \cdot 10^{-5}$
Other fundamental constants			
Number of space-time dimensions		3 + 1	0 [b]
Fine-structure constant [d] or	$\alpha = \frac{e^2}{4\pi\varepsilon_0 \hbar c}$	1/137.035 999 084(21)	$1.5 \cdot 10^{-10}$
e.m. coupling constant	$= g_{em}(m_e^2 c^2)$	$= 0.007\,297\,352\,5693(11)$	$1.5 \cdot 10^{-10}$
Fermi coupling constant [d] or	$G_F/(\hbar c)^3$	$1.166\,3787(6) \cdot 10^{-5}$ GeV^{-2}	$5.1 \cdot 10^{-7}$
weak coupling constant	$\alpha_w(M_Z) = g_w^2/4\pi$	1/30.1(3)	$1 \cdot 10^{-2}$
Strong coupling constant [d]	$\alpha_s(M_Z) = g_s^2/4\pi$	0.1179(10)	$8.5 \cdot 10^{-3}$
Weak mixing angle	$\sin^2 \theta_W(\overline{MS})$	0.231 22(4)	$1.7 \cdot 10^{-4}$
	$\sin^2 \theta_W$ (on shell)	0.222 90(30)	$1.3 \cdot 10^{-3}$
	$= 1 - (m_W/m_Z)^2$		

TABLE 11 (Continued) Basic physical constants.

Quantity	Symbol	Value in SI units	Uncert.[a]
CKM quark mixing matrix	$\lvert V \rvert$	$\begin{pmatrix} 0.97383(24) & 0.2272(10) & 0.00396(9) \\ 0.2271(10) & 0.97296(24) & 0.04221(80) \\ 0.00814(64) & 0.04161(78) & 0.999100(34) \end{pmatrix}$	
Jarlskog invariant	J	$3.08(18) \cdot 10^{-5}$	
PMNS neutrino mixing m.	$\lvert P \rvert$	$\begin{pmatrix} 0.82(2) & 0.55(4) & 0.150(7) \\ 0.37(13) & 0.57(11) & 0.71(7) \\ 0.41(13) & 0.59(10) & 0.69(7) \end{pmatrix}$	
Electron mass	m_e	$9.109\,383\,7015(28) \cdot 10^{-31}$ kg	$3.0 \cdot 10^{-10}$
		$5.485\,799\,090\,65(16) \cdot 10^{-4}$ u	$2.9 \cdot 10^{-11}$
		$0.510\,998\,950\,00(15)$ MeV	$3.0 \cdot 10^{-10}$
Muon mass	m_μ	$1.883\,531\,627(42) \cdot 10^{-28}$ kg	$2.2 \cdot 10^{-8}$
		$105.658\,3755(23)$ MeV	$2.2 \cdot 10^{-8}$
Tau mass	m_τ	$1.776\,82(12)\,\mathrm{GeV}/c^2$ $6.8 \cdot 10^{-5}$	
El. neutrino mass	m_{ν_e}	$< 2\,\mathrm{eV}/c^2$	
Muon neutrino mass	m_{ν_μ}	$< 2\,\mathrm{eV}/c^2$	
Tau neutrino mass	m_{ν_τ}	$< 2\,\mathrm{eV}/c^2$	
Up quark mass	u	$21.6(+0.49/-0.26)\,\mathrm{MeV}/c^2$	
Down quark mass	d	$4.67(+0.48/-0.17)\,\mathrm{MeV}/c^2$	
Strange quark mass	s	$93(+11/-5)\,\mathrm{MeV}/c^2$	
Charm quark mass	c	$1.27(2)\,\mathrm{GeV}/c^2$	
Bottom quark mass	b	$4.18(3)\,\mathrm{GeV}/c^2$	
Top quark mass	t	$172.9(0.4)\,\mathrm{GeV}/c^2$	
Photon mass	γ	$< 2 \cdot 10^{-54}$ kg	
W boson mass	$W^\pm$	$80.379(12)\,\mathrm{GeV}/c^2$	
Z boson mass	Z^0	$91.1876(21)\,\mathrm{GeV}/c^2$	
Higgs mass	H	$125.10(14)\,\mathrm{GeV}/c^2$	
Gluon mass	$\mathrm{g}_{1...8}$	*c.* $0\,\mathrm{MeV}/c^2$	

a. Uncertainty: standard deviation of measurement errors.
b. Measured from to 10^{-19} m to 10^{26} m.
c. Defining constant.
d. All coupling constants depend on the 4-momentum transfer, as explained in the section on renormalization. *Fine-structure constant* is the traditional name for the electromagnetic coupling constant g_{em} in the case of a 4-momentum transfer of $Q^2 = m_e^2 c^2$, which is the smallest one possible. At higher momentum transfers it has larger values, e.g., $g_{\mathrm{em}}(Q^2 = M_W^2 c^2) \approx 1/128$. In contrast, the strong coupling constant has lover values at higher momentum transfers; e.g., $\alpha_s(34\,\mathrm{GeV}) = 0.14(2)$.

Page 131

Why do all these basic constants have the values they have? For any basic constant *with a dimension*, such as the quantum of action $\hbar$, the numerical value has only historical meaning. It is $1.054 \cdot 10^{-34}$ Js because of the SI definition of the joule and the second.

The question why the value of a *dimensional* constant is not larger or smaller therefore always requires one to understand the origin of some *dimensionless* number giving the ratio between the constant and the corresponding *natural unit* that is defined with c, G, k, N_A and $\hbar$. Details and values for the natural units are given in the dedicated section.

Vol. IV, page 208

In other words, understanding the sizes of atoms, people, trees and stars, the duration of molecular and atomic processes, or the mass of nuclei and mountains, implies understanding the ratios between these values and the corresponding natural units. The key to understanding nature is thus the understanding of all measurement ratios, and thus of all dimensionless constants. This quest, including the understanding of the fine-structure constant α itself, is completed only in the final volume of our adventure.

The basic constants yield the following useful high-precision observations.

TABLE 12 Derived physical constants.

QUANTITY	SYMBOL	VALUE IN SI UNITS	UNCERT.
Vacuum permeability	μ_0	1.256 637 062 12(19) µH/m	$1.5 \cdot 10^{-10}$
Vacuum permittivity	$\varepsilon_0 = 1/\mu_0 c^2$	8.854 187 8128(13) pF/m	$1.5 \cdot 10^{-10}$
Vacuum impedance	$Z_0 = \sqrt{\mu_0/\varepsilon_0}$	376.730 313 668(57) Ω	$1.5 \cdot 10^{-10}$
Loschmidt's number	N_L	$2.686\,780\,111... \cdot 10^{25}\ 1/\mathrm{m}^3$	0
at 273.15 K and 101 325 Pa			
Faraday's constant	$F = N_A e$	96 485.332 12... C/mol	0
Universal gas constant	$R = N_A k$	8.314 462 618... J/(mol K)	0
Molar volume of an ideal gas	$V = RT/p$	22.413 969 54... l/mol	0
at 273.15 K and 101 325 Pa			
Rydberg constant [a]	$R_\infty = m_e c \alpha^2/2h$	$10\,973\,731.568\,160(21)\ \mathrm{m}^{-1}$	$1.9 \cdot 10^{-12}$
Conductance quantum	$G_0 = 2e^2/h$	77.480 917 29... µS	0
Magnetic flux quantum	$\varphi_0 = h/2e$	2.067 833 848... fWb	0
Josephson frequency ratio	$2e/h$	483.597 8484... THz/V	0
Von Klitzing constant	$h/e^2 = \mu_0 c/2\alpha$	25 812.807 45... Ω	0
Bohr magneton	$\mu_B = e\hbar/2m_e$	9.274 010 0783(28) yJ/T	$3.0 \cdot 10^{-10}$
Classical electron radius	$r_e = e^2/4\pi\varepsilon_0 m_e c^2$	2.817 940 3262(13) fm	$4.5 \cdot 10^{-10}$
Compton wavelength	$\lambda_C = h/m_e c$	2.426 310 238 67(73) pm	$3.0 \cdot 10^{-10}$
of the electron	$ƛ_c = \hbar/m_e c = r_e/\alpha$	0.386 159 267 96(12) pm	$3.0 \cdot 10^{-10}$
Bohr radius [a]	$a_\infty = r_e/\alpha^2$	52.917 721 0903(80) pm	$1.5 \cdot 10^{-10}$
Quantum of circulation	$h/2m_e$	$3.636\,947\,5516(11)\ \mathrm{cm}^2/\mathrm{s}$	$3.0 \cdot 10^{-10}$
Specific positron charge	e/m_e	175.882 001 076(55) GC/kg	$3.0 \cdot 10^{-10}$
Cyclotron frequency	$f_c/B = e/2\pi m_e$	27.992 489 872(9) GHz/T	$3.0 \cdot 10^{-10}$
of the electron			
Electron magnetic moment	μ_e	−9.284 764 7043(28) yJ/T	$3.0 \cdot 10^{-10}$
	μ_e/μ_B	−1.001 159 652 181 28(18)	$1.7 \cdot 10^{-13}$
	μ_e/μ_N	$-1\,838.281\,971\,88(11) \cdot 10^3$	$6.0 \cdot 10^{-11}$
Electron g-factor	g_e	−2.002 319 304 362 56(35)	$1.7 \cdot 10^{-13}$
Muon–electron mass ratio	m_μ/m_e	206.768 2830(46)	$2.2 \cdot 10^{-8}$
Muon magnetic moment	μ_μ	$-4.490\,448\,30(10) \cdot 10^{-26}\ \mathrm{J/T}$	$2.2 \cdot 10^{-8}$

TABLE 12 (Continued) Derived physical constants.

QUANTITY	SYMBOL	VALUE IN SI UNITS	UNCERT.
Muon g-factor	g_μ	−2.002 331 8418(13)	$6.3 \cdot 10^{-10}$
Atomic mass unit	$1\,\mathrm{u} = m_{^{12}\mathrm{C}}/12$	1.660 539 066 60(50) · 10^{-27} kg	$3.0 \cdot 10^{-10}$
Proton mass	m_p	1.672 621 923 69(51) · 10^{-27} kg	$3.1 \cdot 10^{-10}$
		1.007 276 466 621(53) u	$5.3 \cdot 10^{-11}$
		938.272 088 16(29) MeV	$3.1 \cdot 10^{-10}$
Proton–electron mass ratio	$m_\mathrm{p}/m_\mathrm{e}$	1 836.152 673 43(11)	$6.0 \cdot 10^{-11}$
Specific proton charge	e/m_p	9.578 833 1560(29) · 10^{7} C/kg	$3.1 \cdot 10^{-10}$
Proton Compton wavelength	$\lambda_\mathrm{C,p} = h/m_\mathrm{p}c$	1.321 409 855 39(40) fm	$3.1 \cdot 10^{-10}$
Nuclear magneton	$\mu_\mathrm{N} = e\hbar/2m_\mathrm{p}$	5.050 783 7461(15) · 10^{-27} J/T	$3.1 \cdot 10^{-10}$
Proton magnetic moment	μ_p	1.410 606 797 36(60) · 10^{-26} J/T	$4.2 \cdot 10^{-10}$
	$\mu_\mathrm{p}/\mu_\mathrm{B}$	1.521 032 202 30(46) · 10^{-3}	$3.0 \cdot 10^{-10}$
	$\mu_\mathrm{p}/\mu_\mathrm{N}$	2.792 847 344 63(82)	$2.9 \cdot 10^{-10}$
Proton gyromagnetic ratio	$\gamma_\mathrm{p} = 2\mu_p/h$	42.577 478 518(18) MHz/T	$4.2 \cdot 10^{-10}$
Proton g factor	g_p	5.585 694 6893(16)	$2.9 \cdot 10^{-10}$
Neutron mass	m_n	1.674 927 498 04(95) · 10^{-27} kg	$5.7 \cdot 10^{-10}$
		1.008 664 915 95(43) u	$4.8 \cdot 10^{-10}$
		939.565 420 52(54) MeV	$5.7 \cdot 10^{-10}$
Neutron–electron mass ratio	$m_\mathrm{n}/m_\mathrm{e}$	1 838.683 661 73(89)	$4.8 \cdot 10^{-10}$
Neutron–proton mass ratio	$m_\mathrm{n}/m_\mathrm{p}$	1.001 378 419 31(49)	$4.9 \cdot 10^{-10}$
Neutron Compton wavelength	$\lambda_\mathrm{C,n} = h/m_\mathrm{n}c$	1.319 590 905 81(75) fm	$5.7 \cdot 10^{-10}$
Neutron magnetic moment	μ_n	−0.966 236 51(23) · 10^{-26} J/T	$2.4 \cdot 10^{-7}$
	$\mu_\mathrm{n}/\mu_\mathrm{B}$	−1.041 875 63(25) · 10^{-3}	$2.4 \cdot 10^{-7}$
	$\mu_\mathrm{n}/\mu_\mathrm{N}$	−1.913 042 73(45)	$2.4 \cdot 10^{-7}$
Stefan–Boltzmann constant	$\sigma = \pi^2 k^4/60\hbar^3 c^2$	56.703 744 19... nW/m²K⁴	0
Wien's displacement constant	$b = \lambda_\mathrm{max} T$	2.897 771 955... mmK	0
		58.789 257 57... GHz/K	0
Electron volt	eV	0.160 217 6634... aJ	0
Bits to entropy conversion const.	$k \ln 2$	10^{23} bit = 0.956 994... J/K	0
TNT energy content		3.7 to 4.0 MJ/kg	$4 \cdot 10^{-2}$

a. For infinite mass of the nucleus.

Some useful properties of our local environment are given in the following table.

TABLE 13 Astronomical constants.

QUANTITY	SYMBOL	VALUE
Tropical year 1900 [a]	a	31 556 925.974 7 s
Tropical year 1994	a	31 556 925.2 s
Mean sidereal day	d	$23^h 56' 4.090\,53''$

TABLE 13 (Continued) Astronomical constants.

Quantity	Symbol	Value
Average distance Earth–Sun[b]		149 597 870.691(30) km
Astronomical unit[b]	AU	149 597 870 691 m
Light year, based on Julian year[b]	al	9.460 730 472 5808 Pm
Parsec	pc	30.856 775 806 Pm = 3.261 634 al
Earth's mass	$M_{♁}$	$5.973(1) \cdot 10^{24}$ kg
Geocentric gravitational constant	GM	$3.986\,004\,418(8) \cdot 10^{14}\ \mathrm{m^3/s^2}$
Earth's gravitational length	$l_{♁} = 2GM/c^2$	8.870 056 078(16) mm
Earth's equatorial radius[c]	$R_{♁\mathrm{eq}}$	6378.1366(1) km
Earth's polar radius[c]	$R_{♁\mathrm{p}}$	6356.752(1) km
Equator–pole distance[c]		10 001.966 km (average)
Earth's flattening[c]	$e_{♁}$	1/298.25642(1)
Earth's av. density	$\rho_{♁}$	$5.5\ \mathrm{Mg/m^3}$
Earth's age	$T_{♁}$	4.50(4) Ga = 142(2) Ps
Earth's normal gravity	g	$9.806\,65\ \mathrm{m/s^2}$
Earth's standard atmospher. pressure	p_0	101 325 Pa
Moon's radius	$R_{☾\,\mathrm{v}}$	1738 km in direction of Earth
Moon's radius	$R_{☾\,\mathrm{h}}$	1737.4 km in other two directions
Moon's mass	$M_{☾}$	$7.35 \cdot 10^{22}$ kg
Moon's mean distance[d]	$d_{☾}$	384 401 km
Moon's distance at perigee[d]		typically 363 Mm, historical minimum 359 861 km
Moon's distance at apogee[d]		typically 404 Mm, historical maximum 406 720 km
Moon's angular size[e]		average 0.5181° = 31.08′, minimum 0.49°, maximum 0.55°
Moon's average density	$\rho_{☾}$	$3.3\ \mathrm{Mg/m^3}$
Moon's surface gravity	$g_{☾}$	$1.62\ \mathrm{m/s^2}$
Moon's atmospheric pressure	$p_{☾}$	from 10^{-10} Pa (night) to 10^{-7} Pa (day)
Jupiter's mass	$M_{♃}$	$1.90 \cdot 10^{27}$ kg
Jupiter's radius, equatorial	$R_{♃}$	71.398 Mm
Jupiter's radius, polar	$R_{♃}$	67.1(1) Mm
Jupiter's average distance from Sun	$D_{♃}$	778 412 020 km
Jupiter's surface gravity	$g_{♃}$	$24.9\ \mathrm{m/s^2}$
Jupiter's atmospheric pressure	$p_{♃}$	from 20 kPa to 200 kPa
Sun's mass	$M_{\odot}$	$1.988\,43(3) \cdot 10^{30}$ kg
Sun's gravitational length	$2GM_{\odot}/c^2$	2.953 250 08(5) km
Heliocentric gravitational constant	$GM_{\odot}$	$132.712\,440\,018(8) \cdot 10^{18}\ \mathrm{m^3/s^2}$
Sun's luminosity	$L_{\odot}$	384.6 YW
Solar equatorial radius	$R_{\odot}$	695.98(7) Mm

TABLE 13 (Continued) Astronomical constants.

QUANTITY	SYMBOL	VALUE
Sun's angular size		0.53° average; minimum on fourth of July (aphelion) 1888 ″, maximum on fourth of January (perihelion) 1952 ″
Sun's average density	$\rho_\odot$	1.4 Mg/m^3
Sun's average distance	AU	149 597 870.691(30) km
Sun's age	$T_\odot$	4.6 Ga
Solar velocity around centre of galaxy	$v_{\odot g}$	220(20) km/s
Solar velocity against cosmic background	$v_{\odot b}$	370.6(5) km/s
Sun's surface gravity	$g_\odot$	274 m/s^2
Sun's lower photospheric pressure	$p_\odot$	15 kPa
Distance to Milky Way's centre		8.0(5) kpc = 26.1(1.6) kal
Milky Way's age		13.6 Ga
Milky Way's size		*c.* 10^{21} m or 100 kal
Milky Way's mass		10^{12} solar masses, *c.* $2 \cdot 10^{42}$ kg
Most distant galaxy cluster known	SXDF-XCLJ 0218-0510	$9.6 \cdot 10^9$ al

a. Defining constant, from vernal equinox to vernal equinox; it was once used to define the second. (Remember: π seconds is about a nanocentury.) The value for 1990 is about 0.7 s less, corresponding to a slowdown of roughly 0.2 ms/a. (Watch out: why?) There is even an empirical formula for the change of the length of the year over time. Challenge 427 s Ref. 313

b. The truly amazing precision in the average distance Earth–Sun of only 30 m results from time averages of signals sent from Viking orbiters and Mars landers taken over a period of over twenty years. Note that the International Astronomical Union distinguishes the average distance Earth–Sun from the *astronomical unit* itself; the latter is defined as a fixed and exact length. Also the *light year* is a unit defined as an exact number by the IAU. For more details, see www.iau.org/public/measuring.

c. The shape of the Earth is described most precisely with the World Geodetic System. The last edition dates from 1984. For an extensive presentation of its background and its details, see the www.wgs84.com website. The International Geodesic Union refined the data in 2000. The radii and the flattening given here are those for the 'mean tide system'. They differ from those of the 'zero tide system' and other systems by about 0.7 m. The details constitute a science in itself.

d. Measured centre to centre. To find the precise position of the Moon at a given date, see the www.fourmilab.ch/earthview/moon_ap_per.html page. For the planets, see the page www.fourmilab.ch/solar/solar.html and the other pages on the same site.

e. Angles are defined as follows: 1 degree = 1° = π/180 rad, 1 (first) minute = 1 ′ = 1°/60, 1 second (minute) = 1 ″ = 1 ′/60. The ancient units 'third minute' and 'fourth minute', each 1/60th of the preceding, are not in use any more. ('Minute' originally means 'very small', as it still does in modern English.)

Some properties of nature at large are listed in the following table. (If you want a chal-

Challenge 428 s lenge, can you determine whether any property of the universe itself is listed?)

TABLE 14 Cosmological constants.

Quantity	Symbol	Value
Cosmological constant	Λ	*c.* $1 \cdot 10^{-52}\,\mathrm{m}^{-2}$
Age of the universe[a]	t_0	$4.333(53) \cdot 10^{17}\,\mathrm{s} = 13.8(0.1) \cdot 10^{9}\,\mathrm{a}$
(determined from space-time, via expansion, using general relativity)		
Age of the universe[a]	t_0	over $3.5(4) \cdot 10^{17}\,\mathrm{s} = 11.5(1.5) \cdot 10^{9}\,\mathrm{a}$
(determined from matter, via galaxies and stars, using quantum theory)		
Hubble parameter[a]	H_0	$2.3(2) \cdot 10^{-18}\,\mathrm{s}^{-1} = 0.73(4) \cdot 10^{-10}\,\mathrm{a}^{-1}$
	$= h_0 \cdot 100\,\mathrm{km/s\,Mpc} = h_0 \cdot 1.0227 \cdot 10^{-10}\,\mathrm{a}^{-1}$	
Reduced Hubble parameter[a]	h_0	$0.71(4)$
Deceleration parameter[a]	$q_0 = -(\ddot{a}/a)_0/H_0^2$	$-0.66(10)$
Universe's horizon distance[a]	$d_0 = 3ct_0$	$40.0(6) \cdot 10^{26}\,\mathrm{m} = 13.0(2)\,\mathrm{Gpc}$
Universe's topology		trivial up to $10^{26}\,\mathrm{m}$
Number of space dimensions		3, for distances up to $10^{26}\,\mathrm{m}$
Critical density of the universe	$\rho_c = 3H_0^2/8\pi G$	$h_0^2 \cdot 1.878\,82(24) \cdot 10^{-26}\,\mathrm{kg/m^3}$ $= 0.95(12) \cdot 10^{-26}\,\mathrm{kg/m^3}$
(Total) density parameter[a]	$\Omega_0 = \rho_0/\rho_c$	$1.02(2)$
Baryon density parameter[a]	$\Omega_{B0} = \rho_{B0}/\rho_c$	$0.044(4)$
Cold dark matter density parameter[a]	$\Omega_{CDM0} = \rho_{CDM0}/\rho_c$	$0.23(4)$
Neutrino density parameter[a]	$\Omega_{\nu 0} = \rho_{\nu 0}/\rho_c$	0.001 to 0.05
Dark energy density parameter[a]	$\Omega_{X0} = \rho_{X0}/\rho_c$	$0.73(4)$
Dark energy state parameter	$w = p_X/\rho_X$	$-1.0(2)$
Baryon mass	m_b	$1.67 \cdot 10^{-27}\,\mathrm{kg}$
Baryon number density		$0.25(1)\,/\mathrm{m^3}$
Luminous matter density		$3.8(2) \cdot 10^{-28}\,\mathrm{kg/m^3}$
Stars in the universe	n_s	$10^{22\pm1}$
Baryons in the universe	n_b	$10^{81\pm1}$
Microwave background temperature[b]	T_0	$2.725(1)\,\mathrm{K}$
Photons in the universe	n_γ	10^{89}
Photon energy density	$\rho_\gamma = \pi^2 k^4/15T_0^4$	$4.6 \cdot 10^{-31}\,\mathrm{kg/m^3}$
Photon number density		$410.89\,/\mathrm{cm^3}$ or $400\,/\mathrm{cm^3}(T_0/2.7\,\mathrm{K})^3$
Density perturbation amplitude	$\sqrt{S}$	$5.6(1.5) \cdot 10^{-6}$
Gravity wave amplitude	$\sqrt{T}$	$< 0.71\sqrt{S}$
Mass fluctuations on 8 Mpc	σ_8	$0.84(4)$
Scalar index	n	$0.93(3)$
Running of scalar index	$\mathrm{d}n/\mathrm{d}\ln k$	$-0.03(2)$
Planck length	$l_{Pl} = \sqrt{\hbar G/c^3}$	$1.62 \cdot 10^{-35}\,\mathrm{m}$
Planck time	$t_{Pl} = \sqrt{\hbar G/c^5}$	$5.39 \cdot 10^{-44}\,\mathrm{s}$
Planck mass	$m_{Pl} = \sqrt{\hbar c/G}$	$21.8\,\mathrm{\mu g}$

TABLE 14 (Continued) Cosmological constants.

Quantity	Symbol	Value
Instants in history[a]	t_0/t_{Pl}	$8.7(2.8) \cdot 10^{60}$
Space-time points inside the horizon[a]	$N_0 = (R_0/l_{\mathrm{Pl}})^3 \cdot (t_0/t_{\mathrm{Pl}})$	$10^{244\pm1}$
Mass inside horizon	M	$10^{54\pm1}$ kg

a. The index 0 indicates present-day values.

b. The radiation originated when the universe was 380 000 years old and had a temperature of about 3000 K; the fluctuations ΔT_0 which led to galaxy formation are today about 16 ± 4 μK = $6(2) \cdot 10^{-6}\, T_0$. Page 231

USEFUL NUMBERS

Ref. 314

π	$3.14159\,26535\,89793\,23846\,26433\,83279\,50288\,41971\,69399\,37510_5$
e	$2.71828\,18284\,59045\,23536\,02874\,71352\,66249\,77572\,47093\,69995_9$
γ	$0.57721\,56649\,01532\,86060\,65120\,90082\,40243\,10421\,59335\,93992_3$
ln 2	$0.69314\,71805\,59945\,30941\,72321\,21458\,17656\,80755\,00134\,36025_5$
ln 10	$2.30258\,50929\,94045\,68401\,79914\,54684\,36420\,76011\,01488\,62877_2$
$\sqrt{10}$	$3.16227\,76601\,68379\,33199\,88935\,44432\,71853\,37195\,55139\,32521_6$

CHALLENGE HINTS AND SOLUTIONS

Challenge 1, page 10: Do not hesitate to be demanding and strict. The next edition of the text will benefit from it.

Challenge 2, page 15: A cone or a circular hyperboloid also looks straight from 'all' directions, provided the positioning of the eye is suitably chosen. Therefore, to check planarity, we need not only to turn the object, but also to displace it. The best method to check planarity is to use interference between an arriving and a departing beam of coherent light* with a diameter that covers the whole object. If the interference fringes in such an *interferogram* are straight, the surface is planar.

Vol. III, page 180

Challenge 3, page 16: A finite fraction of infinity is still infinite. Infinity cannot be used as a unit.

Challenge 4, page 17: The time at which the Moon Io enters the shadow in the second measurement occurs about 1000 s later than predicted from the first measurement. Since the Earth is about $3 \cdot 10^{11}$ m further away from Jupiter and Io, we get the usual value for the speed of light.

Challenge 5, page 18: The rain and wind diagrams in Figure 4 suggest to use the *tangent* in equation (2). This is the Galilean expression; however, for light, it would imply velocities above c, and thus cannot be correct. The light diagrams suggest to use the *sine*. This is the correct relativistic expression for the special case of a star precisely above the ecliptic. More general relativistic expressions, for stars of general declination, are easily derived.

Challenge 6, page 18: To compensate for the aberration, the telescope has to be inclined *along* the direction of motion of the Earth; to compensate for parallax, *perpendicularly* to the motion.

Challenge 7, page 19: The drawing shows it. Observer, Moon and Sun form a triangle. When the Moon is half full, the angle at the Moon is a right angle. Thus the distance ratio can be determined, though not easily, as the angle at the observer is very close to a right angle as well.

Challenge 8, page 19: There are Cat's-eyes on the Moon deposited there during the Apollo and Lunokhod missions. They are used to reflect laser 35 ps light pulses sent there through telescopes. The timing of the round trip then gives the distance to the Moon. Of course, absolute distance is not know to high precision, but the variations are. The thickness of the atmosphere is the largest source of error. See the www.csr.utexas.edu/mlrs and ilrs.gsfc.nasa.gov websites.

Page 168

Challenge 9, page 19: Fizeau used a mirror about 8.6 km away. As the picture shows, he only had to count the teeth of his cog-wheel and measure its rotation speed when the light goes in one direction through one tooth and comes back to the next.

Challenge 10, page 20: The shutter time must be shorter than $T = l/c$, in other words, shorter than 30 ps; it was a *gas* shutter, not a solid one. It was triggered by a red light pulse (shown in the photograph) timed by the pulse to be photographed; for certain materials, such as the used gas, strong light can lead to bleaching, so that they become transparent. For more details about the

---

* Generally speaking, two light beams – or any other two waves – are called *coherent* if they have constant phase difference and frequency. Coherence enables and is required for interference.

shutter and its neat trigger technique, see the paper by the authors. For even faster shutters, see also the discussion in volume VI, on page 120.

Challenge 11, page 21: Just take a photograph of a lightning while moving the camera horizontally. You will see that a lightning is made of several discharges; the whole shows that lightning is much slower than light.

If lightning moved only nearly as fast as light itself, the Doppler effect would change its colour depending on the angle at which we look at it, compared to its direction of motion. A nearby lightning would change colour from top to bottom.

Challenge 12, page 23: The fastest lamps were subatomic particles, such as muons, which decay by emitting a photon, thus a tiny flash of light. However, also some stars emit fasts jets of matter, which move with speeds comparable to that of light.

Challenge 13, page 24: The speed of neutrinos is the same as that of light to 9 decimal digits, since neutrinos and light were observed to arrive together, within 12 seconds of each other, after a trip of 170 000 light years from a supernova explosion. Ref. 44

Challenge 14, page 25: Even the direction of the arriving light pulse is hard to measure before it arrives. But maybe one could play on the surface of a black hole? Or under water? Or using a mirror as tennis court? Enjoy the exploration.

Challenge 16, page 29: This is best discussed by showing that other possibilities make no sense.

Challenge 17, page 29: The spatial coordinate of the event at which the light is reflected is $c(k^2-1)T/2$; the time coordinate is $(k^2+1)T/2$. Their ratio must be v. Solving for k gives the result.

Challenge 19, page 31: The motion of radio waves, infrared, ultraviolet and gamma rays is also unstoppable. Another past suspect, the neutrino, has been found to have mass and to be thus in principle stoppable. The motion of gravity is also unstoppable.

Challenge 21, page 35: $\lambda_R/\lambda_S = \gamma$.

Challenge 22, page 35: To change from bright red (650 nm) to green (550 nm), $v = 0.166c$ is necessary.

Challenge 23, page 35: People measure the shift of spectral lines, such as the shift of the so-called Lyman-α line of hydrogen, that is emitted (or absorbed) when a free electron is captured (or ejected) by a proton. It is one of the famous Fraunhofer lines. Vol. IV, page 180

Challenge 24, page 35: The speeds are given by

$$v/c = \frac{(z+1)^2-1}{(z+1)^2+1} \qquad (299)$$

which implies $v(z=-0.1) = 31\,\mathrm{Mm/s} = 0.1c$ *towards* the observer and $v(z=5) = 284\,\mathrm{Mm/s} = 0.95c$ *away* from the observer.

A red-shift of 6 implies a speed of $0.96c$; such speeds appear because, as we will see in the section of general relativity, far away objects recede from us. And high red-shifts are observed only for objects which are extremely far from Earth, and the faster the further they are away. For a red-shift of 6 that is a distance of several thousand million light years.

Challenge 25, page 36: No Doppler effect is seen for a distant observer at rest with respect to the large mass. In other cases there obviously is a Doppler effect, but it is not due to the deflection.

Challenge 26, page 36: Sound speed is not invariant of the speed of observers. As a result, the Doppler effect for sound even confirms – within measurement differences – that time is the *same* for observers moving against each other.

Challenge 29, page 38: Inside colour television tubes (they used higher voltages, typically 30 kV, than black and white ones did), electrons are described by $v/c \approx \sqrt{2\cdot 30/511}$ or $v \approx 0.3c$.

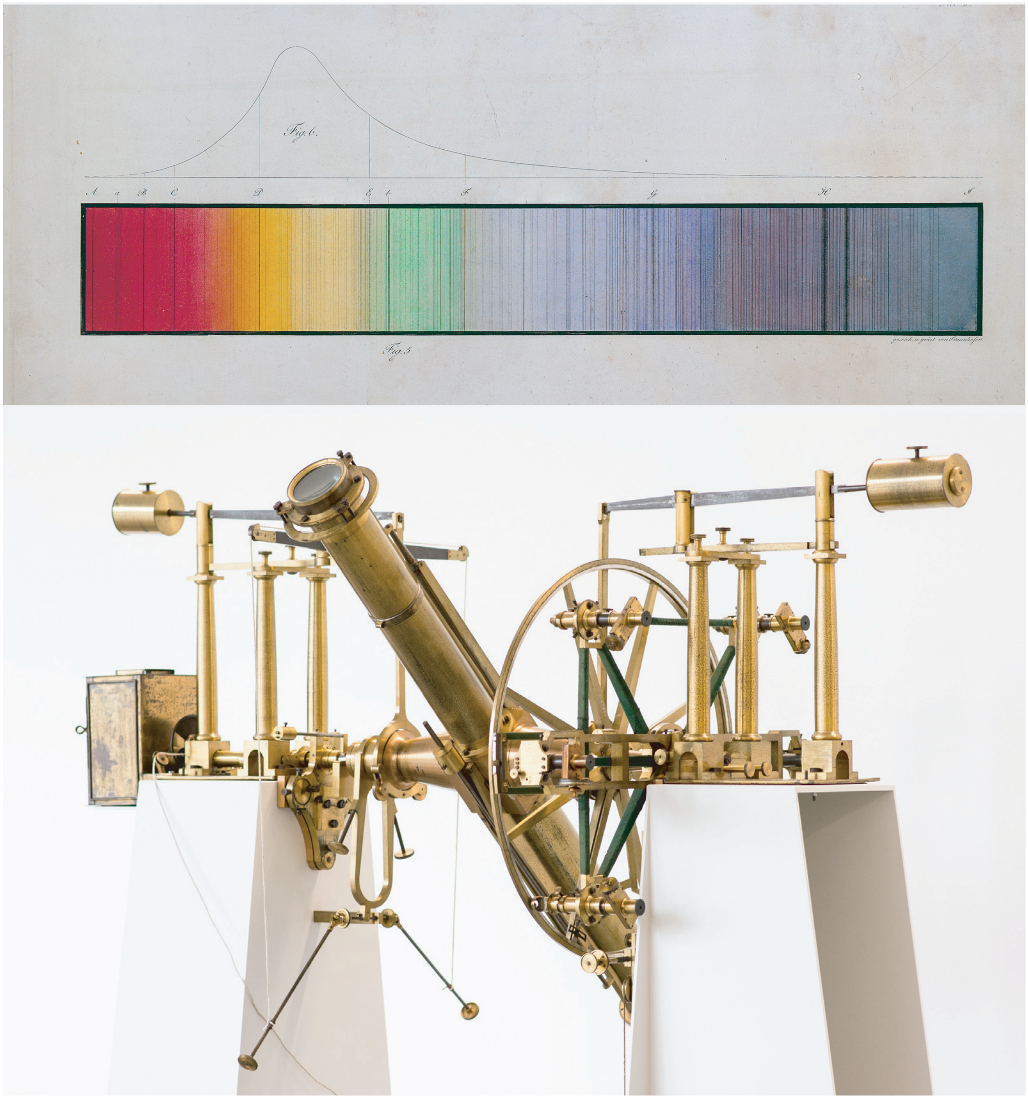


FIGURE 126 The original lines published by Fraunhofer and the meridian instrument that he used (© Fraunhofer Gesellschaft).

Challenge 30, page 38: If you can imagine this, publish it. Readers will be delighted to hear the story.

Challenge 32, page 39: The connection between observer invariance and limit property seems to be generally valid in nature, as shown in chapter 2. However, a complete and airtight argument is not yet at hand. If you have one, publish it! Vol. VI, page 27

Challenge 35, page 42: If the speed of light is the same for all observers, no observer can pretend to be more at rest than another (as long as space-time is flat), because there is no observation from electrodynamics, mechanics or another part of physics that allows such a statement.

Challenge 39, page 43: The human value is achieved in particle accelerators; the value in nature is found in cosmic rays of the highest energies.

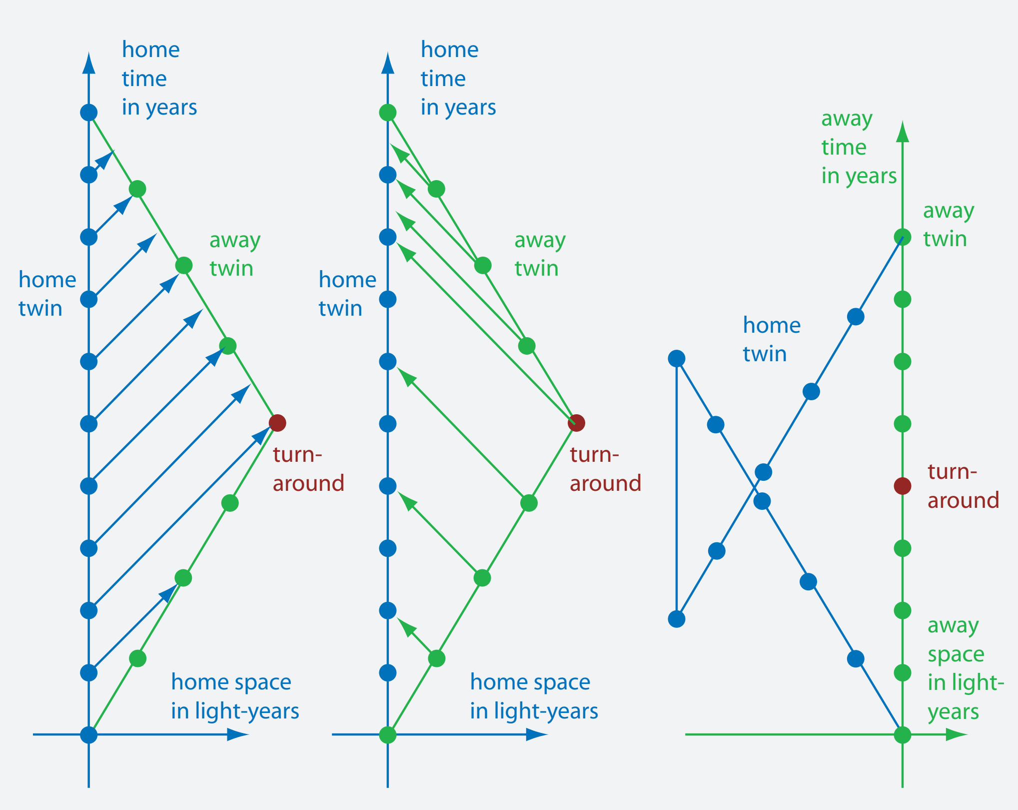


FIGURE 127 The twin paradox: (left and centre) the clock timing for both twins with the signals sent among the twins in the inertial frame of the home twin, and (right) the description by the away twin, in a frame that, however, is not inertial.

Challenge 41, page 44: Redrawing Figure 11 on page 29 for the other observer makes the point.

Challenge 42, page 45: The set of events behaves like a manifold, because it behaves like a four-dimensional space: it has infinitely many points around any given starting point, and distances behave as we are used to, limits behave as we are used to. It differs by one added dimension, and by the sign in the definition of distance; thus, properly speaking, it is a Riemannian manifold.

Challenge 43, page 46: Infinity is obvious, as is openness. Thus the topology equivalence can be shown by imagining that the manifold is made of rubber and wrapped around a sphere.

Challenge 44, page 46: The light cone remains unchanged; thus causal connection as well.

Challenge 47, page 47: In such a case, the division of space-time around an inertial observer into future, past and elsewhere would not hold any more, and the future could influence the past (as seen from another observer).

Challenge 53, page 50: To understand the twin paradox, the best way is to draw a space-time diagram showing how each twin sends a time signal at regular intervals, as seen on his own clock, to his brother. Some examples are given in Figure 127. These time signals show how much he has aged. You will see directly that, during the trip, one twin sends fewer signals than the other.

Challenge 54, page 51: The ratio predicted by naive reasoning is $(1/2)^{(6.4/2.2)} = 0.13$.

Challenge 55, page 51: The time dilation factor for $v = 0.9952c$ is 10.2, giving a proper time of 0.62 μs; thus the ratio predicted by special relativity is $(1/2)^{(0.62/2.2)} = 0.82$.

Challenge 57, page 51: Send a light signal from the first clock to the second clock and back. Take the middle time between the departure and arrival, and then compare it with the time at the reflection. Repeat this a few times. See also Figure 11.

Challenge 59, page 52: Not with present experimental methods.

Challenge 60, page 52: Hint: think about different directions of sight.

Challenge 62, page 53: Hint: be careful with the definition of 'rigidity'.

Challenge 64, page 53: While the departing glider passes the gap, the light cannot stay on at any speed, if the glider is shorter than the gap. This is strange at first sight, because the glider does not light the lamp even at high speeds, even though in the frame of the glider there is contact at both ends. The reason is that in this case there is not enough time to send the signal to the battery that contact is made, so that the current cannot start flowing.

Assume that current flows with speed u, which is of the order of c. Then, as Dirk Van de Moortel showed, the lamp will go off if the glider length l and the gap length d obey $l/d < \gamma(u + v)/u$. See also the cited reference.

For a glider approaching the gap and the lamp, the situation is different: a glider shorter than the gap *can* keep the lamp on all the time, as pointed out by Madhu Rao.

Why are the debates often heated? Some people will (falsely) pretend that the problem is unphysical; other will say that Maxwell's equations are needed. Still others will say that the problem is absurd, because for larger lengths of the glider, the on/off answer depends on the precise speed value. However, this actually is the case in this situation.

Challenge 65, page 54: Yes, the rope breaks; in accelerated cars, distance changes, as shown later on in the text.

Challenge 66, page 54: The submarine will sink. The fast submarine will even be heavier, as his kinetic energy adds to his weight. The contraction effect would make it lighter, as the captain says, but by a smaller amount. The total weight – counting upwards as positive – is given by $F = -mg(\gamma - 1/\gamma)$.

Challenge 67, page 54: A relativistic submarine would instantly melt due to friction with the water. If not, it would fly off the planet because it would move faster than the escape velocity – and would produce several other disasters.

Challenge 68, page 55: A relativistic pearl necklace at constant speed does get shorter, but as usual, the shortening can only be measured, not photographed. At relativistic speeds, the measured sizes of the pearls are flattened ellipsoids. Spheres do not get transformed into spheres. The observed necklace consists of overlapping elliposids.

Challenge 69, page 55: No: think about it!

Challenge 72, page 58: Yes, ageing in a valley is slowed compared to mountain tops. However, the proper sensation of time is not changed. The reason for the appearance of grey hair is not known; if the timing is genetic, the proper time at which it happens is the same in either location.

Challenge 73, page 58: There is no way to put an observer at the specified points. Proper velocity can only be defined for observers, i.e., for entities which can carry a clock. That is not the case for images.

Challenge 74, page 59: Just use plain geometry to show this.

Challenge 75, page 60: Most interestingly, the horizon can easily move faster than light, if you move your head appropriately, as can the end of the rainbow.

Challenge 77, page 63: Light is necessary to determine distance *and* to synchronize clocks; thus there is no way to measure the speed of light from one point to another alone. The reverse motion needs to be included. However, some statements on the one-way speed of light can still be made (see math.ucr.edu/home/baez/physics/Relativity/SR/experiments.html). All experiments on the one-way speed of light performed so far are consistent with an isotropic value that is equal to the two-way velocity. However, no experiment is able to rule out a group of theories in which the one-way speed of light is anisotropic and thus different from the two-way speed. All theories from this group have the property that the *round-trip* speed of light is isotropic in any inertial frame, but the *one-way* speed is isotropic only in a preferred 'ether' frame. In all of these theories, in all inertial frames, the effects of slow clock transport exactly compensate the effects of the anisotropic one-way speed of light. All these theories are experimentally indistinguishable from special relativity. In practice, therefore, the one-way speed of light has been measured and is constant. But a small option remains.

The subtleties of the one-way and two-way speed of light have been a point of discussion for a long time. It has been often argued that a factor different than two, which would lead to a distinction between the one-way speed of light and the two-way speed of light, cannot be ruled out by experiment, as long as the two-way speed of light remains c for all observers.

Ref. 18 Many experiments on the one-way velocity of light are explained and discussed by Zhang. He says in his summary on page 171, that the one-way velocity of light is indeed independent of the light source; however, no experiment really shows that it is equal to the two-way velocity. Ref. 78 Moreover, almost all so-called 'one-way' experiments are in fact still hidden 'two-way' experiments (see his page 150).

Ref. 79 In 2004, Hans Ohanian showed that the question can be settled by discussing how a non-standard one-way speed of light would affect dynamics. He showed that a non-standard one-way speed of light would introduce pseudoaccelerations and pseudoforces (similar to the Coriolis acceleration and force); since these pseudoaccelerations and pseudoforces are not observed, the one-way speed of light is the same as the two-way speed of light.

In short, the issues of the one-way speed of light do not need to worry us here.

Challenge 78, page 65: The expression does not work for a photon hitting a mirror, for example.

Challenge 79, page 65: Teleportation contradicts, in an inertial reference frame, the conservation of the centre of mass. Quick teleportation would lead to strong acceleration of the sending and the receiving environment.

Challenge 85, page 68: The lower collision in Figure 41 shows the result directly, from energy conservation. For the upper collision the result also follows, if one starts from momentum conservation $\gamma m v = \Gamma M V$ and energy conservation $(gamma + 1)m = \Gamma M$.

Challenge 97, page 74: Just turn the left side of Figure 45 a bit in anti-clockwise direction.

Challenge 98, page 75: In collisions between relativistic charges, part of the energy is radiated away as light, so that the particles effectively lose energy.

Challenge 99, page 76: Probably not, as all relations among physical quantities are known now. However, you might check for yourself; one might never know. It is worth to mention that the maximum force in nature was discovered (in this text) after remaining hidden for over 80 years.

Challenge 101, page 79: Write down the four-vectors U' and U and then extract v' as function of v and the relative coordinate speed V. Then rename the variables.

Challenge 102, page 79: *No* example of motion of a massive body has! Only the motion of light waves has null phase 4-velocity and null group 4-velocity, as explained on page 86.

Challenge 107, page 82: For ultrarelativistic particles, like for massless particles, one has $E = pc$.

Challenge 108, page 82: Hint: evaluate P_1 and P_2 in the rest frame of one particle.

Challenge 110, page 83: Use the definition $\boldsymbol{F} = \mathrm{d}\boldsymbol{p}/\mathrm{d}t$ and the relation $\boldsymbol{K}\boldsymbol{U} = 0 = \boldsymbol{F}\boldsymbol{v} - \mathrm{d}E/\mathrm{d}t$ valid for rest-mass preserving forces.

Challenge 112, page 84: The story is told on page 108.

Challenge 116, page 85: This problem is called the *Ehrenfest paradox*. There are many publications about it. Enjoy the exploration!

Challenge 117, page 85: Yes, one can see such an object: the searchlight effect and the Doppler effect do not lead to invisibility. However, part of the object, namely the region rotating away from the observer, may become very dark.

Challenge 119, page 86: If the rotating particle has a magnetic moment, one can send it through an inhomogeneous magnetic field and observe whether the magnetic moment changes direction.

Challenge 121, page 86: No.

Challenge 122, page 86: For a discussion of relativistic angular momentum and a pretty effect, see K. Y. Bliokh & F. Nori, *Relativistic Hall Effect*, Physical Review Letters **108**, p. 120403, 2012, preprint at arxiv.org/abs/1112.5618.

Challenge 125, page 86: The relation for the frequency follows from the definition of the phase.

Challenge 128, page 88: Planck invited Einstein to Berlin and checked his answers with him. The expression $E = \hbar\omega$ for the photon energy implies the invariance of $\hbar$.

Challenge 144, page 96: The energy contained in the fuel must be comparable to the rest mass of the motorbike, multiplied by c^2. Since fuel contains much more mass than energy, that gives a big problem.

Challenge 146, page 96: Constant acceleration and gravity are similar in their effects, as discussed in the section on general relativity.

Challenge 149, page 98: Yes, it is true.

Challenge 150, page 98: It is flat, like a plane.

Challenge 151, page 98: Despite the acceleration towards the centre of the carousel, no horizon appears.

Challenge 153, page 99: Yes; however, the effect is minimal and depends on the position of the Sun. In fact, what is white at one height is not white at another.

Challenge 155, page 100: Locally, light always moves with speed c.

Challenge 156, page 100: Away from Earth, g decreases; it is effectively zero over most of the distance.

Challenge 159, page 102: As shown in the cited reference, the limit follows from the condition $l\gamma^3 a \leqslant c^2$.

Challenge 161, page 102: Yes.

Challenge 162, page 102: Yes. Take $\Delta f \ \Delta t \geqslant 1$ and substitute $\Delta l = c/\Delta f$ and $\Delta a = c/\Delta t$.

Challenge 164, page 105: Though there are many publications pretending to study the issue, there are also enough physicists who notice the impossibility. Measuring a variation of the speed of light is not much far from measuring the one way speed of light: it is not possible. However, the debates on the topic are heated; the issue will take long to be put to rest.

Challenge 166, page 107: The inverse square law of gravity does not comply with the maximum speed principle; it is not clear how it changes when one changes to a moving observer.

Challenge 167, page 112: If you hear about a claim to surpass the force or power limit, let me know.

Challenge 168, page 112: Take a surface moving with the speed of light, or a surface defined with a precision smaller than the Planck length.

Challenge 169, page 117: Also shadows do not remain parallel on curved surfaces. Forgetting this leads to strange mistakes: many arguments allegedly ‘showing’ that men have never been on the moon neglect this fact when they discuss the photographs taken there.

Challenge 170, page 120: If you find one, publish it and then send it to me.

Challenge 172, page 125: This is tricky. Simple application of the relativistic transformation rule for 4-vectors can result in force values above the limit. But in every such case, a horizon has appeared that prevents the observation of this higher value.

Challenge 173, page 125: If so, publish it; then send it to me.

Challenge 174, page 127: For example, it is possible to imagine a surface that has such an intricate shape that it will pass all atoms of the universe at almost the speed of light. Such a surface is not physical, as it is impossible to imagine observers on all its points that move in that way all at the same time.

Challenge 176, page 127: Publish it – and then send it to me.

Challenge 177, page 128: New sources cannot appear from nowhere. Any 'new' power source results from the transformation of other radiation found in the universe already before the appearance.

Challenge 178, page 128: Many do not believe the limits yet; so any proposed counter-example or any additional paradox is worth a publication.

Challenge 181, page 133: If so, publish it; then send it to me.

Challenge 185, page 135: If so, publish it; then send it to me.

Challenge 187, page 137: They are accelerated upwards.

Challenge 188, page 137: In everyday life, (a) the surface of the Earth can be taken to be flat, (b) the vertical curvature effects are negligible, and (c) the lateral length effects are negligible.

Challenge 192, page 138: For a powerful bus, the acceleration is $2\,\mathrm{m/s^2}$; in 100 m of acceleration, this makes a relative frequency change of $2.2 \cdot 10^{-15}$.

Challenge 193, page 139: Yes, light absorption and emission are always lossless conversions of energy into mass.

Challenge 196, page 140: For a beam of light, in both cases the situation is described by an environment in which masses 'fall' against the direction of motion. If the Earth and the train walls were not visible – for example if they were hidden by mist – there would not be any way to determine by experiment which situation is which. Or again, if an observer would be enclosed in a box, he could not distinguish between constant acceleration or constant gravity. (Important: this impossibility only applies if the observer has negligible size!)

Challenge 200, page 141: Length is time times the speed of light. If time changes with height, so do lengths.

Challenge 202, page 141: Both fall towards the centre of the Earth. Orbiting particles are also in free fall; their relative distance changes as well, as explained in the text.

Challenge 205, page 144: Such a graph would need four or even 5 dimensions.

Challenge 206, page 145: The experiments about change of time with height can also be used in this case.

Challenge 208, page 146: The energy due to the rotation can be neglected compared with all other energies in the problem.

Challenge 218, page 151: Different nucleons, different nuclei, different atoms and different molecules have different percentages of binding energies relative to the total mass.

Challenge 220, page 153: In free fall, the bottle and the water remain at rest with respect to each other.

Challenge 221, page 153: Let the device fall. The elastic rubber then is strong enough to pull the ball into the cup. See M. T. Westra, *Einsteins verjaardagscadeau*, Nederlands tijdschrift voor natuurkunde **69**, p. 109, April 2003. The original device also had a spring connected in series to the rubber.

Challenge 222, page 153: Apart from the chairs and tables already mentioned, important anti-gravity devices are suspenders, belts and plastic bags.

Challenge 224, page 154: The same amount.

Challenge 225, page 154: Yes, in gravity the higher twin ages more. The age difference changes with height, and reaches zero for infinite height.

Challenge 226, page 154: The mass flow limit is $c^3/4G$.

Challenge 227, page 154: No, the conveyer belt can be built into the train.

Challenge 228, page 154: They use a spring scale, and measure the oscillation time. From it they deduce their mass. NASA's bureaucracy calls it a BMMD, a body mass measuring device. A photograph is found in the first volume. Vol. I, page 108

Challenge 229, page 154: The apple hits the wall after about half an hour.

Challenge 232, page 155: Approaches with curved light paths, or with varying speed of light do not describe horizons properly.

Challenge 233, page 155: With $\hbar$ as smallest angular momentum one get about 100 Tm.

Challenge 234, page 155: No. The diffraction of the beams does not allow it. Also quantum theory makes this impossible; bound states of massless particles, such as photons, are not stable.

Challenge 236, page 156: The orbital radius is 4.2 Earth radii; that makes *c.* 38 µs every day.

Challenge 237, page 157: To be honest, the experiments are not consistent. They assume that some other property of nature is constant – such as atomic size – which in fact also depends on G. More on this issue on page 292.

Challenge 238, page 157: Of course other spatial dimensions could exist which can be detected only with the help of measurement apparatuses. For example, hidden dimensions could appear at energies not accessible in everyday life.

Challenge 239, page 157: On this tiny effect, see the text by Ohanian, Ref. 113, on page 147.

Challenge 262, page 172: Since there is no negative mass, gravitoelectric fields cannot be neutralized. In contrast, electric fields can be neutralized around a metallic conductor with a Faraday cage.

Challenge 265, page 174: To find the answer, thinking about the electromagnetic analogy helps.

Challenge 274, page 182: One needs to measure the timing of pulses which cross the Earth at different gravitational wave detectors on Earth.

Challenge 248, page 162: They did so during a solar eclipse.

Challenge 275, page 183: No. For the same reasons that such a electrostatic field is not possible.

Challenge 280, page 186: No, a line cannot have intrinsic curvature. A torus is indeed intrinsically curved; it cannot be cut open to yield a flat sheet of paper.

Challenge 285, page 188: No, they cannot be made from a sheet of paper. The curvature is nonzero everywhere.

Challenge 302, page 196: The trace of the Einstein tensor is the negative of the Ricci scalar; it is thus the negative of the trace of the Ricci tensor.

Challenge 306, page 198: The concept of energy makes no sense for the universe, as the concept is only defined for physical systems, and thus not for the universe itself. See also page 256.

Challenge 313, page 204: Indeed, in general relativity gravitational energy cannot be *localized* in space, in contrast to what one expects and requires from an interaction.

Challenge 324, page 208: Errors in the south-pointing carriage are due to the *geometric phase*, an effect that appears in any case of *parallel transport* in three dimensions. It is the same effect that makes Foucault's pendulum turn. Parallel transport is sometimes also called *Fermi-Walker transport*. The geometric phase is explained in detail in the volume on optics. Vol. III, page 139

Challenge 328, page 210: The European Space Agency is exploring the issue. Join them!

Challenge 331, page 220: There is a good chance that some weak form of Sun jets exist; but a detection will not be easy. (The question whether the Milky Way has jets was part of this text since 2006; they have been discovered in 2010.)

Challenge 333, page 224: If you believe that the two amounts differ, you are prisoner of a belief, namely the belief that your ideas of classical physics and general relativity allow you to extrapolate these ideas into domains where they are not valid, such as behind a horizon. At every horizon, quantum effects are so strong that they *invalidate* such classical extrapolations.

Challenge 334, page 224: A few millimetres.

Challenge 335, page 226: If we assume a diameter of 150 μm and a density of 1000 kg/m^3 for the flour particles, then there are about 566 million particles in one kg of flour. A typical galaxy contains 10^{11} stars; that corresponds to 177 kg of flour.

Challenge 336, page 226: Speed is measured with the Doppler effect, usually by looking at the Lyman-alpha line. Distance is much more difficult to explain. Measuring distances is a science on its own, depending on whether one measures distances of stars in the galaxy, to other galaxies, or to quasars. Any book on astronomy or astrophysics will tell more.

Challenge 337, page 226: See the challenge on page 234.

Challenge 339, page 234: The rabbit observes that all other rabbits seem to move away from him.

Challenge 347, page 240: Stand in a forest in winter, and try to see the horizon. If the forest is very deep, you hit tree trunks in all directions. If the forest is finite in depth, you have chance to see the horizon.

Challenge 361, page 255: No. This is an example of how a seemingly exact description of nature can lead to an unscientific statement, a belief, without any relation to reality.

Challenge 362, page 255: Again no. The statement is a pure belief.

Challenge 364, page 257: The universe does not allow observation from outside. It thus has no state properties.

Challenge 375, page 265: At the horizon, light cannot climb upwards.

Challenge 399, page 277: This happens in the same way that the static electric field comes out of a charge. In both cases, the transverse fields do not get out, but the longitudinal fields do. Quantum theory provides the deeper reason. Real radiation particles, which are responsible for free, transverse fields, cannot leave a black hole because of the escape velocity. However, virtual particles can, as their speed is not bound by the speed of light. All static, longitudinal fields are produced by virtual particles. In addition, there is a second reason. Classical field can come out of a black hole because for an outside observer everything that constitutes the black hole is continuously falling, and no constituent has actually crossed the horizon. The field sources thus are not yet out of reach.

Challenge 403, page 278: The description says it all. A visual impression can be found in the room on black holes in the 'Deutsches Museum' in Munich.

Challenge 405, page 278: On the one hand, black holes can occur through collapse of matter. On the other hand, black holes can be seen as a curved horizon.

Challenge 407, page 279: So far, it seems that all experimental consequences from the analogy match observations; it thus seems that we can claim that the night sky is a black hole horizon.

Nevertheless, the question is not settled, and some prominent physicists do not like the analogy. The issue is also related to the question whether nature shows a symmetry between extremely large and extremely small length scales. This topic is expanded in the last leg of our adventure.

Challenge 411, page 282: Any device that uses mirrors requires electrodynamics; without electrodynamics, mirrors are impossible.

Challenge 413, page 286: The hollow Earth theory is correct if usual distances are consistently changed according to $r_{\text{he}} = R_{\text{Earth}}^2/r$. This implies a quantum of action that decreases towards the centre of the hollow sphere. Then there is no way to prefer one description over the other, except for reasons of simplicity.

Challenge 417, page 297: Mass is a measure of the amount of energy. The 'square of mass' makes no sense.

Challenge 420, page 299: Probably the quantity with the biggest variation is mass, where a prefix for $1\,\text{eV}/\text{c}^2$ would be useful, as would be one for the total mass in the universe, which is about 10^{90} times larger.

Challenge 421, page 300: The formula with $n - 1$ is a better fit. Why?

Challenge 424, page 301: No! They are much too precise to make sense. They are only given as an illustration for the behaviour of the Gaussian distribution. Real measurement distributions are not Gaussian to the precision implied in these numbers.

Challenge 425, page 301: About 0.33 m/s. It is *not* 0.333 m/s and it is *not* any longer strings of threes!

Challenge 427, page 307: The slowdown goes *quadratically* with time, because every new slowdown adds to the old one!

Vol. VI, page 112

Challenge 428, page 308: No, only properties of parts of the universe are listed. The universe itself has no properties, as shown in detail in the last part of this adventure.

Challenge 429, page 349: This could be solved with a trick similar to those used in the irrationality of each of the two terms of the sum, but nobody has found one.

Challenge 430, page 349: There are still discoveries to be made in modern mathematics, especially in topology, number theory and algebraic geometry.

BIBLIOGRAPHY

> A man will turn over half a library to make one book.
>
> Samuel Johnson*

1 Aristotle, *On sense and the sensible*, section 1, part 1, 350 bce. Cited in Jean-Paul Dumont, *Les écoles présocratiques*, Folio Essais, Gallimard, p. 157, 1991. Cited on page 15.

2 The original Latin text of Descartes' letter, dated 22 August 1634, can be read online on gallica.bnf.fr/ark:/12148/bpt6k20740j/f419.image. Cited on page 16.

3 Anonyme, *Demonstration touchant le mouvement de la lumière trouvé par M. Römer de l'Academie Royale des Sciences*, Journal des Scavans pp. 233–236, 1676. An English summary is found in O. C. Rømer, *A demonstration concerning the motion of light*, Philosophical Transactions of the Royal Society **136**, pp. 893–894, 1677. You can read the two papers at dbhs.wvusd.k12.ca.us/webdocs/Chem-History/Roemer-1677/Roemer-1677.html. Cited on page 16.

4 F. Tuinstra, *Rømer and the finite speed of light*, Physics Today 57, pp. 16–17, December 2004. Cited on page 17.

5 The history of the measurement of the speed of light can be found in chapter 19 of the text by Francis A. Jenkins & Harvey E. White, *Fundamentals of Optics*, McGraw-Hill, New York, 1957. Cited on page 17.

6 On the way to perform such measurements, see Sydney G. Brewer, *Do-it-yourself Astronomy*, Edinburgh University Press, 1988. Kepler himself never measured the distances of planets to the Sun, but only *ratios* of planetary distances. The parallax of the Sun from two points of the Earth is at most 8.79″; it was first measured in the eighteenth century. Cited on page 19.

7 For a detailed discussion, see note 32 of A. Gualandi & F. Bònoli, *The search for stellar parallaxes and the discovery of the aberration of light: the observational proofs of the Earth's revolution, Eustachio Manfredi, and the 'Bologna case'*, Journal for the History of Astronomy **20**, pp. 155–172, 2009. Indeed, the term 'aberration' does not appear in Bradley's publication on the issue. Cited on page 18.

8 Aristarchus of Samos, *On the sizes and the distances of the Sun and the Moon*, c. 280 bce, in Michael J. Crowe, *Theories of the World From Antiquity to the Copernican Revolution*, Dover, 1990. Cited on page 19.

9 J. Frercks, *Creativity and technology in experimentation: Fizeau's terrestrial determination of the speed of light*, Centaurus **42**, pp. 249–287, 2000. See also the beautiful web-

---

* Samuel Johnson (1709–1784), famous English poet and intellectual.

site on reconstrutions of historical science experiments at www.uni-oldenburg.de/histodid/forschung/nachbauten. Cited on page 19.

10 The way to make pictures of light pulses with an ordinary photographic camera, without any electronics, is described by M. A. Duguay & A. T. Mattick, *Ultrahigh speed photography of picosecond light pulses and echoes*, Applied Optics **10**, pp. 2162–2170, 1971. The picture on page 20 is taken from it. Cited on page 20.

11 You can learn the basics of special relativity with the help of the web; the simplest and clearest introduction is part of the Karlsruhe physics course, downloadable at www.physikdidaktik.uni-karlsruhe.de. You can also use the physics.syr.edu/research/relativity/RELATIVITY.html web page as a starting point; the page mentions many of the English-language relativity resources available on the web. Links in other languages can be found with search engines. Cited on page 22.

12 See the classic paper G. J. Fishman, *Gamma ray observations of the Crab pulsar – past, present, future*, Annals of the New York Academy of Sciences **655**, pp. 309–318, 1992. Cited on page 22.

13 Observations of gamma-ray bursts show that the speed of light does not depend on the lamp speed to within one part in 10^{20}, as shown by the well-known paper by B. E. Schaefer, *Severe limits on variations of the speed of light with frequency*, Physical Review Letters **82**, pp. 4964–4966, 1999, preprint at arxiv.org/abs/astro-ph/9810479. The result was confirmed most impressively with a bright gamma-ray burst observed in 2009. In this gamma-ray burst, after travelling for ten thousand million years, photons of frequencies that differed by a factor $3 \cdot 10^{10}$ – gamma rays and visible light – still arrived within less than a second from each other. See A.A. Abdo & al., (Fermi GBM/LAT collaborations) *Testing Einstein's special relativity with Fermi's short hard gamma-ray burst GRB090510*, preprint at arxiv.org/abs/0908.1832. Cited on pages 22, 23, 322, and 324.

14 M. Füllekrug, *Probing the speed of light with radio waves at extremely low frequencies*, Physical Review Letters **93**, p. 043901, 2004. Cited on page 23.

15 See, e.g., C. Will, *Theory and Experiment in Gravitational Physics*, revised edition, Cambridge University Press, 1993. See also Ref. 27. Cited on pages 23 and 27.

16 The discussions using binary stars are W. de Sitter, *A proof of the constancy of the speed of light*, Proceedings of the Section of the Sciences – Koninklijke Academie der Wetenschappen **15**, pp. 1297–1298, 1913, W. de Sitter, *On the constancy of the speed of light*, Proceedings of the Section of the Sciences – Koninklijke Academie der Wetenschappen **16**, pp. 395–396, 1913, W. de Sitter, *Ein astronomischer Beweis für die Konstanz der Lichtgeschwindigkeit*, Physikalische Zeitschrift **14**, p. 429, 1913, W. de Sitter, *Über die Genauigkeit, innerhalb welcher die Unabhängigkeit der Lichtgeschwindigkeit von der Bewegung der Quelle behauptet werden kann*, Physikalische Zeitschrift **14**, p. 1267, 1913, For a more recent version, see K. Brecher, *Is the speed of light independent of the velocity of the source?*, Physics Letters **39**, pp. 1051–1054, Errata 1236, 1977. Cited on page 23.

17 The strong limits from observations of gamma-ray bursts were explored in Ref. 13.

Measuring the light speed from rapidly moving stars is another possible test; see the previous reference. Some of these experiments are not completely watertight; a speculation about electrodynamics, due to Ritz, maintains that the speed of light is c only when measured with respect to the source; the light from stars, however, passes through the atmosphere, and its speed might thus be reduced to c.

The famous experiment with light emitted from rapid pions at CERN is not subject to this criticism. It is described in T. Alväger, J. M. Bailey, F. J. M. Farley, J. Kjellman & I. Wallin, *Test of the second postulate of relativity in the GeV region*, Physics Letters **12**, pp. 260–262, 1964. See also T. Alväger & al., *Velocity of high-energy gamma rays*, Arkiv för Fysik **31**, pp. 145–157, 1965.

Another precise experiment at extreme speeds is described by G. R. Kalbfleisch, N. Baggett, E. C. Fowler & J. Alspector, *Experimental comparison of neutrino, anti-neutrino, and muon velocities*, Physical Review Letters **43**, pp. 1361–1364, 1979. Cited on page 23.

18 An overview of experimental results is given in Yuan Zhong Zhang, *Special Relativity and its Experimental Foundations*, World Scientific, 1998. Cited on pages 23, 32, 40, 51, 68, 315, and 326.

19 The beginning of the modern theory of relativity is the famous paper by Albert Einstein, *Zur Elektrodynamik bewegter Körper*, Annalen der Physik **17**, pp. 891–921, 1905. It still well worth reading, and every physicist should have done so. The same can be said of the famous paper, probably written after he heard of Olinto De Pretto's idea, found in Albert Einstein, *Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?*, Annalen der Physik **18**, pp. 639–641, 1905. See also the review Albert Einstein, *Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen*, Jahrbuch der Radioaktivität und Elektronik **4**, pp. 411–462, 1907. These papers are now available in many languages. A later, unpublished review is available in facsimile and with an English translation as Albert Einstein, Hanoch Gutfreund, ed., *Einstein's 1912 Manuscript on the Theory of Relativity*, George Braziller, 2004. Also interesting is the later paper Albert Einstein, *Elementary derivation of the equivalence of mass and energy*, American Mathematical Society Bulletin **41**, pp. 223–230, 1935. All papers and letters of Einstein are now available, both in their original language and in English translation, at einsteinpapers.press.princeton.edu. Cited on pages 24, 27, and 76.

20 Jean van Bladel, *Relativity and Engineering*, Springer, 1984. Cited on page 24.

21 Albert Einstein, *Mein Weltbild*, edited by Carl Selig, Ullstein Verlag, 1998. Cited on page 26.

22 There is even a book on the topic: Hans C. Ohanian, *Einstein's Mistakes: The Human Failings of Genius*, Norton, 2009. On 26 December 1915, Einstein wrote, in a letter to Paul Ehrenfest: ''Es ist bequem mit dem Einstein, jedes Jahr widerruft er, was er das vorige Jahr geschrieben hat.'' A few weeks later, on 17 January 1916, Einstein wrote to Hendrik Lorentz: ''Die Serie meiner Gravitationsarbeiten ist eine Kette von Irrwegen, die aber doch allmählich dem Ziele näher führten.'' See for, example, the book by Henoch Gutfreund & Jürgen Renn, editors, *The Road to Relativity*, Princeton University Press, 2015. Cited on page 26.

23 Albrecht Fölsing, *Albert Einstein – eine Biographie*, Suhrkamp, p. 237, 1993. Cited on pages 27 and 41.

24 Einstein's beautiful introduction, almost without formulae, is Albert Einstein, *Über die spezielle und allgemeine Relativitätstheorie*, Vieweg, 1917 and 1997. For a text with all required mathematics, see Albert Einstein, *The Meaning of Relativity*, Methuen, 1921 and 1956. The posthumous edition also contains Einstein's last printed words on the theory, in the appendix. See also the German text Albert Einstein, *Grundzüge der Relativitätstheorie*, Springer, 1921 expanded 1954, and republished 2002, which also contains the relevant mathematics. Cited on page 27.

25 Julian Schwinger, *Einstein's Legacy*, Scientific American, 1986. Edwin F. Taylor & John A. Wheeler, *Spacetime Physics – Introduction to Special Relativity*, second edition, Freeman, 1992. See also Nick M. J. Woodhouse, *Special Relativity*, Springer, 2003. Cited on pages 27 and 87.

26 Wolfgang Rindler, *Relativity – Special, General and Cosmological*, Oxford University Press, 2001. This is a beautiful book by one of the masters of the field. Cited on pages 27 and 85.

27 R. J. Kennedy & E. M. Thorndike, *Experimental establishment of the relativity of time*, Physical Review **42**, pp. 400–418, 1932. See also H. E. Ives & G. R. Stilwell, *An experimental study of the rate of a moving atomic clock*, Journal of the Optical Society of America **28**, pp. 215–226, 1938, and **31**, pp. 369–374, 1941. For a modern, high-precision versions, see C. Braxmeier, H. Müller, O. Pradl, J. Mlynek, A. Peters & S. Schiller, *New tests of relativity using a cryogenic optical resonator*, Physical Review Letters **88**, p. 010401, 2002. The newest result is in P. Antonini, M. Okhapkin, E. Göklü & S. Schiller, *Test of constancy of speed of light with rotating cryogenic optical resonators*, Physical Review A **71**, p. 050101, 2005, preprint at arxiv.org/abs/gr-qc/0504109, and the subsequent arxiv.org/abs/physics/0510169. See also P. Antonini, M. Okhapkin, E. Göklü & S. Schiller, *Reply to "Comment on 'Test of constancy of speed of light with rotating cryogenic optical resonators' "*, Physical Review A **72**, p. 066102, 2005, preprint at arxiv.org/abs/physics/0602115. Cited on pages 28 and 322.

28 The slowness of the speed of light inside the Sun is due to the frequent scattering of photons by solar matter. The most serious estimate is by R. Mitalas & K. R. Sills, *On the photon diffusion time scale for the Sun*, The Astrophysical Journal **401**, pp. 759–760, 1992, They give a photon escape time of 0.17 Ma, an average photon free mean path of 0.9 mm, an average speed of 0.97 cm/s and a speed at the centre that is ten times smaller. Cited on page 28.

29 L. Vestergaard Hau, S. E. Harris, Z. Dutton & C. H. Behroozi, *Light speed reduction to 17 meters per second in an ultracold atomic gas*, Nature **397**, pp. 594–598, 1999. See also C. Liu, Z. Dutton, C. H. Behroozi & L. Vestergaard Hau, *Observation of coherent optical information storage in an atomic medium using halted light pulses*, Nature **409**, pp. 490–493, 2001, and the comment E. A. Cornell, *Stopping light in its track*, **409**, pp. 461–462, 2001. However, despite the claim, the light pulses have *not* been halted. Cited on page 28.

30 The method of explaining special relativity by drawing a few lines on paper is due to Hermann Bondi, *Relativity and Common Sense: A New Approach to Einstein*, Dover, New York, 1980. See also Dierck-Ekkehard Liebscher, *Relativitätstheorie mit Zirkel und Lineal*, Akademie-Verlag Berlin, 1991. Cited on page 28.

31 S. Reinhardt & al., *Test of relativistic time dilation with fast optical clocks at different velocities*, Nature Physics 3, pp. 861–864, 2007. Cited on page 30.

32 Rod S. Lakes, *Experimental limits on the photon mass and cosmic vector potential*, Physical Review Letters **80**, pp. 1826–1829, 1998. A maximum photon mass of 10^{-47} kg was also deduced from gamma-ray bursts in the paper Ref. 13. Cited on page 32.

33 F. Tuinstra, *De lotgevallen van het dopplereffect*, Nederlands tijdschrift voor natuurkunde 75, p. 296, August 2009. Cited on page 33.

34 R. W. McGowan & D. M. Giltner, *New measurement of the relativistic Doppler shift in neon*, Physical Review Letters **70**, pp. 251–254, 1993. Cited on page 35.

35 R. Lambourne, *The Doppler effect in astronomy*, Physics Education **32**, pp. 34–40, 1997, Cited on page 35.

36 D. Kiefer & al., *Relativistic electron mirrors from nanoscale foils for coherent frequency upshift to the extreme ultraviolet*, Nature Communications 4, p. 1763, 2013. The exploration of relativistic charge systems promises interesting results and applications in the coming decade. Cited on page 35.

37 See, for example, the paper by T. Wilken & al., *A spectrograph for exoplanet observations calibrated at the centimetre-per-second level*, Nature **485**, pp. 611–614, 2012. Cited on page 36.

38 The present record for clock synchronization seems to be 1 ps for two clocks distant 3 km from each other. See A. Valencia, G. Scarcelli & Y. Shih, *Distant clock synchronization using entangled photon pairs*, Applied Physics Letters **85**, pp. 2655–2657, 2004, or arxiv.org/abs/quant-ph/0407204. Cited on page 36.

39 J. Frenkel & T. Kontorowa, *Über die Theorie der plastischen Verformung*, Physikalische Zeitschrift der Sowietunion **13**, p. 1, 1938. F.C. Frank, *On the equations of motion of crystal dislocations*, Proceedings of the Physical Society A **62**, pp. 131–134, 1949. J. Eshelby, *Uniformly moving dislocations*, Proceedings of the Physical Society A **62**, pp. 307–314, 1949. See also G. Leibfried & H. Dietze, Zeitschrift für Physik **126**, p. 790, 1949. A general introduction can be found in A. Seeger & P. Schiller, *Kinks in dislocation lines and their effects in internal friction in crystals*, Physical Acoustics **3A**, W. P. Mason, ed., Academic Press, 1966. See also the textbooks by Frank R.N. Nabarro, *Theory of Crystal Dislocations*, Oxford University Press, 1967, or J. P. Hirth & J. Lothe, *Theory of Dislocations*, McGraw Hill, 1968. Cited on page 36.

40 This beautiful graph is taken from Z.G.T. Guiragossian, G.B. Rothbart, M.R. Yearian, R. Gearhart & J.J. Murray, *Relative velocity measurements of electrons and gamma rays at 15 GeV*, Physical Review Letters **34**, pp. 335–338, 1975. Cited on page 37.

41 A provocative attempt to explain the lack of women in physics in general is made in Margaret Wertheim, *Pythagoras' Trousers – God, Physics and the Gender Wars*, Fourth Estate, 1997. Cited on page 37.

42 To find out more about the best-known crackpots, and their ideas, send an email to majordomo@zikzak.net with the one-line body 'subscribe psychoceramics'. Cited on page 37.

43 The accuracy of Galilean mechanics was discussed by Simon Newcomb already in 1882. For details, see Steven Weinberg, *Gravitation and Cosmology*, Wiley, 1972. Cited on page 38.

44 The speed of neutrinos is the same as that of light to 9 decimal digits. This is explained by Leo Stodolsky, *The speed of light and the speed of neutrinos*, Physics Letters B **201**, p. 353, 1988. An observation of a small mass for the neutrino has been published by the Japanese Super-Kamiokande collaboration, in Y. Fukuda & al., *Evidence for oscillation of atmospheric neutrinos*, Physical Review Letters **81**, pp. 1562–1567, 1998. The newer results published by the Canadian Sudbury Neutrino Observatory, as Q.R. Ahmad & al., *Direct evidence for neutrino flavor transformation from neutral-current interactions in the Sudbury Neutrino Observatory*, Physical Review Letters **89**, p. 011301, 2002, also confirm that neutrinos have a mass in the 1 eV region. Cited on pages 39 and 311.

45 B. Rothenstein & G. Eckstein, *Lorentz transformations directly from the speed of light*, American Journal of Physics **63**, p. 1150, 1995. See also the comment by E. Kapuścik, *Comment on "Lorentz transformations directly from the speed of light" by B. Rothenstein and G. Eckstein*, American Journal of Physics **65**, p. 1210, 1997. Cited on page 40.

46 See e.g. the 1922 lectures by Lorentz at Caltech, published as H. A. Lorentz, *Problems of Modern Physics*, edited by H. Bateman, Ginn and Company, page 99, 1927. Cited on page 40.

47 Max Born, *Die Relativitätstheorie Einsteins*, Springer, 2003, a new, commented edition of the original text of 1920. Cited on page 40.

48 A. A. Michelson & E. W. Morley, *On the relative motion of the Earth and the luminiferous ether*, American Journal of Science (3rd series) **34**, pp. 333–345, 1887. Michelson published many other papers on the topic after this one. Cited on page 40.

49 The newest result is Ch. Eisele, A. Yu. Nevsky & S. Schiller, *Laboratory test of the isotropy of light propagation at the 10^{-17} level*, Physics Review Letters **103**, p. 090401, 2009. See also the older experiment at S. Schiller, P. Antonini & M. Okhapkin, *A precision test of the isotropy of the speed of light using rotating cryogenic cavities*, arxiv.org/abs/physics/0510169. See also the institute page at www.exphy.uni-duesseldorf.de/ResearchInst/WelcomeFP.html. Cited on page 41.

50 H. A. Lorentz, *De relative beweging van de aarde en dem aether*, Amst. Versl. **1**, p. 74, 1892, and also H. A. Lorentz, *Electromagnetic phenomena in a system moving with any velocity smaller than that of light*, Amst. Proc. **6**, p. 809, 1904, or Amst. Versl. **12**, p. 986, 1904. Cited on page 44.

51 A general refutation of such proposals is discussed by S. R. Mainwaring & G. E. Stedman, *Accelerated clock principles*, Physical Review A **47**, pp. 3611–3619, 1993. Experiments on *muons* at CERN in 1968 showed that accelerations of up to 10^{20} m/s^2 have no effect, as explained by D. H. Perkins, *Introduction to High Energy Physics*, Addison-Wesley, 1972, or by J. Bailey & al., Il Nuovo Cimento **9A**, p. 369, 1972. Cited on page 44.

52 W. Rindler, *General relativity before special relativity: an unconventional overview of relativity theory*, American Journal of Physics **62**, pp. 887–893, 1994. Cited on page 45.

53 Steven K. Blau, *Would a topology change allow Ms. Bright to travel backward in time?*, American Journal of Physics **66**, pp. 179–185, 1998. Cited on page 47.

54 On the 'proper' formulation of relativity, see for example D. Hestenes, *Proper particle mechanics*, Journal of Mathematical Physics **15**, pp. 1768–1777, 1974. See also his numerous other papers, his book David Hestenes, *Spacetime Algebra*, Gordon and Breach, 1966, and his webpage modelingnts.la.asu.edu. A related approach is W. E. Baylis, *Relativity in introductory physics*, preprint at arxiv.org/abs/physics/0406158. Cited on page 49.

55 The simple experiment to take a precise clock on a plane, fly it around the world and then compare it with an identical one left in place was first performed by J. C. Hafele & R. E. Keating, *Around-the-world atomic clocks: predicted relativistic time gains*, Science **177**, pp. 166–167, and *Around-the-world atomic clocks: observed relativistic time gains*, pp. 168–170, 14 July 1972. See also Ref. 18. Cited on pages 49 and 140.

56 A readable introduction to the change of time with observers, and to relativity in general, is Roman U. Sexl & Herbert Kurt Schmidt, *Raum-Zeit-Relativität*, 2. Auflage, Vieweg & Sohn, Braunschweig, 1991. Cited on page 49.

57 Most famous is the result that moving muons stay younger, as shown for example by D. H. Frisch & J. B. Smith, *Measurement of the relativistic time dilation using μ-mesons*, American Journal of Physics **31**, pp. 342–355, 1963. For a full pedagogical treatment of the twin paradox, see E. Sheldon, *Relativistic twins or sextuplets?*, European Journal of Physics **24**, pp. 91–99, 2003. Cited on page 50.

58 Paul J. Nahin, *Time Machines – Time Travel in Physics, Metaphysics and Science Fiction*, Springer Verlag and AIP Press, second edition, 1999. Cited on page 50.

59 The first muon experiment was B. Rossi & D. B. Hall, *Variation of the rate of decay of mesotrons with momentum*, Physical Review **59**, pp. 223–228, 1941. 'Mesotron' was the old name for muon. Cited on page 51.

60 J. Bailey & al., *Final report on the CERN muon storage ring including the anomalous magnetic moment and the electric dipole moment of the muon, and a direct test of relativistic time dilation*, Nuclear Physics B **150**, pp. 1–75, 1979. Cited on page 51.

61 A. Harvey & E. Schucking, *A small puzzle from 1905*, Physics Today, pp. 34–36, March 2005. Cited on page 51.

62 Search for 'fuel' and 'relativistic rocket' on the internet. Cited on page 51.

63 W. Rindler, *Length contraction paradox*, American Journal of Physics **29**, pp. 365–366, 1961. For a variation without gravity, see R. Shaw, *Length contraction paradox*, American Journal of Physics **30**, p. 72, 1962. Cited on page 53.

64 H. van Lintel & C. Gruber, *The rod and hole paradox re-examined*, European Journal of Physics **26**, pp. 19–23, 2005. Cited on page 53.

65 See the clear discussion by C. Iyer & G. M. Prabhu, *Differing observations on the landing of the rod into the slot*, American Journal of Physics **74**, pp. 998–1001, 2006, preprint at arxiv.org/abs/0809.1740. Cited on page 53.

66 This situation is discussed by G. P. Sastry, *Is length contraction paradoxical?*, American Journal of Physics **55**, 1987, pp. 943–946. This paper also contains an extensive literature list covering variants of length contraction paradoxes. Cited on page 53.

67 S. P. Boughn, *The case of the identically accelerated twins*, American Journal of Physics **57**, pp. 791–793, 1989. Cited on pages 54 and 55.

68 J. M. Supplee, *Relativistic buoyancy*, American Journal of Physics 57 **1**, pp. 75–77, January 1989. See also G. E. A. Matsas, *Relativistic Arquimedes law for fast moving bodies and the general-relativistic resolution of the 'submarine paradox'*, Physical Review D **68**, p. 027701, 2003, or arxiv.org/abs/gr-qc/0305106. Cited on page 54.

69 The distinction was first published by J. Terrell, *Invisibility of Lorentz contraction*, Physical Review **116**, pp. 1041–1045, 1959, and R. Penrose, *The apparent shape of a relativistically moving sphere*, Proceedings of the Cambridge Philosophical Society **55**, pp. 137–139, 1959. Cited on page 55.

70 G. R. Rybicki, *Speed limit on walking*, American Journal of Physics **59**, pp. 368–369, 1991. Cited on page 58.

71 See, for example, R. J. Nemiroff, Q. Zhong & E. Lilleskov, *Lights illuminate surfaces superluminally*, preprint available at arXiv.org/abs/1506.02643; see also the related and beautiful paper R. J. Nemiroff, *Superluminal spot pair events in astronomical settings: sweeping beams*, preprint available at arXiv.org/abs/1412.7581. Cited on page 60.

72 The first examples of such astronomical observations were provided by A.R. Whitney & al., *Quasars revisited: rapid time variations observed via very-long-baseline interferometry*, Science **173**, pp. 225–230, 1971, and by M.H. Cohen & al., *The small-scale structure of radio galaxies and quasi-stellar sources at 3.8 centimetres*, Astrophysical Journal **170**, pp. 207–217, 1971. See also T. J. Pearson, S. C. Unwin, M. H. Cohen, R. P. Linfield, A. C. S. Readhead, G. A. Seielstad, R. S. Simon & R. C. Walker, *Superluminal expansion of quasar 3C 273*, Nature **290**, pp. 365–368, 1981. An overview is given in J. A. Zensus & T. J. Pearson, editors, *Superluminal radio sources*, Cambridge University Press, 1987. Another measurement, using very long baseline interferometry with radio waves on jets emitted from a binary star (thus not a quasar), was shown on the cover of Nature: I. F. Mirabel & L. F. Rodríguez, *A superluminal source in the galaxy*, Nature **371**, pp. 46–48, 1994. A more recent example was reported in Science News **152**, p. 357, 6 December 1997.

Pedagogical explanations are given by D. C. Gabuzda, *The use of quasars in teaching* *introductory special relativity*, American Journal of Physics **55**, pp. 214–215, 1987, and by Ref. 119 on pages 89-92. Cited on page 60.

73 A somewhat old-fashioned review paper on tachyons mentioning the doubling issue, on pages 52 and 53, is E. Recami, *Classical tachyons and possible applications*, Rivista del Nuovo Cimento **9**, pp. 1–178, 1986. It also discusses a number of paradoxes. By the way, a simple animation claiming to show a flying tachyon can be found in Wikipedia. Cited on page 61.

74 O.M. Bilaniuk & E.C. Sudarshan, *Particles beyond the light barrier*, Physics Today **22**, pp. 43–51, 1969, and O.M.P. Bilaniuk, V.K. Deshpande & E.C.G. Sudarshan, *'Meta' relativity*, American Journal of Physics **30**, pp. 718–723, 1962. See also the book E. Recami, editor, *Tachyons, Monopoles and Related Topics*, North-Holland, 1978. Cited on page 61.

75 J.P. Costella, B.H.J. McKellar, A.A. Rawlinson & G.J. Stephenson, *The Thomas rotation*, American Journal of Physics **69**, pp. 837–847, 2001. Cited on page 62.

76 Planck wrote this in a letter in 1908. Cited on page 63.

77 See for example S.S. Costa & G.E.A. Matsas, *Temperature and relativity*, preprint available at arxiv.org/abs/gr-qc/9505045. Cited on page 63.

78 One of the latest of these debatable experiments is T.P. Krisher, L. Maleki, G.F. Lutes, L.E. Primas, R.T. Logan, J.D. Anderson & C.M. Will, *Test of the isotropy of the one-way speed of light using hydrogen-maser frequency standards*, Physical Review D **42**, pp. 731–734, 1990. Cited on pages 63 and 315.

79 H.C. Ohanian, *The role of dynamics in the synchronization problem*, American Journal of Physics 72, pp. 141–148, 2004. Cited on pages 63 and 315.

80 R.C. Tolman & G.N. Lewis, *The principle of relativity and non-Newtonian mechanics*, Philosophical Magazine **18**, pp. 510–523, 1909, and R.C. Tolman, *Non-Newtonian mechanics: the mass of a moving body*, Philosophical Magazine **23**, pp. 375–380, 1912. Cited on page 65.

81 S. Rainville, J.K. Thompson, E.G. Myers, J.M. Brown, M.S. Dewey, E.G. Kessler, R.D. Deslattes, H.G. Börner, M. Jentschel, P. Mutti & D.E. Pritchard, *World year of physics: a direct test of* $E = mc^2$, Nature **438**, pp. 1096–1097, 2005. Cited on page 71.

82 This information is due to a private communication by Frank DiFilippo; part of the story is given in F. DiFilippo, V. Natarajan, K.R. Boyce & D.E. Pritchard, *Accurate atomic masses for fundamental metrology*, Physical Review Letters 73, pp. 1481–1484, 1994. These measurements were performed with Penning traps; a review of the possibilities they offer is given by R.C. Thompson, *Precision measurement aspects of ion traps*, Measurement Science and Technology **1**, pp. 93–105, 1990. The most important experimenters in the field of single particle levitation were awarded the Nobel Prize in 1989. One of the Nobel Prize lectures can be found in W. Paul, *Electromagnetic traps for neutral and charged particles*, Reviews of Modern Physics **62**, pp. 531–540, 1990. Cited on page 71.

83 J.L. Synge, *Relativity: The Special Theory*, North-Holland, 1956, pp. 208–213. More about antiparticles in special relativity can be found in J.P. Costella, B.H.J. McKellar & A.A. Rawlinson, *Classical antiparticles*, American Journal of Physics **65**, pp. 835–841, 1997. See also Ref. 103. Cited on page 72.

84 M. Cannoni, *Lorentz invariant relative velocity and relativistic binary collisions*, preprint at arxiv.org/abs/1605.00569. Cited on page 74.

85 A. Papapetrou, *Drehimpuls- und Schwerpunktsatz in der relativistischen Mechanik*, Praktika Acad. Athenes **14**, p. 540, 1939, and A. Papapetrou, *Drehimpuls- und Schwerpunktsatz in der Diracschen Theorie*, Praktika Acad. Athenes **15**, p. 404, 1940. See also M. H. L. Pryce, *The mass-centre in the restricted theory of relativity and its connexion with the quantum theory of elementary particles*, Proceedings of the Royal Society in London, A **195**, pp. 62–81, 1948. Cited on page 74.

86 The references preceding Einstein's $E = c^2 m$ are: S. Tolver Preston, *Physics of the Ether*, E. & F.N. Spon, 1875, J. H. Poincaré, *La théorie de Lorentz et le principe de réaction*, Archives néerlandaises des sciences exactes et naturelles **5**, pp. 252–278, 1900, O. De Pretto, *Ipotesi dell'etere nella vita dell'universo*, Reale Istituto Veneto di Scienze, Lettere ed Arti **tomo LXIII**, parte 2, pp. 439–500, Febbraio 1904, F. Hasenöhrl, Berichte der Wiener Akademie **113**, p. 1039, 1904, F. Hasenöhrl, *Zur Theorie der Strahlung in bewegten Körpern*, Annalen der Physik **15**, pp. 344–370, 1904, F. Hasenöhrl, *Zur Theorie der Strahlung in bewegten Körpern – Berichtigung*, Annalen der Physik **16**, pp. 589–592, 1905. Hasenöhrl died in 1915, De Pretto in 1921. All these papers were published before the famous paper by Albert Einstein, *Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?*, Annalen der Physik **18**, pp. 639–641, 1905. Cited on page 76.

87 Umberto Bartocci, *Albert Einstein e Olinto De Pretto: la vera storia della formula più famosa del mondo*, Ultreja, 1998. Cited on page 76.

88 For a clear overview of the various sign conventions in general relativity, see the front cover of the text by Misner, Thorne and Wheeler, Ref. 98. We use the gravitational sign conventions of the text by Ohanian and Ruffini, Ref. 113. Cited on page 78.

89 A jewel among the texts on special relativity is the booklet by Ulrich E. Schröder, *Spezielle Relativitätstheorie*, Verlag Harri Deutsch, 1981. Cited on pages 80 and 83.

90 An improved proposal for the definition of 4-jerk that also cites this book is J. G. Russo & P. K. Townsend, *Relativistic kinematics and stationary motions*, Journal of Physics A: Mathematical and Theoretical **42**, p. 445402, 2009, preprint at arxiv.org/abs/0902.4243. Cited on page 80.

91 G. Stephenson & C. W. Kilmister, *Special Relativity for Physicists*, Longmans, London, 1965. See also W. N. Matthews, *Relativistic velocity and acceleration transformations from thought experiments*, American Journal of Physics 73, pp. 45–51, 2005, and the subsequent J. M. Lévy, *A simple derivation of teh Lorentz transformation and of the accompanying velocity and acceleration changes*, American Journal of Physics 75, pp. 615–618, 2007. Cited on page 80.

92 A readable article showing a photocopy of a letter by Einstein making this point is Lev B. Okun, *The concept of mass*, Physics Today, pp. 31–36, June 1989. The topic is not without controversy, as the letters by readers following that article show; they are found in Physics Today, pp. 13–14 and pp. 115–117, May 1990. The topic is still a source of debates. Cited on page 82.

93 Christian Møller, *The Theory of Relativity*, Clarendon Press, 1952, 1972. This standard textbook has been translated in several languages. Cited on page 83.

94 The famous *no-interaction theorem* states that there is no way to find a Lagrangian that only depends on particle variables, is Lorentz invariant *and* contains particle interactions. It was shown by D. G. Currie, T. F. Jordan & E. C. G. Sudarshan, *Relativistic invariance and Hamiltonian theories of interacting particles*, Review of Modern Physics **35**, pp. 350–375, 1963. Cited on page 83.

95 P. Ehrenfest, *Gleichförmige Rotation starrer Körper und Relativitätstheorie*, Physikalische Zeitschrift **10**, pp. 918–928, 1909. Ehrenfest (incorrectly) suggested that this meant that relativity cannot be correct. A good modern summary of the issue can be found in M. L. Ruggiero, *The relative space: space measurements on a rotating platform*, arxiv.org/abs/gr-qc/0309020. Cited on page 85.

96 R. J. Low, *When moving clocks run fast*, European Journal of Physics **16**, pp. 228–229, 1995. Cited on pages 91, 92, and 93.

97 E. A. Desloge & R. J. Philpott, *Uniformly accelerated reference frames in special relativity*, American Journal of Physics **55**, pp. 252–261, 1987. Cited on page 94.

98 The impossibility of defining rigid coordinate frames for non-uniformly accelerating observers is discussed by Charles Misner, Kip Thorne & John A. Wheeler, *Gravitation*, Freeman, p. 168, 1973. Cited on pages 94 and 329.

99 R. H. Good, *Uniformly accelerated reference frame and twin paradox*, American Journal of Physics **50**, pp. 232–238, 1982. Cited on pages 94, 95, and 99.

100 J. D. Hamilton, *The uniformly accelerated reference frame*, American Journal of Physics **46**, pp. 83–89, 1978. Cited on page 95.

101 The best and cheapest mathematical formula collection remains the one by K. Rottmann, *Mathematische Formelsammlung*, BI Hochschultaschenbücher, 1960. Cited on page 95.

102 C. G. Adler & R. W. Brehme, *Relativistic solutions to a falling body in a uniform gravitation field*, American Journal of Physics **59**, pp. 209–213, 1991. Cited on page 96.

103 See for example the excellent lecture notes by D. J. Raymond, *A radically modern approach to freshman physics*, on the www.physics.nmt.edu/~raymond/teaching.html website. Cited on pages 96 and 328.

104 Edward A. Desloge, *The gravitational red-shift in a uniform field*, American Journal of Physics **58**, pp. 856–858, 1990. Cited on page 99.

105 L. Mishra, *The relativistic acceleration addition theorem*, Classical and Quantum Gravity **11**, pp. L97–L102, 1994. Cited on page 100.

106 Edwin F. Taylor & A. P. French, *Limitation on proper length in special relativity*, American Journal of Physics **51**, pp. 889–893, 1983. Cited on page 102.

107 Clear statements against a varying speed of light are made by Michael Duff in several of his publications, such as M. J. Duff, *Comment on time-variation of fundamental constants*, arxiv.org/abs/hep-th/0208093. The opposite point of view, though incorrect, has been proposed by John Moffat and by João Magueijo, but also by various other authors. Cited on page 104.

108 The quote is form a letter of Gibbs to the American Academy of Arts and Sciences, in which he thanks the Academy for their prize. The letter was read in a session of the Academy and thus became part of the proceedings: J. W. Gibbs, *Proceedings of the American Academy of Arts and Sciences*, **16**, p. 420, 1881. Cited on page 108.

109 It seems that the first published statement of the maximum force as a *fundamental principle* was around the year 2000, in this text, in the chapter on gravitation and relativity. The author discovered the maximum force principle, not knowing the work of others, when searching for a way to derive the results of the last part of this adventure that would be so simple that it would convince even a secondary-school student. In the year 2000, the author told his friends in Berlin about his didactic approach.

Vol. VI, page 57

The *concept* of a maximum force was first proposed, most probably, by Venzo de Sabbata and C. Sivaram in 1993. Also this physics discovery was thus made much too late. In 1995, Corrado Massa took up the idea. Independently, Ludwik Kostro in 1999, Christoph Schiller just before 2000 and Gary Gibbons in the years before 2002 arrived at the same concept. Gary Gibbons was inspired by a book by Oliver Lodge; he explains that the maximum force value follows from general relativity; he does not make a statement about the converse, nor do the other authors. The statement of maximum force as a fundamental principle seems original to Christoph Schiller.

The temporal order of the first papers on maximum force seems to start with a first mention in E. A. RAUSCHER, *The Minkowski metric for a multidimensional geometry*, Lett. Nuovo Cimento 7, pp. 361–377, 1973, "F can be considered an upper bound on force", and a further mention in H.-J. TREDER, *The planckions as largest elementary particles and as smallest test bodies*, Foundations of Physics **15**, pp. 161–166, 1985. Then came the dedicated paper V. DE SABBATA & C. SIVARAM, *On limiting field strengths in gravitation*, Foundations of Physics Letters **6**, pp. 561–570, 1993. It was followed by C. MASSA, *Does the gravitational constant increase?*, Astrophysics and Space Science **232**, pp. 143–148, 1995, and by L. KOSTRO & B. LANGE, *Is c^4/G the greatest possible force in nature?*, Physics Essays **12**, pp. 182–189, 1999. The next references are the paper by G. W. GIBBONS, *The maximum tension principle in general relativity*, Foundations of Physics **32**, pp. 1891–1901, 2002, preprint at arxiv.org/abs/hep-th/0210109 – though he developed the ideas before that date – and the older versions of the present text CHRISTOPH SCHILLER, *Motion Mountain – The Adventure of Physics*, free pdf available at www.motionmountain.net. Then came C. SCHILLER, *Maximum force and minimum distance: physics in limit statements*, preprint at arxiv.org/abs/physics/0309118, and C. SCHILLER, *General relativity and cosmology derived from principle of maximum power or force*, International Journal of Theoretical Physics **44**, pp. 1629–1647, 2005, preprint at arxiv.org/abs/physics/0607090. See also R. BEIG, G. W. GIBBONS & R. M. SCHOEN, *Gravitating opposites attract*, Classical and Quantum Gravity **26**, p. 225013, 2009. preprint at arxiv.org/abs/09071193.

In 2016, Gary Gibbons was not yet convinced maximum force or power can be seen as a *fundamental* physical principle from which general relativity can be *deduced* – though he sees it as a promising conjecture. Cited on pages 108, 113, 118, 124, 133, and 148.

110 See the fundamental paper by A. DISESSA, *Momentum flow as an alternative perspective in elementary mechanics*, **48**, p. 365, 1980, and A. DISESSA, *Erratum: "Momentum flow as an alternative perspective in elementary mechanics" [Am. J. Phys. 48, 365 (1980)]*, **48**, p. 784, 1980. Also the excellent physics textbook by FRIEDRICH HERRMANN, *The Karlsruhe Physics Course*, makes this point extensively; it is free to download in English, Spanish, Russian, Italian and Chinese at www.physikdidaktik.uni-karlsruhe.de/index_en.html. Cited on page 110.

111 C. SCHILLER, *Maximum force and minimum distance: physics in limit statements*, preprint at arxiv.org/abs/physics/0309118; the ideas are also part of the sixth volume of this text, which is freely downloadable at www.motionmountain.net. Cited on pages 111, 113, 124, and 133.

112 The analysis of the first detected gravitational wave event, called GW150914, is presented in B.P. ABBOTT & al., (LIGO Scientific Collaboration and Virgo Collaboration) *Observation of gravitational waves from a binary black hole merger*, Physical Review Letters **116**, p. 061102, 2016, also available for free download at journals.aps.org/prl/pdf/10.1103/PhysRevLett.116.061102. Additional, more detailed papers on the event and the ones that followed in 2015 and 2017 can be found via the website www.ligo.caltech.edu. A recent example is arxiv.org/abs/1706.01812. Cited on pages 113 and 182.

113 H. C. OHANIAN & REMO RUFFINI, *Gravitation and Spacetime*, W.W. Norton & Co., 1994. Another textbook that talks about the power limit is IAN R. KENYON, *General Relativity*, Oxford University Press, 1990. The maximum power is also discussed in

L. Kostro, *The quantity c^5/G interpreted as the greatest possible power in nature*, Physics Essays **13**, pp. 143–154, 2000. Cited on pages 113, 123, 124, 127, 131, 318, 329, and 340.

114 An overview of the literature on analog model of general relativity can be found on Matt Visser's website www.physics.wustl.edu/~visser/Analog/bibliography.html. Cited on page 113.

115 See for example Wolfgang Rindler, *Relativity – Special, General and Cosmological*, Oxford University Press, 2001, p. 70 ff, or Ray d'Inverno, *Introducing Einstein's Relativity*, Clarendon Press, 1992, p. 36 ff. Cited on page 115.

116 See for example A. Ashtekar, S. Fairhust & B. Krishnan, *Isolated horizons: Hamiltonian evolution and the first law*, arxiv.org/abs/gr-qc/0005083. Cited on page 115.

117 T. Jacobson, *Thermodynamics of spacetime: the Einstein equation of state*, Physical Review Letters 75, pp. 1260–1263, 1995 or arxiv.org/abs/gr-qc/9504004. Cited on page 116.

118 See for example Ekkehart Kröner, *Kontinuumstheorie der Versetzungen und Eigenspannungen*, Springer, 1958, volume 5 of the series 'Ergebnisse der angewandten Mathematik'. Kröner shows the similarity between the equations, methods and results of solid-state continuum physics and those of general relativity, including the Ricci formalism. Cited on pages 119 and 209.

119 See the excellent book Edwin F. Taylor & John A. Wheeler, *Spacetime Physics – Introduction to Special Relativity*, second edition, Freeman, 1992. Cited on pages 120 and 328.

120 This counter-example was suggested by Steve Carlip. Cited on page 122.

121 E. R. Caianiello, Lettere al Nuovo Cimento **41**, p. 370, 1984. Cited on page 124.

122 J. D. Barrow & G. W. Gibbons, *Maximum tension: with and without a cosmological constant*, Monthly Notices of the Royal Astronomical Society **446**, pp. 3874–3877, 2014, preprint at arxiv.org/abs/1408.1820. Cited on page 125.

123 A notable exception is the physics teching group in Karlsruhe, who has always taught force in the correct way. See F. Herrmann, *Mengenartige Größen im Physikunterricht*, Physikalische Blätter **54**, pp. 830–832, September 1998. See also the lecture notes on general introductory physics on the website www.physikdidaktik.uni-karlsruhe.de/skripten. Cited on page 129.

124 R. Penrose, *Naked singularities*, Annals of the New York Academy of Sciences **224**, pp. 125–134, 1973. Cited on page 130.

125 G. Huisken & T. Ilmanen, *The Riemannian Penrose inequality*, International Mathematics Research Notices **59**, pp. 1045–1058, 1997. S. A. Hayward, *Inequalities relating area, energy, surface gravity and charge of black holes*, Physical Review Letters **81**, pp. 4557–4559, 1998. Cited on page 130.

126 C. Will, *The Confrontation between General Relativity and Experiment*, Living Reviews in Relativity 17, 2014, available freely at www.livingreviews.org/lrr-2014-4. An older and more extensive reference is Clifford M. Will, *Was Einstein Right? – Putting General Relativity to the Test*, Oxford University Press, 1993. See also his paper arxiv.org/abs/gr-qc/9811036. Cited on pages 131 and 336.

127 The measurement results by the WMAP satellite are summarized on the website map.gsfc.nasa.gov/m_mm.html; the papers are available at lambda.gsfc.nasa.gov/product/map/current/map_bibliography.cfm. Cited on page 132.

128 The simplest historical source is Albert Einstein, Sitzungsberichte der Preussischen Akademie der Wissenschaften II pp. 844–846, 1915. It is the first explanation of the general theory of relativity, in only three pages. The theory is then explained in detail in the famous article Albert Einstein, *Die Grundlage der allgemeinen Relativitätstheorie*, Annalen der Physik **49**, pp. 769–822, 1916. The historic references can be found in German and English in John Stachel, ed., *The Collected Papers of Albert Einstein*, Volumes 1–9, Princeton University Press, 1987–2004.

Below is a selection of English-language textbooks for deeper study, in ascending order of depth and difficulty:

- An entertaining book without any formulae, but nevertheless accurate and detailed, is the paperback by Igor Novikov, *Black Holes and the Universe*, Cambridge University Press, 1990.
- Almost no formulae, but loads of insight, are found in the enthusiastic text by John A. Wheeler, *A Journey into Gravity and Spacetime*, W.H. Freeman, 1990.
- An excellent presentation is Edwin F. Taylor & John A. Wheeler, *Exploring Black Holes: Introduction to General Relativity*, Addison Wesley Longman, 2000.
- Beauty, simplicity and shortness are the characteristics of Malcolm Ludvigsen, *General Relativity, a Geometric Approach*, Cambridge University Press, 1999.
- Good explanation is the strength of Bernard Schutz, *Gravity From the Ground Up*, Cambridge University Press, 2003.
- A good overview of experiments and theory is given in James Foster & J. D. Nightingale, *A Short Course in General Relativity*, Springer Verlag, 2nd edition, 1998.
- A pretty text is Sam Lilley, *Discovering Relativity for Yourself*, Cambridge University Press, 1981.
- A modern text is by Ray d'Inverno, *Introducing Einstein's Relativity*, Clarendon Press, 1992. It includes an extended description of black holes and gravitational radiation, and regularly refers to present research.
- A beautiful, informative and highly recommended text is H. C. Ohanian & Remo Ruffini, *Gravitation and Spacetime*, W.W. Norton & Co., 1994.
- A well written and modern book, with emphasis on the theory, by one of the great masters of the field is Wolfgang Rindler, *Relativity – Special, General and Cosmological*, Oxford University Press, 2001.
- A classic is Steven Weinberg, *Gravitation and Cosmology*, Wiley, 1972.
- The passion of general relativity can be experienced also in John Klauder, ed., *Magic without Magic: John Archibald Wheeler – A Collection of Essays in Honour of His Sixtieth Birthday*, W.H. Freeman & Co., 1972.
- An extensive text is Kip S. Thorne, *Black Holes and Time Warps – Einstein's Outrageous Legacy*, W.W. Norton, 1994.
- The most mathematical – and toughest – text is Robert M. Wald, *General Relativity*, University of Chicago Press, 1984.
- Much information about general relativity is available on the internet. As a good starting point for US-American material, see the math.ucr.edu/home/baez/physics/ website.

There is still a need for a large and modern textbook on general relativity, with colour material, that combines experimental and theoretical aspects. For texts in other languages, see the next reference. Cited on pages 136, 162, 164, 202, and 203.

129 A beautiful German teaching text is the classic G. Falk & W. Ruppel, *Mechanik, Relativität, Gravitation – ein Lehrbuch*, Springer Verlag, third edition, 1983.

A practical and elegant booklet is Ulrich E. Schröder, *Gravitation – Einführung in die allgemeine Relativitätstheorie*, Verlag Harri Deutsch, Frankfurt am Main, 2001.

A modern reference is TORSTEN FLIESSBACH, *Allgemeine Relativitätstheorie*, Akademischer Spektrum Verlag, 1998.

Excellent is HUBERT GOENNER, *Einführung in die spezielle und allgemeine Relativitätstheorie*, Akademischer Spektrum Verlag, 1996.

In Italian, there is the beautiful, informative, but expensive H. C. OHANIAN & REMO RUFFINI, *Gravitazione e spazio-tempo*, Zanichelli, 1997. It is highly recommended. A modern update of that book would be without equals. Cited on pages 136, 162, 164, 179, 181, 203, and 338.

130 P. MOHAZZABI & J. H. SHEA, *High altitude free fall*, American Journal of Physics **64**, pp. 1242–1246, 1996. As a note, due to a technical failure Kittinger had his hand in (near) vacuum during his ascent, without incurring any permanent damage. On the consequences of human exposure to vacuum, see the www.sff.net/people/geoffrey.landis/vacuum.html website. Cited on page 137.

131 This story is told by W. G. UNRUH, *Time, gravity, and quantum mechanics*, preprint available at arxiv.org/abs/gr-qc/9312027. Cited on page 137.

132 H. BONDI, *Gravitation*, European Journal of Physics **14**, pp. 1–6, 1993. Cited on page 138.

133 J. W. BRAULT, Princeton University Ph.D. thesis, 1962. See also J. L. SNIDER, Physical Review Letters **28**, pp. 853–856, 1972, and for the star Sirius see J.L. GREENSTEIN & al., Astrophysical Journal **169**, p. 563, 1971. Cited on pages 140 and 290.

134 See the detailed text by JEFFREY CRELINSTEN, *Einstein's Jury – The Race to Test Relativity*, Princeton University Press, 2006, which covers all researchers involved in the years from 1905 to 1930. Cited on page 140.

135 The famous paper is R. V. POUND & G. A. REBKA, *Apparent weight of photons*, Physical Review Letters **4**, pp. 337–341, 1960. A higher-precision version was published by R. V. POUND & J. L. SNIDER, Physical Review Letters **13**, p. 539, 1964, and R. V. POUND & J. L. SNIDER, Physical Review B **140**, p. 788, 1965. Cited on pages 140 and 290.

136 R.F.C. VESSOT & al., *Test of relativistic gravitation with a space-borne hydrogen maser*, Physical Review Letters **45**, pp. 2081–2084, 1980. The experiment was performed in 1976; there are more than a dozen co-authors involved in this work, which involved shooting a maser into space with a scout missile to a height of *c.* 10 000 km. Cited on page 140.

137 L. BRIATORE & S. LESCHIUTTA, *Evidence for Earth gravitational shift by direct atomic-time-scale comparison*, Il Nuovo Cimento **37B**, pp. 219–231, 1977. Cited on page 140.

138 More information about tides can be found in E. P. CLANCY, *The Tides*, Doubleday, New York, 1969. Cited on page 142.

139 The expeditions had gone to two small islands, namely to Sobral, north of Brazil, and to Principe, in the gulf of Guinea. The results of the expedition appeared in *The Times* before they appeared in a scientific journal. Today this would be called unprofessional. The results were published as F. W. DYSON, A. S. EDDINGTON & C. DAVIDSON, Philosophical Transactions of the Royal Society (London) **220A**, p. 291, 1920, and Memoirs of the Royal Astronomical Society **62**, p. 291, 1920. Cited on page 143.

140 D. KENNEFICK, *Testing relativity from the 1919 eclipse – a question of bias*, Physics Today pp. 37–42, March 2009. This excellent article discusses the measurement errors in great detail. The urban legend that the star shifts were so small on the negatives that they implied large measurement errors is wrong – it might be due to a lack of respect on the part of some physicists for the abilities of astronomers. The 1979 reanalysis of the measurement confirm that such small shifts, smaller than the star image diameter, are reliably measurable. In fact, the 1979 reanalysis of the data produced a smaller error bar than the 1919 analysis. Cited on page 143.

141 A good source for images of space-time is the text by G. F. R. Ellis & R. Williams, *Flat and Curved Space-times*, Clarendon Press, Oxford, 1988. Cited on page 144.

142 J. Droste, *Het veld van een enkel centrum in Einstein's theorie der zwaartekracht, en de beweging van een stoffelijk punt*, Verslag gew. Vergad. Wiss. Amsterdam **25**, pp. 163–180, 1916. Cited on page 146.

143 The name *black hole* was introduced in 1967 at a pulsar conference, as described in his autobiography by John A. Wheeler, *Geons, Black Holes, and Quantum Foam: A Life in Physics*, W.W. Norton, 1998, pp. 296–297: 'In my talk, I argued that we should consider the possibility that at the center of a pulsar is a gravitationally completely collapsed object. I remarked that one couldn't keep saying "gravitationally completely collapsed object" over and over. One needed a shorter descriptive phrase. "How about black hole?" asked someone in the audience. I had been searching for just the right term for months, mulling it over in bed, in the bathtub, in my car, whenever I had quiet moments. Suddenly, this name seemed exactly right. When I gave a more formal ... lecture ... a few weeks later on, on December 29, 1967, I used the term, and then included it into the written version of the lecture published in the spring of 1968 ... I decided to be casual about the term "black hole", dropping it into the lecture and the written version as if it were an old familiar friend. Would it catch on? Indeed it did. By now every schoolchild has heard the term.'

The widespread use of the term began with the article by R. Ruffini & J. A. Wheeler, *Introducing the black hole*, Physics Today **24**, pp. 30–41, January 1971.

In his autobiography, Wheeler also writes that the expression 'black hole has no hair' was criticized as 'obscene' by Feynman. This is a bizarre comment, given that Feynman used to write his papers in topless bars. Cited on pages 146, 262, 263, and 269.

144 L. B. Kreuzer, *Experimental measurement of the equivalence of active and passive gravitational mass*, Physical Review **169**, pp. 1007–1012, 1968. With a clever experiment, he showed that the gravitational masses of fluorine and of bromine are equal. Cited on page 147.

145 A good and accessible book on the topic is David Blair & Geoff McNamara, *Ripples on a cosmic sea*, Allen & Unwin, 1997. Cited on page 147.

146 That bodies fall along geodesics, independently of their mass, the so-called weak equivalence principle, has been checked by many experiments, down to the 10^{-13} level. The most precise experiments use so-called *torsion balances*. See, for example, the website of the Eőt-Wash group at www.npl.washington.edu/eotwash/experiments/experiments.html. Cited on page 151.

147 So far, the experiments confirm that electrostatic and (strong) nuclear energy fall like matter to within one part in 10^{8}, and weak (nuclear) energy to within a few per cent. This is summarized in Ref. 151. Cited on page 151.

148 J. Soldner, *Berliner Astronomisches Jahrbuch auf das Jahr 1804*, 1801, p. 161. Cited on page 152.

149 See for example K. D. Olum, *Superluminal travel requires negative energies*, Physical Review Letters **81**, pp. 3567–3570, 1998, or M. Alcubierre, *The warp drive: hyper-fast travel within general relativity*, Classical and Quantum Gravity **11**, pp. L73–L77, 1994. See also Chris Van Den Broeck, *A warp drive with more reasonable total energy requirements*, Classical and Quantum Gravity **16**, pp. 3973–3979, 1999. Cited on page 155.

150 See the *Astronomical Almanac*, and its *Explanatory Supplement*, H.M. Printing Office, London and U.S. Government Printing Office, Washington, 1992. For the information about various time coordinates used in the world, such as barycentric coordinate time, the time at the barycentre of the solar system, see also the tycho.usno.navy.mil/systime.html web page. It also contains a good bibliography. Cited on page 155.

151 An overview is given in Clifford Will, *Theory and Experiment in Gravitational Physics*, chapter 14.3, revised edition, Cambridge University Press, 1993. Despite being a standard reference, Will's view of the role of tides and the role of gravitational energy within the principle of equivalence has been criticised by other researchers. See also Ref. 126. Cited on pages 156, 162, and 335.

152 The calculation omits several smaller effects, such as rotation of the Earth and red-shift. For the main effect, see Edwin F. Taylor, *'The boundaries of nature: special and general relativity and quantum mechanics, a second course in physics' – Edwin F. Taylor's acceptance speech for the 1998 Oersted Medal presented by the American Association of Physics Teachers, 6 January 1998*, American Journal of Physics **66**, pp. 369–376, 1998. Cited on page 157.

153 A. G. Lindh, *Did Popper solve Hume's problem?*, Nature **366**, pp. 105–106, 11 November 1993, Cited on page 157.

154 See the paper P. Kaaret, S. Piraino, P. F. Bloser, E. C. Ford, J. E. Grindlay, A. Santangelo, A. P. Smale & W. Zhang, *Strong Field Gravity and X-Ray Observations of 4U1820-30*, Astrophysical Journal **520**, pp. L37–L40, 1999, or at arxiv.org/abs/astro-ph/9905236. The beautiful graphics at the research.physics.uiuc.edu/CTA/movies/spm website illustrate this star system. Cited on page 157.

155 R. J. Nemiroff, *Visual distortions near a black hole and a neutron star*, American Journal of Physics **61**, pp. 619–632, 1993. Cited on page 157.

156 The equality was first tested with precision by R. von Eötvös, Annalen der Physik & Chemie **59**, p. 354, 1896, and by R. von Eötvös, V. Pekár, E. Fekete, *Beiträge zum Gesetz der Proportionalität von Trägheit und Gravität*, Annalen der Physik 4, Leipzig **68**, pp. 11–66, 1922. Eötvös found agreement to 5 parts in 10^9. More experiments were performed by P. G. Roll, R. Krotkow & R. H. Dicke, *The equivalence of inertial and passive gravitational mass*, Annals of Physics (NY) **26**, pp. 442–517, 1964, one of the most interesting and entertaining research articles in experimental physics, and by V. B. Braginsky & V. I. Panov, Soviet Physics – JETP **34**, pp. 463–466, 1971. Modern results, with errors less than one part in 10^{12}, are by Y. Su & al., *New tests of the universality of free fall*, Physical Review **D50**, pp. 3614–3636, 1994. Several future experiments have been proposed to test the equality in space to less than one part in 10^{16}. Cited on pages 158 and 290.

157 Nigel Calder, *Einstein's Universe*, Viking, 1979. Weizmann and Einstein once crossed the Atlantic on the same ship. Cited on page 160.

158 L. Lerner, *A simple calculation of the deflection of light in a Schwarzschild gravitational field*, American Journal of Physics **65**, pp. 1194–1196, 1997. Cited on page 161.

159 A. Einstein, *Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes*, Annalen der Physik 35, p. 898, 1911. Cited on page 162.

160 O. Titov, *Testing of general relativity with geodetic VLBI*, preprint at https://arxiv.org/abs/1702.06647. Cited on pages 162 and 163.

161 I. I. Shapiro, & al., *Fourth test of general relativity*, Physical Review Letters **13**, pp. 789–792, 1964. Cited on page 163.

162 I. I. Shapiro, & al., *Fourth test of general relativity: preliminary results*, Physical Review Letters **20**, pp. 1265–1269, 1968. Cited on page 163.

163 J. H. Taylor, *Pulsar timing and relativistic gravity*, Proceedings of the Royal Society, London A **341**, pp. 117–134, 1992. Cited on pages 164 and 166.

164 B. Bertotti, I. Ciufolini & P. L. Bender, *New test of general relativity: measurement of De Sitter geodetic precession rate for lunar perigee*, Physical Review Letters **58**, pp. 1062–1065, 1987. Later it was confirmed by I.I. Shapiro & al., *Measurement of the De Sitter precession of the moon: a relativistic three body effect*, Physical Review Letters **61**, pp. 2643–2646, 1988. Cited on pages 167 and 290.

165 The Thirring effect was predicted in H. Thirring, *Über die Wirkung rotierender ferner Massen in der Einsteinschen Gravitationstheorie*, Physikalische Zeitschrift **19**, pp. 33–39, 1918, and in H. Thirring, *Berichtigung zu meiner Arbeit: "Über die Wirkung rotierender Massen in der Einsteinschen Gravitationstheorie"*, Physikalische Zeitschrift **22**, p. 29, 1921. The Thirring–Lense effect was predicted in J. Lense & H. Thirring, *Über den Einfluß der Eigenrotation der Zentralkörper auf die Bewegung der Planeten und Monde nach der Einsteinschen Gravitationstheorie*, Physikalische Zeitschrift **19**, pp. 156–163, 1918. Cited on page 169.

166 W. de Sitter, *On Einstein's theory of gravitation and its astronomical consequences*, Monthly Notes of the Royal Astrononmical Society 77, pp. 155–184, p. 418E, 1916. For a discussion of De Sitter precession and Thirring–Lense precession, see also B. R. Holstein, *Gyroscope precession in general relativity*, American Journal of Physics **69**, pp. 1248–1256, 2001. Cited on page 169.

167 The work is based on the LAGEOS and LAGEOS II satellites and is told in I. Ciufolini, *The 1995–99 measurements of the Thirring–Lense effect using laser-ranged satellites*, Classical and Quantum Gravity **17**, pp. 2369–2380, 2000. See also I. Ciufolini & E. C. Pavlis, *A confirmation of the general relativistic prediction of the Lense–Thirring effect*, Nature **431**, pp. 958–960, 2004. See, however, the next reference. Cited on pages 170 and 290.

168 See the interesting, detailed and disturbing discussion by L. Iorio, *On some critical issues of the LAGEOS-based tests of the Lense–Thirring effect*, Journal of Modern Physics **2**, pp. 210–218, 2011, preprint available at arxiv.org/abs/1104.4464. Cited on pages 170 and 290.

169 On the Gravity Probe B satellite experiment, see the web page einstein.stanford.edu/highlights/status1.html and the papers cited there. Cited on pages 167 and 170.

170 The detection of the Thirring–Lense effect in binary pulsars is presented in R. D. Blandford, *Lense–Thirring precession of radio pulsars*, Journal of Astrophysics and Astronomy **16**, pp. 191–206, 1995. Cited on page 170.

171 G. Holzmüller, Zeitschrift für Mathematik und Physik **15**, p. 69, 1870, F. Tisserand, Comptes Rendus 75, p. 760, 1872, and Comptes Rendus **110**, p. 313, 1890. Cited on page 170.

172 B. Mashhoon, *Gravitoelectromagnetism: a brief review*, arxiv.org/abs/gr-qc/0311030, and B. Mashhoon, *Gravitoelectromagnetism*, arxiv.org/abs/gr-qc/0011014. See also its extensive reference list on gravitomagnetism. Cited on page 171.

173 A. Tartaglia & M. L. Ruggiero, *Gravito-electromagnetism versus electromagnetism*, European Journal of Physics **25**, pp. 203–210, 2004. Cited on page 171.

174 D. Bedford & P. Krumm, *On relativistic gravitation*, American Journal of Physics **53**, pp. 889–890, 1985, and P. Krumm & D. Bedford, *The gravitational Poynting vector and energy transfer*, American Journal of Physics 55, pp. 362–363, 1987. Cited on pages 172 and 179.

175 M. Kramer & al., *Tests of general relativity from timing the double pulsar*, preprint at arxiv.org/abs/astro-ph/0609417. Cited on pages 174 and 290.

176 The discussion of gravitational waves by Poincaré is found in H. Poincaré, *Sur la dynamique de l'électron*, Comptes Rendus de l'Académie des Sciences **140**, pp. 1504–1508, 1905, which can be read online at www.academie-sciences.fr/pdf/dossiers/Poincare/Poincare_pdf/Poincare_CR1905.pdf. Einstein's prediction from an approximation of general relativity, eleven years later, is found in A. Einstein, *Näherungsweise Integration der Feldgleichungen der Gravitation*, Sitzungsberichte der Königlich-Preußischen Akademie der Wissenschaften pp. 688–696, 1916. The first fully correct prediction of gravitational waves is A. Einstein & N. Rosen, *On gravitational waves*, Journal of the Franklin Institute **223**, pp. 43–54, 1937. Despite heated discussions, until his death, Nathan Rosen continued not to believe in the existence of gravitational waves. On the story about the errors of Einstein and Rosen about the reality of the waves, see also en.wikipedia.org/wiki/Sticky_bead_argument. Cited on page 174.

177 The history of how the existence gravitational was proven in all its conceptual details is told in the excellent paper C. D. Hill & P. Nurowski, *How the green light was given for gravitational wave search*, preprint at arxiv.org/abs/1608.08673. Cited on page 174.

178 This is told in John A. Wheeler, *A Journey into Gravity and Spacetime*, W.H. Freeman, 1990. Cited on page 174.

179 See, for example, K. T. McDonald, *Answer to question #49. Why c for gravitational waves?*, American Journal of Physics **65**, pp. 591–592, 1997, and section III of V. B. Braginsky, C. M. Caves & K. S. Thorne, *Laboratory experiments to test relativistic gravity*, Physical Review D **15**, pp. 2047–2068, 1992. Cited on page 176.

180 A proposal to measure the speed of gravity is by S. M. Kopeikin, *Testing the relativistic effect of the propagation of gravity by Very Long Baseline Interferometry*, Astrophysical Journal **556**, pp. L1–L5, 2001, and the experimental data is E. B. Formalont & S. M. Kopeikin, *The measurement of the light deflection from Jupiter: experimental results*, Astrophysical Journal **598**, pp. 704–711, 2003. See also S. M. Kopeikin, *The post-Newtonian treatment of the VLBI experiment on September 8, 2002*, Physics Letters A **312**, pp. 147–157, 2003, or arxiv.org/abs/gr-qc/0212121. Several arguments against the claim were published, such as C. M. Will, *Propagation speed of gravity and the relativistic time delay*, arxiv.org/abs/astro-ph/0301145, and S. Samuel, *On the speed of gravity and the v/c corrections to the Shapiro time delay*, arxiv.org/abs/astro-ph/0304006. The discussion went on, as shown in S. M. Kopeikin & E. B. Formalont, *Aberration and the fundamental speed of gravity in the Jovian deflection experiment*, Foundations of Physics **36**, pp. 1244–1285, 2006, preprint at arxiv.org/abs/astro-ph/0311063. Both sides claim to be right: the experiment claims to deduce the speed of gravity from the lack of a tangential component of the light deflection by the gravity of Jupiter, and the critical side claims that the speed of gravity does not enter in this measurement. If we compare the situation with analogous systems in transparent fluids or solids, which also show no tangential deflection component, we conclude that neither the measurement nor the proposal allow us to deduce information on the speed of gravity. A similar conclusion, but based on other arguments, is found on physics.wustl.edu/cmw/SpeedofGravity.html. Cited on page 177.

181 For an introduction to gravitational waves, see B. F. Schutz, *Gravitational waves on the back of an envelope*, American Journal of Physics **52**, pp. 412–419, 1984. Cited on page 177.

182 The quadrupole formula is explained clearly in the text by Goenner. See Ref. 129. Cited on page 179.

183 The beautiful summary by Daniel Kleppner, *The gem of general relativity*, Physics Today **46**, pp. 9–11, April 1993, appeared half a year before the authors of the cited work, Joseph Taylor and Russel Hulse, received the Nobel Prize in Physics for the discovery of millisecond pulsars. A more detailed review article is J. H. Taylor, *Pulsar timing and relativistic gravity*, Philosophical Transactions of the Royal Society, London A **341**, pp. 117–134, 1992. The original paper is J. H. Taylor & J. M. Weisberg, *Further experimental tests of relativistic gravity using the binary pulsar PSR 1913+16*, Astrophysical Journal **345**, pp. 434–450, 1989. See also J. M. Weisberg, J. H. Taylor & L. A. Fowler, *Pulsar PSR 1913+16 sendet Gravitationswellen*, Spektrum der Wissenschaft, pp. 53–61, December 1981. Cited on page 180.

184 D. R. Lorimer, *Binary and millisecond pulsars*, in www.livingreviews.org/lrr-2005-7, and J. M. Weisberg & J. H. Taylor, *The relativistic binary pulsar B1913+16: thirty years of observations and analysis*, pp. 25–31, in F. A. Rasio & I. H. Stairs, editors, *Binary Radio Pulsars*, Proceedings of a meeting held at the Aspen Center for Physics, USA, 12 Janaury – 16 January 2004, volume 328 of ASP Conference Series, Astronomical Society of the Pacific, 2005. Cited on page 180.

185 W. B. Bonnor & M. S. Piper, *The gravitational wave rocket*, Classical and Quantum Gravity **14**, pp. 2895–2904, 1997, or arxiv.org/abs/gr-qc/9702005. Cited on page 184.

186 Wolfgang Rindler, *Essential Relativity*, Springer, revised second edition, 1977. Cited on page 186.

187 This is told (without the riddle solution) on p. 67, in Wolfgang Pauli, *Relativitätstheorie*, Springer Verlag, Berlin, 2000, the edited reprint of a famous text originally published in 1921. The reference is H. Vermeil, *Notiz über das mittlere Krümmungsmaß einer n-fach ausgedehnten Riemannschen Mannigfaltigkeit*, Göttinger Nachrichten, mathematische–physikalische Klasse p. 334, 1917. Cited on page 187.

188 M. Santander, L. M. Nieto & N. A. Cordero, *A curvature based derivation of the Schwarzschild metric*, American Journal of Physics **65**, pp. 1200–1209, 1997. Cited on pages 191, 193, and 194.

189 Michael H. Soffel, *Relativity in Astronomy, Celestial Mechanics and Geodesy*, Springer Verlag, 1989. Cited on page 192.

190 Richard P. Feynman, Fernando B. Morinigo, William G. Wagner & Brian Hatfield, *Feynman Lectures on Gravitation*, Westview Press, 1995. Cited on page 192.

191 J. C. Baez & E. F. Bunn, *The meaning of Einstein's equation*, American Journal of Physics 73, pp. 644–652, 2005. Cited on page 195.

192 Y. Wang & M. Tegmark, *New dark energy constraints from supernovae, microwave background and galaxy clustering*, Physical Review Letters **92**, p. 241302, 2004, or arxiv.org/astro-ph/0403292. Cited on page 199.

193 Arguments for the emptiness of general covariance are given by John D. Norton, *General covariance and the foundations of general relativity*, Reports on Progress in Physics **56**, pp. 791–858, 1993. The opposite point, including the discussion of 'absolute elements', is made in the book by J. L. Anderson, *Principles of Relativity Physics*, chapter 4, Academic Press, 1967. Cited on page 199.

194 For a good introduction to mathematical physics, see the famous three-women text in two volumes by Yvonne Choquet-Bruhat, Cecile DeWitt-Morette & Margaret Dillard-Bleick, *Analysis, Manifolds, and Physics*, North-Holland, 1996 and 2001. The first edition of this classic appeared in 1977. Cited on page 200.

195 C. G. Torre & I. M. Anderson, *Symmetries of the Einstein equations*, Physical Review Letters **70**, pp. 3525–3529, 1993, or arxiv.org/abs/gr-qc/9302033. Cited on page 202.

196 H. Nicolai, *Gravitational billiards, dualities and hidden symmetries*, arxiv.org//abs/gr-qc/0506031. Cited on page 202.

197 The original paper is R. Arnowitt, S. Deser & C. Misner, *Coordinate invariance and energy expressions in general relativity*, Physical Review **122**, pp. 997–1006, 1961. Cited on page 203.

198 See for example H. L. Bray, *Black holes, geometric flows, and the Penrose inequality in general relativity*, Notices of the AMS **49**, pp. 1372–1381, 2002. Cited on page 203.

199 See for example R.A. Knop & al., *New constraints on Ω_M, Ω_Λ, and w from an independent set of eleven high-redshift supernovae observed with HST*, Astrophysical Journal **598**, pp. 102–137, 2003. Cited on page 203.

200 See for example the paper by K. Dalton, *Gravity, geometry and equivalence*, preprint to be found at arxiv.org/abs/gr-qc/9601004, and L. Landau & E. Lifshitz, *The Classical Theory of Fields*, Pergamon, 4th edition, 1975, p. 241. Cited on page 204.

201 A recent overview on the experimental tests of the universality of free fall is that by R. J. Hughes, *The equivalence principle*, Contemporary Physics **4**, pp. 177–191, 1993. Cited on page 206.

202 The equivalence of the various definitions of the Riemann tensor is explained in most texts on general relativity; see Ref. 113. Cited on page 207.

203 K. Tangen, *Can the Pioneer anomaly have a gravitational origin?*, arxiv.org/abs/gr-qc/0602089. Cited on page 208.

204 H. Dittus & C. Lämmerzahl, *Die Pioneer-Anomalie*, Physik Journal **5**, pp. 25–31, January 2006. Cited on page 209.

205 Black hole analogues appear in acoustics, fluids and several other fields. This is an ongoing research topic. See, for example, M. Novello, S. Perez Bergliaffa, J. Salim, V. De Lorenci & R. Klippert, *Analog black holes in flowing dielectrics*, preprint at arxiv.org/abs/gr-qc/0201061, T. G. Philbin, C. Kuklewicz, S. Robertson, S. Hill, F. Konig & U. Leonhardt, *Fiber-optical analog of the event horizon*, Science **319**, pp. 1367–1379, 2008, O. Lahav, A. Itah, A. Blumkin, C. Gordon & J. Steinhauer, *A sonic black hole in a density-inverted Bose–Einstein condensate*, arxiv.org/abs/0906.1337. Cited on page 209.

206 This famous quote is the first sentence of the final chapter, the 'Beschluß', of Immanuel Kant, *Kritik der praktischen Vernunft*, 1797. Cited on page 211.

207 About the myths around the stars and the constellations, see the text by G. Fasching, *Sternbilder und ihre Mythen*, Springer Verlag, 1993. On the internet there are also the beautiful www.astro.wisc.edu/~dolan/constellations/ and www.astro.uiuc.edu/~kaler/sow/sow.html websites. Cited on page 211.

208 Aetius, Opinions, III, I, 6. See Jean-Paul Dumont, *Les écoles présocratiques*, Folio Essais, Gallimard, 1991, p. 445. Cited on page 211.

209 A. Mellinger, *A color all-sky panorama of the Milky Way*, preprint at arxiv.org/abs/0908.4360. Cited on page 212.

210 P. Jetzer, *Gravitational microlensing*, Naturwissenschaften **86**, pp. 201–211, 1999. Measurements using orbital speeds around the Galaxy gives agree with this value. Cited on pages 214 and 221.

211 Dirk Lorenzen, *Geheimnivolles Universum – Europas Astronomen entschleiern das Weltall*, Kosmos, 2002. See also the beautiful website of the European Southern Observatory at www.eso.org. Cited on page 214.

212 A beautiful introduction to modern astronomy was Paolo Maffei, *I mostri del cielo*, Mondadori Editore, 1976. Cited on page 220.

213 See for example A. N. Cox, ed., *Allen's Astrophysical Quantities*, AIP Press and Springer Verlag, 2000. An overview of optical observations is given by the Sloan Digital Sky Survey at skyserver.sdss.org. More details about the universe can be found in the beautiful text by W. J. Kaufmann & R. A. Fredman, *Universe*, fifth edition, W.H. Freeman & Co., 1999. The most recent discoveries are best followed on the sci.esa.int and hubble.nasa.gov websites. Cited on page 220.

214 D. R. Lorimer, A. J. Faulkner, A. G. Lyne, R. N. Manchester, M. Kramer, M. A. McLaughlin, G. Hobbs, A. Possenti, I. H. Stairs, F. Camilo, M. Burgay, N. D'Amico, A. Corongiu & F. Crawford, *The Parkes multibeam pulsar survey: VI. Discovery and timing of 142 pulsars and a Galactic population analysis*, Monthly Notices of the Royal Astronomical Society preprint at arxiv.org/abs/astro-ph/0607640. Cited on page 221.

215 D. Figer, *An upper limit to the masses of stars*, Nature **434**, pp. 192–194, 2005. Cited on page 222.

216 G. Basri, *The discovery of brown dwarfs*, Scientific American **282**, pp. 77–83, April 2001. Cited on page 222.

217 See the well-written paper by P. M. Woods & C. Thompson, *Soft gamma repeaters and anomalous X-ray pulsars: magnetar candidates*, preprint at arxiv.org/abs/astro-ph/0406133. Cited on page 222.

218 B. M. Gaensler, N. M. McClure-Griffiths, M. S. Oey, M. Haverkorn, J. M. Dickey & A. J. Green, *A stellar wind bubble coincident with the anomalous X-ray pulsar 1E 1048.1-5937: are magnetars formed from massive progenitors?*, The Astrophysical Journal (Letters) **620**, pp. L95–L98, 2005, or arxiv.org/abs/astro-ph/0501563. Cited on page 222.

219 Opposition to the cosmological principle is rare, as experimental data generally supports it. Local deviations are discussed by various cosmologists; the issue is still open. See, for example, D. Wiltshire, *Gravitational energy and cosmic acceleration*, preprint at arxiv.org/abs/0712.3982, and D. Wiltshire, *Dark energy without dark energy*, preprint at arxiv.org/abs/0712.3984. Cited on page 226.

220 C. Wirtz, Scientia **38**, p. 303, 1925, and K. Lundmark, *The motions and the distances of the spiral nebulae*, Monthly Notices of the Royal Astronomical Society **85**, pp. 865–894, 1925. See also G. Stromberg, *Analysis of radial velocities of globular clusters and non-galactic nebulae*, Astrophysical Journal **61**, pp. 353–362, 1925. Cited on page 226.

221 G. Gamow, *The origin of the elements and the separation of galaxies*, Physical Review **74**, p. 505, 1948. Cited on page 227.

222 A. G. Doroshkevich, & I. D. Novikov, Dokl. Akad. Nauk. SSSR **154**, p. 809, 1964. It appeared translated into English a few months later. The story of the prediction was told by Penzias in his Nobel lecture. Cited on page 228.

223 Arno A. Penzias & Robert W. Wilson, *A measurement of excess antenna temperature at 4080 Mcs*, Astrophysical Journal **142**, pp. 419–421, 1965. Cited on page 228.

224 See for example, D. Prialnik, *An Introduction to the Theory of Stellar Structure and Evolution*, Cambridge University Press, 2000. Cited on page 229.

225 Star masses are explored in D. Figier, *An upper limit to the masses of stars*, Nature **434**, pp. 192–194, 2005. Cited on page 229.

226 Macrobius, *Somnium Scipionis*, XIV, 19. See Jean-Paul Dumont, *Les écoles présocratiques*, Folio Essais, Gallimard, 1991, p. 61. Cited on page 230.

227 On the remote history of the universe, see the excellent texts by G. Börner, *The Early Universe – Facts & Fiction*, Springer Verlag, 3rd edition, 1993, or Barry Parker, *Creation – The Story of the Origin and the Evolution of the Universe*, Plenum Press, 1988. For an excellent popular text, see M. Longair, *Our Evolving Universe*, Cambridge University Press, 1996. Cited on page 230.

228 The first oxygen seems to have appeared in the atmosphere, produced by microorganisms, 2.32 thousand million years ago. See A. Becker & al., *Dating the rise of atmospheric oxygen*, Nature **427**, pp. 117–120, 2003. Cited on page 232.

229 Gabriele Walker, *Snowball Earth – The Story of the Great Global Catastrophe That Spawned Life as We Know It*, Crown Publishing, 2003. Cited on page 232.

230 K. Knie, *Spuren einer Sternexplosion*, Physik in unserer Zeit **36**, p. 8, 2005. The first step of this connection is found in K. Knie, G. Korschinek, T. Faestermann, E. A. Dorfi, G. Rugel & A. Wallner, ^{60}*Fe anomaly in a deep-sea manganese crust and implications for a nearby supernova source*, Physics Review Letters **93**, p. 171103, 2004, the second step in N. D. Marsh & H. Svensmark, *Low cloud properties influenced by cosmic rays*, Physics Review Letters **85**, pp. 5004–5007, 2000, and the third step in P. B. de Menocal, *Plio-Pleistocene African climate*, Science **270**, pp. 53–59, 1995. Cited on page 233.

231 A. Friedman, *Über die Krümmung des Raumes*, Zeitschrift für Physik **10**, pp. 377–386, 1922, and A. Friedmann, *Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes*, Zeitschrift für Physik **21**, pp. 326–332, 1924. (In the Latin transliteration, the author aquired a second 'n' in his second paper.) Cited on page 234.

232 H. Knutsen, *Darkness at night*, European Journal of Physics **18**, pp. 295–302, 1997. Cited on pages 240 and 242.

233 See for example P.D. Peşić, *Brightness at night*, American Journal of Physics **66**, pp. 1013–1015, 1998. Cited on page 242.

234 Paul Wesson, *Olbers' paradox and the spectral intensity of extra-galactic background light*, Astrophysical Journal **367**, p. 399, 1991. Cited on page 242.

235 Steven Weinberg, *Gravitation and Cosmology*, John Wiley, 1972. An excellent book written with a strong personal touch and stressing most of all the relation with experimental data. It does not develop a strong feeling for space-time curvature, and does not address the basic problems of space and time in general relativity. Excellent for learning how to actually calculate things, but less for the aims of our present adventure. Cited on pages 242 and 281.

236 Supernova searches are being performed by many research groups at the largest optical and X-ray telescopes. A famous example is the Supernova Cosmology project described at supernova.lbl.gov. Cited on page 244.

237 The experiments are discussed in detail in the excellent review by D. Giulini & N. Straumann, *Das Rätsel der kosmischen Vakuumenergiedichte und die beschleunigte Expansion des Universums*, Physikalische Blätter **556**, pp. 41–48, 2000. See also N. Straumann, *The mystery of the cosmic vacuum energy density and the accelerated expansion of the universe*, European Journal of Physics **20**, pp. 419–427, 1999. Cited on pages 244 and 291.

238 A. Harvey & E. Schucking, *Einstein's mistake and the cosmological contant*, American Journal of Physics **68**, pp. 723–727, 2000. Cited on page 245.

239 The author of the bible explains *rain* in this way, as can be deduced from its very first page, Genesis 1: 6-7. Cited on page 245.

240 Up to his death, Fred Hoyle defended his belief that the universe is not described by the big bang, but by a steady flow; see G. Burbidge, F. Hoyle & J. V. Narlikar, *A different approach to cosmology*, Physics Today **52**, pp. 38–44, 1999. The team has also written a book with the same title, published in 2000. The newest book on the topic is J. V. Narlikar & G. Burbidge, *Facts and speculations in Cosmology*, Cambridge University Press, 2008, well worth reading because it is one of the rare books that are thought-provoking. Cited on page 246.

241 Stephen W. Hawking & G. F. R. Ellis, *The Large Scale Structure of Space-Time*, Cambridge University Press, Cambridge, 1973. Among other things, this reference text discusses the singularities of space-time, and their necessity in the history of the universe. Cited on pages 246, 283, and 346.

242 Augustine, *Confessions*, 398, writes in Book XI: 'My answer to those who ask 'What was god doing before he made Heaven and Earth?' is not 'He was preparing Hell for people who pry into mysteries'. This frivolous retort has been made before now, so we are told, in order to evade the point of the question. But it is one thing to make fun of the questioner and another to find the answer. So I shall refrain from giving this reply. [...] Before God made heaven and earth, he did not make anything at all. [...] But if before Heaven and Earth there was no time, why is it demanded what you [god] did then? For there was no "then" when there was no time.' (Book XI, chapter 12 and 13). Cited on page 248.

243 Stephen Hawking, *A Brief History of Time – From the Big Bang to Black Holes*, 1988. Reading this bestseller is almost a must for any physicist, as it is a frequent topic at dinner parties. Cited on page 248.

244 Star details are explained in many texts on stellar structure and evolution. See for example Rudolf Kippenhahn & Alfred Weigert, *Stellar Structure and Evolution*, Springer, 1990. Cited on page 250.

245 J. Pelt, R. Kayser, S. Refsdal & T. Schramm, *The light curve and the time delay of QSO 0957+561*, Astronomy and Astrophysics **305**, p. 97, 1996. Cited on page 252.

246 F. Zwicky, *Nebulae as gravitational lenses*, Physical Review Letters **51**, p. 290, and F. Zwicky, *On the probability to detect nebulae which act as gravitational lenses*, p. 679, 1937. The negative view by Einstein is found in A. Einstein, *Lens-like action of a star by the deviatioin of light in the gravitational field*, Science **84**, pp. 506–507, 1936. A review on gravitational lensing can even be found online, in the paper by J. Wambsganss, *Gravitational lensing in astronomy*, Living Reviews in Relativity **1-12**, pp. 1–80, 1998, to be found on the www.livingreviews.org/Articles/Volume1/1998-12wamb website.

There is also the book by P. Schneider, J. Ehlers & E. E. Falco, *Gravitational Lenses*, Springer Verlag, Berlin, 1992. Cited on page 252.

247 M. Lachièze-Rey & J.-P. Luminet, *Cosmic topology*, Physics Reports **254**, pp. 135–214, 1995. See also B. F. Roukema, *The topology of the universe*, arxiv.org/abs/astro-ph/0010185 preprint. Cited on page 254.

248 Steve Carlip clarified this point. Cited on page 255.

249 G. F. R. Ellis & T. Rothman, *Lost horizons*, American Journal of Physics **61**, pp. 883–893, 1993. Cited on page 255.

250 A. Guth, *Die Geburt des Kosmos aus dem Nichts – Die Theorie des inflationären Universums*, Droemer Knaur, 1999. Cited on page 256.

251 Entropy values for the universe have been discussed by Ilya Prigogine, *Is Future Given?*, World Scientific, 2003. This was his last book. For a different approach, see G. A. Mena Marugán & S. Carneiro, *Holography and the large number hypothesis*, arxiv.org/abs/gr-qc/0111034. This paper also repeats the often heard statement that the universe has an entropy that is much smaller than the theoretical maximum. The maximum is often estimated to be in the range of 10^{100} k to 10^{120} k. Other authors give 10^{84} k. In 1974, Roger Penrose also made statements about the entropy of the universe. However, it is more correct to state that the entropy of the universe is not a useful quantity, because the universe is not a physical system. Cited on page 257.

252 C. L. Bennet, M. S. Turner & M. White, *The cosmic rosetta stone*, Physics Today **50**, pp. 32–38, November 1997. The cosmic background radiation differs from black hole radiation by less than 0.005 %. Cited on page 258.

253 The lack of expansion in the solar system is explained in detail in E. F. Bunn & D. W. Hogg, *The kinematic origin of the cosmological redshift*, American Journal of Physics 77, pp. 688–694, 2009. Cited on page 258.

254 A pretty article explaining how one can make experiments to find out how the human body senses rotation even when blindfolded and earphoned is described by M.-L. Mittelstaedt & H. Mittelstaedt, *The effect of centrifugal force on the perception of rotation about a vertical axis*, Naturwissenschaften **84**, pp. 366–369, 1997. Cited on page 259.

255 No dependence of *inertial mass* on the distribution of surrounding mass has ever been found in experiments. See, for example, R. H. Dicke, *Experimental tests of Mach's principle*, 7, pp. 359–360, 1961. Cited on page 259.

256 The present status is given in the conference proceedings by Julian Barbour & Herbert Pfister, eds., *Mach's Principle: From Newton's Bucket to Quantum Gravity*, Birkhäuser, 1995. Various formulations of Mach's principle – in fact, 21 different ones – are compared on page 530.

In a related development, in 1953, Dennis Sciama published a paper in which he argues that inertia of a particle is due to the gravitational attraction of all other matter in the universe. The paper is widely quoted, but makes no new statements on the issue. See D. W. Sciama, *On the origin of inertia*, Monthly Notices of the Royal Astronomical Society **113**, pp. 34–42, 1953. Cited on page 259.

257 Information on the rotation of the universe is given in A. Kogut, G. Hinshaw & A. J. Banday, *Limits to global rotation and shear from the COBE DMR four-year sky maps*, Physical Review D 55, pp. 1901–1905, 1997. Earlier information is found in J. D. Barrow, R. Juszkiewicz & D. H. Sonoda, *Universal rotation: how large can it be?*, Monthly Notices of the Royal Astronomical Society **213**, pp. 917–943, 1985. See also J. D. Barrow, R. Juszkiewicz & D. H. Sonoda, *Structure of the cosmic microwave background*, Nature **309**, pp. 397–402, 1983, or E. F. Bunn, P. G. Fereira & J. Silk, *How anisotropic is the universe?*, Physical Review Letters 77, pp. 2883–2886, 1996. Cited on page 260.

258 The issue has been discussed within linearized gravity by Richard Tolman, in his textbook *Relativity, Thermodynamics, and Cosmology*, Clarendon Press, 1934, on pp. 272–290. The exact problem has been solved by A. Peres, *Null electromagnetic fields in general relativity theory*, Physical Review **118**, pp. 1105–1110, 1960, and by W. B. Bonnor, *The gravitational field of light*, Commun. Math. Phys. **13**, pp. 163–174, 1969. See also N. V. Mitskievic & K. K. Kumaradtya, *The gravitational field of a spinning pencil of light*, Journal of Mathematical Physics **30**, pp. 1095–1099, 1989, and P. C. Aichelburg & R. U. Sexl, *On the gravitational field of a spinning particle*, General Relativity and Gravitation 2, pp. 303–312, 1971. Cited on page 260.

259 See the delightful popular account by Igor Novikov, *Black Holes and the Universe*, Cambridge University Press, 1990. The consequences of light decay were studied by M. Bronshtein, *Die Ausdehnung des Weltalls*, Physikalische Zeitschrift der Sowjetunion **3**, pp. 73–82, 1933. Cited on pages 261 and 267.

260 C. L. Carilli, K. M. Menten, J. T. Stocke, E. Perlman, R. Vermeulen, F. Briggs, A. G. de Bruyn, J. Conway & C. P. Moore, *Astronomical constraints on the cosmic evolution of the fine structure constant and possible quantum dimensions*, Physical Review Letters **85**, pp. 5511–5514, 25 December 2000. Cited on page 261.

261 The observations of black holes at the centre of galaxies and elsewhere are summarised by R. Blandford & N. Gehrels, *Revisiting the black hole*, Physics Today **52**, pp. 40–46, June 1999. Cited on pages 262, 273, and 274.

262 An excellent and entertaining book on black holes, without any formulae, but nevertheless accurate and detailed, is the paperback by Igor Novikov, *Black Holes and the Universe*, Cambridge University Press, 1990. See also Edwin F. Taylor & John A. Wheeler, *Exploring Black Holes: Introduction to General Relativity*, Addison Wesley Longman 2000.

For a historical introduction, see the paper by R. Ruffini, *The physics of gravitationally collapsed objects*, pp. 59–118, in *Neutron Stars, Black Holes and Binary X-Ray Sources*, Proceedings of the Annual Meeting, San Francisco, Calif., February 28, 1974, Reidel Publishing, 1975. Cited on page 262.

263 J. Michell, *On the means of discovering the distance, magnitude, etc of the fixed stars*, Philosophical Transactions of the Royal Society London **74**, p. 35, 1784, reprinted in S. Detweiler, *Black Holes – Selected Reprints*, American Association of Physics Teachers, 1982. Cited on page 262.

264 The beautiful paper is R. Oppenheimer & H. Snyder, *On continued gravitational contraction*, Physical Review **56**, pp. 455–459, 1939. Cited on page 265.

265 R. P. Kerr, *Gravitational field of a spinning mass as an example of algebraically special metrics*, Physical Review Letters **11**, pp. 237–238, 1963. Cited on page 269.

266 E. T. Newman, E. Couch, R. Chinnapared, A. Exton, A. Prakash & R. Torrence, *Metric of a rotating, charged mass*, Journal of Mathematical Physics **6**, pp. 918–919, 1965. Cited on page 269.

267 For a summary, see P. O. Mazur, *Black hole uniqueness theorems*, pp. 130–157, in M. A. H. MacCallum, editor, *General Relativity and Gravitation*, Cambridge University Press, 1987, or the update at arxiv.org/abs/hep-th/0101012. See also D. C. Robinson, *Four decades of black hole uniqueness theorems*, available at www.mth.kcl.ac.uk/staff/dc_robinson/blackholes.pdf Cited on page 269.

268 H. P. Künzle & A. K. M. Masood-ul-Alam, *Spherically symmetric static SU(2) Einstein-Yang-Mills fields*, Journal of Mathematical Physics **31**, pp. 928–935, 1990. Cited on page 269.

269 An example of research that shows the tendency of gravitational radiation to produce spherical shapes when black holes collide is L. Rezzolla, R. P. Macedo & J. L. Jaramillo, *Understanding the "anti kick" in the merger of binary black holes*, Physical Review Letters **104**, p. 221101, 2010. Cited on pages 270 and 290.

270 R. Penrose & R. M. Floyd, *Extraction of rotational energy from a black hole*, Nature **229**, pp. 177–179, 1971. Cited on page 271.

271 The mass–energy relation for a rotating black hole is due to D. Christodoulou, *Reversible and irreversible transformations in black hole physics*, Physical Review Letters 25, pp. 1596–1597, 1970. For a general, charged and rotating black hole it is due to D. Christodoulou & R. Ruffini, *Reversible transformations of a charged black hole*, Physical Review D 4, pp. 3552–3555, 1971. Cited on page 272.

272 J. D. Bekenstein, *Black holes and entropy*, Physical Review D7, pp. 2333–2346, 1973. Cited on page 272.

273 On the topic of black holes in the early universe, there are only speculative research papers, as found, for example, on arxiv.org. The issue is not settled yet. Cited on page 273.

274 For information about black holes formation via star collapse, see the Wikipedia article at en.wikipedia.org/wikie/Stellar_black_hole. Cited on page 273.

275 Frederick Lamb, APS meeting 1998 press conference: *Binary star 4U1820-30, 20 000 light years from Earth*, Physics News Update, April 27, 1998. Cited on page 274.

276 The first direct evidence for matter falling into a black hole was published in early 2001 by NASA astronomers led by Joseph Dolan. Cited on page 274.

277 For a readable summary of the Penrose–Hawking singularity theorems, see J. Natàrio, *Relativity and singularities – a short introduction for mathematicians*, preprint at arxiv.org/abs/math.DG/0603190. Details can be found in Ref. 241. Cited on page 274.

278 For an overview of cosmic censorship, see T. P. Singh, *Gravitational collapse, black holes and naked singularities*, arxiv.org/abs/gr-qc/9805066, or R. M. Wald, *Gravitational collapse and cosmic censorship*, arxiv.org/abs/gr-qc/9710068. The original idea is due to R. Penrose, *Gravitational collapse: the role of general relativity*, Rivista del Nuovo Cimento **1**, pp. 252–276, 1969. Cited on page 275.

279 The paradox is discussed in M. A. Abramowicz, *Black holes and the centrifugal force paradox*, Scientific American **266**, pp. 74–81, March 1993, and in the comment by D. N. Page, *Relative alternatives*, Scientific American **266**, p. 5, August 1993. See also M. A. Abramowicz & E. Szuszkiewicz, *The wall of death*, American Journal of Physics **61**, pp. 982–991, 1993, and M. A. Abramowicz & J. P. Lasota, *On traveling round without feeling it and uncurving curves*, American Journal of Physics **54**, pp. 936–939, 1986. Cited on page 277.

280 J. Ehlers, *Introduction – Survey of Problems*, pp. 1–10, in J. Ehlers, editor, *Sistemi gravitazionali isolati in relatività generale*, Rendiconti della scuola internazionale di fisica “Enrico Fermi”, LXVII° corso, Società Italiana di Fisica/North Holland, 1979. Cited on page 278.

281 G. J. Stoney, *On the physical units of nature*, Philosophical Magazine **11**, pp. 381–391, 1881. Cited on page 282.

282 The geometrodynamic clock is discussed in D. E. Brahm & R. P. Gruber, *Limitations of the geometrodynamic clock*, General Relativity and Gravitation **24**, pp. 297–303, 1992. The clock itself was introduced by R. F. Marzke, in his Ph.D. thesis *The theory of measurement in general relativity*, 1959, with John Wheeler as thesis adviser. Cited on page 282.

283 R. Geroch, *Einstein algebras*, Commun. Math. Phys. **26**, pp. 271–275, 1972. Cited on page 283.

284 A. Macdonald, *Einstein's hole argument*, American Journal of Physics **69**, pp. 223–225, 2001. Cited on page 284.

285 Roman U. Sexl, *Die Hohlwelttheorie*, Der mathematisch-naturwissenschaftliche Unterricht **368**, pp. 453–460, 1983. Roman U. Sexl, *Universal conventionalism and space-time.*, General Relativity and Gravitation **1**, pp. 159–180, 1970. See also Roman U. Sexl, *Die Hohlwelttheorie*, in Arthur Scharmann & Herbert Schramm, editors, *Physik,* *Theorie, Experiment, Geschichte, Didaktik – Festschrift für Wilfried Kuhn zum 60. Geburtstag am 6. Mai 1983*, Aulis Verlag Deubner, 1984, pp. 241–258. Cited on page 285.

286 T. Damour, *Experimental tests of relativistic gravity*, arxiv.org/abs/gr-qc/9904057. It is the latest in a series of his papers on the topic; the first was T. Damour, *Was Einstein 100 % right?*, arxiv.org/abs/gr-qc/9412064. Cited on pages 289 and 290.

287 H. Dittus, F. Everitt, C. Lämmerzahl & G. Schäfer, *Die Gravitation im Test*, Physikalische Blätter **55**, pp. 39–46, 1999. Cited on page 289.

288 For theories competing with general relativity, see for example the extensive and excellent review by C. M. Will, *The confrontation between general relativity and experiment*, Living Reviews of Relativity 2001-2014, electronic version at www.livingreviews.org/lrr-2001-4, update at www.livingreviews.org/lrr-2006-3 and preprint at arxiv.org/abs/1403.7377. For example, the absence of the Nordtvedt effect, a hypothetical 28-day oscillation in the Earth–Moon distance, which was looked for by laser ranging experiments without any result, eliminated several competing theories. This effect, predicted by Kenneth Nordtvedt, would only appear if the gravitational energy in the Earth–Moon system would fall in a different way than the Earth and the Moon themselves. For a summary of the measurements, see J. Müller, M. Schneider, M. Soffel & H. Ruder, *Testing Einstein's theory of gravity by analyzing lunar laser ranging data*, Astrophysical Journal Letters **382**, pp. L101–L103, 1991. Cited on page 289.

289 See S. Bässler & al., *Improved test of the equivalence principle for gravitational selfenergy*, Physical Review Letters **83**, pp. 3585–3588, 1999. See also C. M. Will, *Gravitational radiation and the validity of general relativity*, Physics Today **52**, p. 38, October 1999. Cited on page 290.

290 The inverse square dependence has been checked down to 60 μm, as reported by E. Adelberger, B. Heckel & C. D. Hoyle, *Testing the gravitational inverse-square law*, Physics World **18**, pp. 41–45, 2005. Cited on page 290.

291 Almost everything of importance in general relativity is published in the free and excellent internet-based research journal *Living Reviews in Relativity*, to be found at the www.livingreviews.org website. The other important journal in the field is *Classical and Quantum Gravity*. In astrophysics, the central publication is *Astronomy & Astrophysics*. Cited on page 290.

292 The study of chaos in Einstein's field equations is just beginning. See, e.g., L. Bombelli, F. Lombardo & M. Castagnino, *Chaos in Robertson-Walker cosmology*, arxiv.org/abs/gr-qc/9707051. Cited on page 291.

293 The ESA satellite called 'Planck' has measured the polarization of the cosmic microwave background. This will provide more details on galaxy formation. Cited on page 291.

294 A good introduction to the topic of gamma-ray bursts is S. Klose, J. Greiner & D. Hartmann, *Kosmische Gammastrahlenausbrüche – Beobachtungen und Modelle*, Teil I und II, Sterne und Weltraum March and April 2001. Cited on page 291.

295 The field solution database is built around the work of A. Karlhede, which allows one to distinguish between solutions with a limited amount of mathematical computation. Cited on page 291.

296 Beautiful simulated images of wormholes are available, for example on the wonderful website www.tempolimit-lichtgeschwindigkeit.de. However, quantum effects forbid their existence, so that no such image is included here. A basic approach is the one by T. Diemer & M. Hadley, *Charge and the topology of spacetime*, Classical and Quantum Gravity **16**, pp. 3567–3577, 1999, or arxiv.org/abs/gr-qc/9905069 and M. Hadley, *Spin half in clas-*

sical general relativity, Classical and Quantum Gravity **17**, pp. 4187–4194, 2000, or arxiv.org/abs/gr-qc/0004029. Cited on page 291.

297 An important formulation of relativity is A. Ashtekar, *New variables for classical and quantum gravity*, Physical Review Letters 57, pp. 2244–2247, 1986. Cited on page 291.

298 For a review on inflation and early universe, see D. Baumann, *TASI lectures on inflation*, preprint at arxiv.org/abs/0907.5424. Cited on page 291.

299 A well written text on the connections between the big bang and particle physics is by I. L. Rozental, *Big Bang – Big Bounce, How Particles and Fields Drive Cosmic Evolution*, Springer, 1988. For another connection, see M. Nagano & A. A. Watson, *Observations and implications of the ultrahigh energy cosmic rays*, Reviews of Modern Physics **72**, pp. 689–732, 2000. Cited on page 291.

300 Teaching will benefit in particular from new formulations, from concentration on principles and their consequences, as has happened in special relativity, from simpler descriptions at the weak field level, and from future research in the theory of general relativity. The newer textbooks cited above are all steps in these directions. Cited on page 292.

301 G. E. Prince & M. Jerie, *Generalising Raychaudhuri's equation*, in *Differential Geometry and Its Applications*, Proc. Conf., Opava (Czech Republic), August 27-31, 2001, Silesian University, Opava, 2001, pp. 235–242. Cited on page 293.

302 Torsion is presented in R. T. Hammond, *New fields in general relativity*, Contemporary Physics **36**, pp. 103–114, 1995. Cited on page 293.

303 A well-known approach is that by Bekenstein; he proposes a modification of general relativity that modifies univesal, $1/r^2$ gravity at galactic distances. This is done in order to explain the hundreds of measured galactic rotation curves that seem to require such a modification. (This approach is called *modified Newtonian dynamics* or *MOND*.) An introduction is given by Jacob D. Bekenstein, *The modified Newtonian dynamics – MOND – and its implications for new physics*, Contemporary Physics **47**, pp. 387–403, 2006, preprint at arxiv.org/abs/astro-ph/0701848v2. Cited on page 293.

304 *Le Système International d'Unités*, Bureau International des Poids et Mesures, Pavillon de Breteuil, Parc de Saint Cloud, 92310 Sèvres, France. All new developments concerning SI units are published in the journal *Metrologia*, edited by the same body. Showing the slow pace of an old institution, the BIPM launched a website only in 1998; it is now reachable at www.bipm.fr. See also the www.utc.fr/~tthomass/Themes/Unites/index.html website; this includes the biographies of people who gave their names to various units. The site of its British equivalent, www.npl.co.uk/npl/reference, is much better; it provides many details as well as the English-language version of the SI unit definitions. Cited on page 295.

305 The bible in the field of time measurement is the two-volume work by J. Vanier & C. Audoin, *The Quantum Physics of Atomic Frequency Standards*, Adam Hilge, 1989. A popular account is Tony Jones, *Splitting the Second*, Institute of Physics Publishing, 2000.

The site opdafl.obspm.fr/www/lexique.html gives a glossary of terms used in the field. For precision *length* measurements, the tools of choice are special lasers, such as mode-locked lasers and frequency combs. There is a huge literature on these topics. Equally large is the literature on precision *electric current* measurements; there is a race going on for the best way to do this: counting charges or measuring magnetic forces. The issue is still open. On *mass* and atomic mass measurements, see page 71. On high-precision *temperature* measurements, see Volume I, on page 548. Cited on page 296.

306 The unofficial SI prefixes were first proposed in the 1990s by Jeff K. Aronson of the University of Oxford, and might come into general usage in the future. See New Scientist **144**, p. 81, 3 December 1994. Other, less serious proposals also exist. Cited on page 297.

307 See the review by L. Ju, D. G. Blair & C. Zhao, *The detection of gravitational waves*, Reports on Progress in Physics **63**, pp. 1317–1427, 2000. Cited on page 299.

308 See the clear and extensive paper by G. E. Stedman, *Ring laser tests of fundamental physics and geophysics*, Reports on Progress in Physics **60**, pp. 615–688, 1997. Cited on page 299.

309 J. Short, *Newton's apples fall from grace*, New Scientist **2098**, p. 5, 6 September 1997. More details can be found in R. G. Keesing, *The history of Newton's apple tree*, Contemporary Physics **39**, pp. 377–391, 1998. Cited on page 300.

310 The various concepts are even the topic of a separate international standard, ISO 5725, with the title *Accuracy and precision of measurement methods and results*. A good introduction is John R. Taylor, *An Introduction to Error Analysis: the Study of Uncertainties in Physical Measurements*, 2nd edition, University Science Books, Sausalito, 1997. Cited on page 300.

311 P. J. Mohr & B. N. Taylor, *CODATA recommended values of the fundamental physical constants: 1998*, Reviews of Modern Physics **59**, p. 351, 2000. This is the set of constants resulting from an international adjustment and recommended for international use by the Committee on Data for Science and Technology (CODATA), a body in the International Council of Scientific Unions, which brings together the International Union of Pure and Applied Physics (IUPAP), the International Union of Pure and Applied Chemistry (IUPAC) and other organizations. The website of IUPAC is www.iupac.org. Cited on page 302.

312 Some of the stories can be found in the text by N. W. Wise, *The Values of Precision*, Princeton University Press, 1994. The field of high-precision measurements, from which the results on these pages stem, is a world on its own. A beautiful introduction to it is J. D. Fairbanks, B. S. Deaver, C. W. Everitt & P. F. Michaelson, eds., *Near Zero: Frontiers of Physics*, Freeman, 1988. Cited on page 302.

313 For details see the well-known astronomical reference, P. Kenneth Seidelmann, *Explanatory Supplement to the Astronomical Almanac*, 1992. Cited on page 307.

314 See the corresponding reference in the first volume.

Note that little is known about the basic properties of some numbers; for example, it is still not known whether $\pi + e$ is a rational number or not! (It is believed that it is not.) Do you want to become a mathematician? Cited on page 309.

Challenge 429 r

Challenge 430 s

CREDITS

Acknowledgements

Many people who have kept their gift of curiosity alive have helped to make this project come true. Most of all, Peter Rudolph and Saverio Pascazio have been – present or not – a constant reference for this project. Fernand Mayné, Ata Masafumi, Roberto Crespi, Serge Pahaut, Luca Bombelli, Herman Elswijk, Marcel Krijn, Marc de Jong, Martin van der Mark, Kim Jalink, my parents Peter and Isabella Schiller, Mike van Wijk, Renate Georgi, Paul Tegelaar, Barbara and Edgar Augel, M. Jamil, Ron Murdock, Carol Pritchard, Richard Hoffman, Stephan Schiller, Franz Aichinger and, most of all, my wife Britta have all provided valuable advice and encouragement.

Many people have helped with the project and the collection of material. In particular, I thank Steve Carlip, Corrado Massa, Tom Helmond, Gary Gibbons, Ludwik Kostro, Heinrich Neumaier, Peter Brown and David Thornton for interesting discussions on maximum force. Most useful was the help of Mikael Johansson, Bruno Barberi Gnecco, Lothar Beyer, the numerous improvements by Bert Sierra, the detailed suggestions by Claudio Farinati, the many improvements by Eric Sheldon, the detailed suggestions by Andrew Young, the continuous help and advice of Jonatan Kelu, the corrections of Elmar Bartel, and in particular the extensive, passionate and conscientious help of Adrian Kubala.

Important material was provided by Bert Peeters, Anna Wierzbicka, William Beaty, Jim Carr, John Merrit, John Baez, Frank DiFilippo, Jonathan Scott, Jon Thaler, Luca Bombelli, Douglas Singleton, George McQuarry, Tilman Hausherr, Brian Oberquell, Peer Zalm, Martin van der Mark, Vladimir Surdin, Julia Simon, Antonio Fermani, Don Page, Stephen Haley, Peter Mayr, Allan Hayes, Norbert Dragon, Igor Ivanov, Doug Renselle, Wim de Muynck, Steve Carlip, Tom Bruce, Ryan Budney, Gary Ruben, Chris Hillman, Olivier Glassey, Jochen Greiner, squark, Martin Hardcastle, Mark Biggar, Pavel Kuzin, Douglas Brebner, Luciano Lombardi, Franco Bagnoli, Lukas Fabian Moser, Dejan Corovic, Paul Vannoni, John Haber, Saverio Pascazio, Klaus Finkenzeller, Leo Volin, Jeff Aronson, Roggie Boone, Lawrence Tuppen, Quentin David Jones, Arnaldo Uguzzoni, Frans van Nieuwpoort, Alan Mahoney, Britta Schiller, Petr Danecek, Ingo Thies, Vitaliy Solomatin, Carl Offner, Nuno Proença, Elena Colazingari, Paula Henderson, Daniel Darre, Wolfgang Rankl, John Heumann, Joseph Kiss, Martha Weiss, Antonio González, Antonio Martos, André Slabber, Ferdinand Bautista, Zoltán Gácsi, Pat Furrie, Michael Reppisch, Enrico Pasi, Thomas Köppe, Martin Rivas, Herman Beeksma, Tom Helmond, John Brandes, Vlad Tarko, Nadia Murillo, Ciprian Dobra, Romano Perini, Harald van Lintel, Andrea Conti, François Belfort, Dirk Van de Moortel, Heinrich Neumaier, Jarosław Królikowski, John Dahlman, Fathi Namouni, Paul Townsend, Sergei Emelin, Freeman Dyson, S.R. Madhu Rao, David Parks, Jürgen Janek, Daniel Huber, Alfons Buchmann, William Purves, Pietro Redondi, Sergei Kopeikin, Damoon Saghian, Zach Joseph Espiritu, David Thornton, Miles Mutka, Fabrizio Bònoli, plus a number of people who wanted to remain unnamed.

The software tools were refined with extensive help on fonts and typesetting by Michael Zedler and Achim Blumensath and with the repeated and valuable support of Donald Arseneau; help

came also from Ulrike Fischer, Piet van Oostrum, Gerben Wierda, Klaus Böhncke, Craig Upright, Herbert Voss, Andrew Trevorrow, Danie Els, Heiko Oberdiek, Sebastian Rahtz, Don Story, Vincent Darley, Johan Linde, Joseph Hertzlinger, Rick Zaccone, John Warkentin, Ulrich Diez, Uwe Siart, Will Robertson, Joseph Wright, Enrico Gregorio, Rolf Niepraschk and Alexander Grahn.

The typesetting and book design is due to the professional consulting of Ulrich Dirr. The typography was much improved with the help of Johannes Küster and his Minion Math font. The design of the book and its website also owe much to the suggestions and support of my wife Britta.

I also thank the lawmakers and the taxpayers in Germany, who, in contrast to most other countries in the world, allow residents to use the local university libraries.

From 2007 to 2011, the electronic edition and distribution of the Motion Mountain text was generously supported by the Klaus Tschira Foundation.

FILM CREDITS

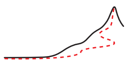
The impressive film of a light pulse bouncing of a mirror on page 21 is copyright and courtesy of Wang Lihong and Washington University at St. Louis. The beautiful animation of a dice flying at relativistic speed, on page 57, is copyright and courtesy by Ute Kraus. It can be found on her splendid website www.tempolimit-lichtgeschwindigkeit.de, which provides many other films of relativistic motions and the related publications. The beautiful animation of an observer accelerating in a desert, on page 91, is copyright Anthony Searle and Australian National University, and courtesy of Craig Savage. It is from the wonderful website at www.anu.edu.au/Physics/Savage/TEE. Also the equally beautiful animation of an observer accelerating between houses, on page 93, is copyright Anthony Searle and Australian National University, and courtesy of Craig Savage. It is from the equally wonderful website at www.anu.edu.au/Physics/Searle. The spectacular animation on page 263 is courtesy and copyright of the European Southern Observatory ESO and found on its website www.eso.org/public/news/eso0846/.

IMAGE CREDITS

The photograph of the east side of the Langtang Lirung peak in the Nepalese Himalayas, shown on the front cover, is courtesy and copyright by Kevin Hite and found on his blog thegettingthere.com. The photograph of the night sky on page 14 is copyright and courtesy of Anthony Ayiomamitis; it is found on his wonderful website www.perseus.gr. The photograph of the reconstruction of Fizeau's experiment on page 20 is copyright by AG Didaktik und Geschichte der Physik, Universität Oldenburg, and courtesy of Jan Frercks, Peter von Heering and Daniel Osewold. The photograph of a light pulse on page 20 is courtesy and copyright of Tom Mattick. On page 24 the photographs of electrical devices are courtesy Miele and EasyGlide. On page 30, the lithium images are courtesy and copyright of the TSR relativity team at the Max Planck Gesellschaft. On page 32 the Doppler images are copyright and courtesy of Maurice Gavin and NASA. On page 33, the Doppler sonar system images are copyright and courtesy of Wikimedia, Hörmann AG and Medison. On page 34, the wave graphic is copyright Pbroks13 and courtesy Wikimedia. On page 41, the image of the historical Michelson experiment is courtesy and copyright of the Astrophysikalisches Institut Potsdam, and the images of the modern high-precision experiment are copyright and courtesy of Stephan Schiller. The relativistic views on page 55 and 57 are courtesy and copyright of Daniel Weiskopf. The relativistic images of the travel through the simplified Stonehenge on page 56 are copyright of Nicolai Mokros and courtesy of Norbert Dragon. On page 68, the photograph of the HARP experiment is courtesy and copyright of CERN. On page 69, the photographs about the bubble chamber are courtesy and copyright of CERN. The stalactite photograph on page 108 is courtesy and copyright of Richard Cindric and found on the website www.kcgrotto.org; the photograph of Saturn is courtesy NASA. The illustration of spatial curvature on page 145 is courtesy and copyright of Farooq Ahmad Bhat. On page 150, the volcano photograph is copyright and courtesy of Marco Fulle and found on the wonderful website www.stromboli.net. On page 181, the VIRGO photographs are courtesy and copyright of INFN. On page 168, the photographs about lunar reflectors are copyright and courtesy NASA and Wikimedia; the photograph of the Nice observatory is courtesy and copyright of Observatoire de la Côte d'Azur. The figures of galaxies on pages 214, 212, 215, 215, 213, 219, 216, 243, 253 and 254 are courtesy of NASA. The photo of the night sky on page 212 is copyright and courtesy of Axel Mellinger; more details on the story of this incredible image is found on his website at home.arcor.de/axel.mellinger. The picture of the universe on page 213 is courtesy of Thomas Jarret, IPAC and Caltech, and is found on the spider.ipac.caltech.edu/staff/jarret/lss/index.html website. The photograph of the molecular cloud on page 216 is courtesy and copyright of the European Southern Observatory ESO; it was also featured on the antwrp.gsfc.nasa.gov/apod/ap030202.html website. On page 217, the photopgraphs of the Very Large Telescopes are copyright and courtesy of ESO. On page 218, the photographs of the XMM-Newton satellite and of the Planck satellite are copyright and courtesy of ESA and found on the fascinting website www.esa.int. The Hubble diagram on page 227 is courtesy of Saul Perlmutter and the Supernova Cosmology Project. The maps of the universe on page 225 and the Hertzsprung–Russell diagram on page 229 are courtesy and copyright of Richard Powell, and taken from his websites www.anzwers.org/free/universe and www.atlasoftheuniverse.com. On page 230, the photograph of M15 is copyright and courtesy of ESA and NASA. The photograph on page 240 is courtesy and copyright of Wally Pacholka and found on the wonderful website www.twanlight.org that collects pictures of the world at night. On page page 241, the tree image is copyright of Aleks G and courtesy of Wikimedia, whereas the Hubble deep sky image is courtesy of NASA and ESA and found at apod.nasa.gov/apod/ap140605.html. On page 243, the Planck data map is courtesy and copyright of Planck/ESA. The simulated view of a black hole on page 264 is copyright and courtesy of Ute Kraus and can be found on her splendid website www.tempolimit-lichtgeschwindigkeit.de. On page 285, the image of the hollow Earth theory is courtesy of Helmut Diel and was drawn by Isolde Diel. On page 312, the drawing of the Fraunhofer lines is copyright and courtesy of Fraunhofer Gesellschaft. The photograph on the back cover, of a basilisk running over water, is courtesy and copyright by the Belgian group TERRA vzw and found on their website www.terravzw.org. All drawings are copyright by Christoph Schiller. If you suspect that your copyright is not correctly given or obtained, this has not been done on purpose; please contact me in this case.

NAME INDEX

SUBJECT INDEX



MOTION MOUNTAIN

The Adventure of Physics – Vol. II

Relativity and Cosmology

What is the most fantastic voyage possible?
Are shadows faster than light?
Can light be accelerated?
How does empty space bend and how do we measure it?
How stiff is empty space?
What are black holes?
What can we see and discover with the best telescopes?
What is the history of the universe?
What are the maximum force and power values in nature?

Answering these and other questions on motion, this series gives an entertaining and mind-twisting introduction into modern physics – one that is surprising and challenging on every page. Starting from everyday life, the adventure provides an overview of modern results in mechanics, heat, electromagnetism, relativity, quantum physics and unification.

Christoph Schiller, PhD Université Libre de Bruxelles, is a physicist and physics popularizer. He wrote this book for his children and for all students, teachers and readers interested in physics, the science of motion.